



# On the optimal control of SIR model with Erlang-distributed infectious period: isolation strategies

Luca Bolzoni<sup>1</sup> · Rossella Della Marca<sup>1</sup>  · Maria Groppi<sup>2</sup>

Received: 18 March 2021 / Revised: 22 July 2021 / Accepted: 8 September 2021 /  
Published online: 22 September 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Mathematical models are formal and simplified representations of the knowledge related to a phenomenon. In classical epidemic models, a major simplification consists in assuming that the infectious period is exponentially distributed, then implying that the chance of recovery is independent on the time since infection. Here, we first attempt to investigate the consequences of relaxing this assumption on the performances of time-variant disease control strategies by using optimal control theory. In the framework of a basic susceptible–infected–removed (SIR) model, an Erlang distribution of the infectious period is considered and optimal isolation strategies are searched for. The objective functional to be minimized takes into account the cost of the isolation efforts per time unit and the sanitary costs due to the incidence of the epidemic outbreak. Applying the Pontryagin’s minimum principle, we prove that the optimal control problem admits only bang–bang solutions with at most two switches. In particular, the optimal strategy could be postponing the starting intervention time with respect to the beginning of the outbreak. Finally, by means of numerical simulations, we show how the shape of the optimal solutions is affected by the different distributions of the infectious period, by the relative weight of the two cost components, and by the initial conditions.

---

✉ Rossella Della Marca  
rossella.dellamarca@izsler.it

Luca Bolzoni  
luca.bolzoni@izsler.it

Maria Groppi  
maria.groppi@unipr.it

<sup>1</sup> Risk Analysis and Genomic Epidemiology Unit, Istituto Zooprofilattico Sperimentale della Lombardia e dell’Emilia Romagna, Via dei Mercati 13, 43126 Parma, Italy

<sup>2</sup> Department of Mathematical, Physical and Computer Sciences, University of Parma, Parco Area delle Scienze 53/A, 43124 Parma, Italy

**Keywords** SIR model · Epidemics · Non-exponential distribution · Erlang distribution · Optimal control · Bang–bang control

**Mathematics Subject Classification** 49K15 · 92D30

## 1 Introduction

Infectious diseases are responsible for significant social and economic losses worldwide. Several recent examples, one for all the ongoing pandemic caused by SARS-CoV-2 virus, pointed out the need for suitable public health interventions for effective disease control and eradication. Also in the case of livestock and wildlife, infectious diseases lead to huge economic losses to farmers (DEFRA/DCMS 2002) and endanger wild species' survival (McCallum and Dobson 1995), in addition to representing potential risks for spillover in human (Daszak et al. 2000).

In the last 20 years, we have witnessed to an increase of the role of mathematical models in improving the understanding and making predictions on the infection dynamics and in informing the policies for disease management (Ferguson et al. 2001; Gatto et al. 2020; Gumel et al. 2004). The susceptible–infected–removed (SIR) class of models represent the most commonly used framework to investigate epidemiological systems through ordinary differential equations (Anderson and May 1992). In SIR models, the host population units are subdivided into different compartments according to their epidemiological status as: susceptibles ( $S$ ), i.e. healthy units that can be infected, infected ( $I$ ), i.e. diseased units that can infect other ones, and removed ( $R$ ), i.e. units who died or recovered from the infection. The units of observation may represent single individuals or groups (e.g. farms in case of livestock diseases).

The classical SIR models are built using different simplified assumptions on the epidemiological processes. One of the major simplifications consists in assuming a constant rate of leaving the infected class, namely an exponential distribution of the infectious period. Effectively, this means that the chance of a unit to recover during any given time period does not depend on the duration of the time that it has already been infected. Whilst such an assumption may provide significant mathematical convenience and be reasonably realistic in some situations, most often the assumption is violated, then requiring the inclusion of more realistic distributions of infectious periods in epidemic models (Bailey 1954; Gough 1977; Simpson 1952).

The consequences of assuming non-exponential distributions of the infectious period in SIR models on (i) the qualitative (Krylova and Earn 2013; Lloyd 2001a) and quantitative (Wearing et al. 2005; Feng et al. 2016) predictions of the infection dynamics, (ii) the estimate of epidemiological parameters (Wearing et al. 2005), and (iii) the effectiveness of time-invariant intervention policies (Feng et al. 2007; Wang et al. 2017) have received some attention in recent years. However, no attempts have been made, to our knowledge, to investigate the consequences of assuming non-exponential distributions of the infectious period on the performances of time-variant disease control strategies by using optimal control theory (Pontryagin et al. 1962).

Here, by investigating the problem of minimizing the epidemic burden by using isolation control, we will show that different assumptions on the shape of the infectious

period distribution may lead to qualitatively and quantitatively different shapes of the optimal control function, thus critically affecting the effectiveness of disease control policies.

The paper is organized as follows. In Sect. 2, we present an SIR model with Erlang-distributed infectious period and isolation control (Erlang distributions are special cases of gamma distributions (Evans and Peacock 1993)). An optimal control problem aimed at minimizing the control costs and the epidemic size is formulated in Sect. 3, where also necessary and sufficient conditions of optimality are derived. The profiles of admissible optimal isolation strategies are analytically characterized in Sect. 4. Then, we provide numerical simulations in order to understand the sensitivity of the optimal solutions to: (i) different assumptions on the distribution of the infectious period; (ii) relative cost of the two components of the objective functional; and, (iii) initial conditions (Sect. 5). Finally, discussion of the presented results and concluding remarks are given in Sect. 6.

## 2 Model formulation

The inclusion of non-exponential distributions of the infectious period in epidemic models means that the chance of recovery depends on the time since infection, and hence the model needs to keep track of this information. Although this can be achieved in several ways, e.g. via an integro-differential equation formulation (Hethcote and Tudor 1980; Keeling and Grenfell 1997) or a partial differential equation formulation (Anderson and May 1992), a significant body of research has been focused on the *method of stages* (see, just to mention a few papers, Cunniffe et al. 2012; Lloyd 2001a, b; Sherborne et al. 2015), first introduced—to the best of our knowledge—by Cox and Miller (1965). It consists in assuming that the infectious period obeys a distribution of the type Erlang( $n, n\gamma$ ), where  $n \in \mathbb{N}_+$  and  $1/\gamma > 0$  is the mean duration of infection (the Erlang variate is a gamma variate with integer shape parameter (Evans and Peacock 1993)). The corresponding probability density function is

$$f(x; n, n\gamma) = \frac{(n\gamma)^n}{(n-1)!} x^{n-1} e^{-n\gamma x}, \quad x \geq 0.$$

The popularity of this assumption comes from the fact that an Erlang distribution with shape parameter  $n$  is nothing else than the distribution of a sum of  $n$  independent exponentially distributed random variables of identical mean  $1/(n\gamma)$ . The case  $n = 1$  reproduces the exponential distribution; as the parameter  $n$  increases, the infectious period distribution becomes more closely centered on its mean, with  $n \rightarrow +\infty$  corresponding to all units having exactly the same duration of infection.

Thus, the use of the Erlang distribution in an epidemic model is equivalent to split the infected compartment in a finite number of fictitious stages in series: newly infected units enter the first stage, pass through each successive stage, and recover as they leave the  $n$ th stage. These multiple stages of infection can be also used to represent periods of increased or decreased risk of transmitting the disease (Ma and Earn 2006).

By adopting the method of stages, we consider the following SIR model with an Erlang( $n, n\gamma$ )-distributed infectious period and including a bounded control function  $u(t) \in U = [0, u_{max}]$  representing the rate of isolation of infected units at time  $t$ :

$$\begin{aligned} \dot{S} &= -\beta SI, \\ \dot{I}_1 &= \beta SI - (n\gamma + u(t))I_1, \\ \dot{I}_j &= n\gamma I_{j-1} - (n\gamma + u(t))I_j, \quad j = 2, \dots, n. \end{aligned} \tag{1}$$

Here, the upper dot denotes the time derivative,  $S$  is the number of susceptible units at time  $t$  and  $I$  is the number of infected units at time  $t$ , which equals the sum of the infected units,  $I_j$ , in each  $j$ th stage, with  $j = 1, \dots, n$ , namely

$$I = \sum_{j=1}^n I_j.$$

The units of observation in model (1) may represent either individuals (in the case of human diseases) or farms (in the case of livestock diseases). The parameters are positive constants:  $\beta$  represents the transmission rate and  $\gamma$  represents the removal rate.

The assumption of upper boundedness for  $u(t)$  reflects the realistic scenario of practical limitations on the maximum rate at which the control strategy may be employed. Such a restriction is due to several factors, including the limited number of health officers, the time necessary for detecting the target individuals or farms, the reduced capacity of infrastructures and so on. Specifically, by setting an upper bound on  $u(t)$ , we are assuming a limitation on the maximum isolation effort relative to the target population instead of on the overall isolation activities performed per time unit (which are represented by  $u(t)I$ ).

For convenience of notation, let us denote  $\mu = n\gamma$  and adopt the following change of variables:

$$Y_k = \sum_{j=1}^k I_j$$

namely,  $Y_k$  is the cumulative of the infected units in the first  $k$  stages of infection. In particular,  $Y_n = I$ . One can easily check that system (1) becomes

$$\begin{aligned} \dot{S} &= -\beta SY_n, \\ \dot{Y}_1 &= \beta SY_n - (\mu + u(t))Y_1, \\ \dot{Y}_k &= \beta SY_n - u(t)Y_k - \mu(Y_k - Y_{k-1}), \quad k = 2, \dots, n. \end{aligned} \tag{2}$$

Then, the general controlled system in compact form reads:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + u(t)\mathbf{g}(\mathbf{x}),$$

where  $\mathbf{x} = (S, Y_1, \dots, Y_n)^T$  is the state-variables vector, and  $\mathbf{f}$  and  $\mathbf{g}$  are given by

$$\mathbf{f} = \begin{pmatrix} -\beta SY_n \\ \beta SY_n - \mu Y_1 \\ \vdots \\ \beta SY_n - \mu(Y_n - Y_{n-1}) \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} 0 \\ -Y_1 \\ \vdots \\ -Y_n \end{pmatrix}. \tag{3}$$

We denote with  $T$  the terminal time of planning (i.e. time horizon). In general, it is not clear how large  $T$  must be in order to coincide with a reasonable interpretation of the end of the epidemic. To overcome the problem, by following the approach adopted by Bolzoni et al. (2019), Bolzoni et al. (2017) and Hansen and Day (2011), we choose  $T$  as the disease *extinction time*. Formally, let  $\varepsilon$  be a positive constant such that  $\varepsilon < 1$ , then

$$T = \inf \{t \in \mathbb{R}_+ | Y_n(t) = \varepsilon\}, \tag{4}$$

which implies that the optimal control problem will have *free* final time. To make the definition (4) meaningful, we choose the initial number of infected units  $Y_n(0) = I(0)$  strictly greater than  $\varepsilon$ . As a consequence,  $T$  is the first time at which the state variable  $Y_n$  drops to  $\varepsilon$ . Using this approach, we also avoid the undesirable possibility of subsequent peaks of infection caused by a fractional number of infected units (Hansen and Day 2011). From a biological point of view, by setting  $\varepsilon < 1$  as the condition to identify the extinction time means assuming the infection goes extinct when there is less than one infected unit in the host population.

To model equations (2), we associate the following epidemiologically well-posed initial conditions:

$$S(0) = S_0 > 0, \quad Y_k(0) = Y_{k,0} \geq 0, \quad k = 1, \dots, n - 1, \quad Y_n(0) = Y_{n,0} > \varepsilon. \tag{5}$$

Note that, by simply looking at the first and the last equation in (2), one gets that the set

$$\mathcal{D} = \left\{ \mathbf{x} \in \mathbb{R}_+^{n+1} \mid S \leq S_0, \quad S + Y_n \leq S_0 + Y_{n,0} \right\} \tag{6}$$

is positively invariant, i.e. solutions of (2) starting in  $\mathcal{D}$  remain in  $\mathcal{D}$  for all  $t \geq 0$ . Therefore, it is not restrictive to limit our analyses to region  $\mathcal{D}$ .

### 3 The optimal control problem

This section is devoted to the formalization of our control problem. Specifically, we set an optimal control problem involving two different and coexisting goals, namely minimization of:

- costs due to the isolation efforts per time unit:

$$\int_0^T A_0 u(t) dt;$$

- costs due to the number of individuals or farms that have been infected during the whole outbreak:

$$\int_0^T A_1 \beta S Y_n dt;$$

where  $A_0, A_1 \in \mathbb{R}_+$  are the balancing weight factors, representing the costs per time unit of isolation efforts and the costs of a single new infection, respectively. Of course,  $A_1$  must be strictly positive for the problem being meaningful (otherwise, we are searching for the control that minimizes its own cost). Hence, it is not restrictive to set  $A = A_0/A_1 \geq 0$ , with  $A$  representing the relative cost per time unit of isolation efforts over the cost of a single new infection.

By setting the costs of isolation per time unit as  $A_0 u(t)$ , we are assuming that the control costs are proportional to control effort instead of to the overall control activities, which are represented by  $u(t) Y_n$ . This assumption fits the scenarios in which the procedures of isolation are forerun by surveillance and testing activities aimed at detecting the target populations, as in the cases of test-and-isolate and test-and-cull control strategies (Bolzoni et al. 2019).

In sum, we are considering the following optimal control (OC) problem, aiming at minimizing simultaneously control efforts and epidemic size:

**(OC)** Find an optimal control function  $u^*(t)$  that minimizes the objective functional

$$J = \int_0^T (A u(t) + \beta S Y_n) dt, \tag{7}$$

with  $T$  given by (4), on the admissible set

$$\Omega = \left\{ u(\cdot) \in L^1(0, T) \mid u(t) \in U = [0, u_{max}] \right\}, \tag{8}$$

subject to (2) with initial conditions (5).

The existence of the solution for the OC problem is guaranteed since the requirements of classical existence theorems are satisfied. More precisely, we refer to the formulation of Filippov theorem and its corresponding corollary given by Agrachev and Sachkov (2004) for proving the following theorem.

**Theorem 1** *There exists an optimal control solution  $u^*(t)$  to the OC problem.*

**Proof** We list the requirements of the corollary of Filippov existence theorem in Agrachev and Sachkov (2004):

1. the control set  $U = [0, u_{max}]$  is compact;
2. the sets  $A_U(\mathbf{x}) = \{\mathbf{f}(\mathbf{x}) + u(t)\mathbf{g}(\mathbf{x}) \mid u(t) \in U\}$ ,  $\mathbf{x} \in \mathfrak{D}$  (with  $\mathfrak{D}$  as in (6)), are convex;
3. there exists a constant  $C$  such that:

$$|\mathbf{f}(\mathbf{x}) + u(t)\mathbf{g}(\mathbf{x})| \leq C(1 + |\mathbf{x}|),$$

with  $\mathbf{x} \in \mathfrak{D}$  and  $u(t) \in U$ .

Condition 1 is trivial. Since  $\mathbf{x} \in \mathfrak{D}$  and  $u(t) \in U$ , the state and the control variables are *a priori* bounded and also Condition 3 is verified.

To check that Condition 2 holds, firstly we take  $\mathbf{x} \in \mathfrak{D}$  and prove that the line connecting two points of  $A_U(\mathbf{x})$  entirely lies in  $A_U(\mathbf{x})$ .

Let us take  $y_1, y_2 \in A_U(\mathbf{x})$ , then there exist  $u_1(t), u_2(t) \in U$  such that  $y_j = \mathbf{f}(\mathbf{x}) + u_j(t)\mathbf{g}(\mathbf{x}), j = 1, 2$ . One easily gets:

$$\alpha y_1 + (1 - \alpha) y_2 = \mathbf{f}(\mathbf{x}) + (\alpha u_1(t) + (1 - \alpha) u_2(t)) \mathbf{g}(\mathbf{x}).$$

Since  $U$  is convex,  $\alpha u_1(t) + (1 - \alpha) u_2(t) \in U, \forall \alpha \in [0, 1]$ , and then

$$\alpha y_1 + (1 - \alpha) y_2 \in A_U(\mathbf{x}), \forall \alpha \in [0, 1].$$

As regards the characterization of the optimal control, according to Pontryagin’s minimum principle (Chachuat 2007; Pontryagin et al. 1962; Schmitendorf 1976), the problem of finding the time-dependent control variable  $u^*(t)$  that minimizes the objective functional (7) is equivalent to the problem of minimizing the Hamiltonian:

$$H(\mathbf{x}, u(t), \boldsymbol{\lambda}) = Au(t) + \beta SY_n + \boldsymbol{\lambda}^T (\mathbf{f}(\mathbf{x}) + u(t)\mathbf{g}(\mathbf{x})), \tag{9}$$

given by the sum of the integrand of (7) and the scalar product of  $\boldsymbol{\lambda}$  and the right-hand side of the state system (2). Here  $\boldsymbol{\lambda} = (\lambda_S, \lambda_{Y_1}, \dots, \lambda_{Y_n})^T$  denotes as usual the *adjoint* variables vector at time  $t$ , which is the solution of the adjoint system

$$\dot{\boldsymbol{\lambda}} = - \frac{\partial H(\mathbf{x}^*, u^*(t), \boldsymbol{\lambda})}{\partial \mathbf{x}}, \tag{10}$$

with the *transversality conditions*

$$\lambda_S(T) = 0, \quad \lambda_{Y_k}(T) = 0, \quad k = 1, \dots, n - 1, \quad \lambda_{Y_n}(T) \text{ free.} \tag{11}$$

The notation with superscript  $*$  is used to denote the solution to the state system (2) corresponding to the optimal control  $u^*(t)$ . The above transversality conditions are consequence of the equality terminal constraint on the state variable  $Y_n$  given by (4). For analytical details see Chachuat (2007), Pontryagin et al. (1962) and Schmitendorf (1976).

Further,  $H$  is constantly equal to zero along the optimal solution:

$$H(\mathbf{x}^*, u^*(t), \boldsymbol{\lambda}) \equiv 0.$$

In sum, the optimal control  $u^*(t)$  that solves the control problem is the time-dependent function that minimizes the Hamiltonian (9) on the admissible set (8):

$$H(\mathbf{x}^*, u^*(t), \boldsymbol{\lambda}) = \min_{u(\cdot) \in \Omega} H(\mathbf{x}^*, u(t), \boldsymbol{\lambda}),$$

where  $(\mathbf{x}^*, u^*(t), \boldsymbol{\lambda})$  must satisfy

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) + u(t)\mathbf{g}(\mathbf{x}), \\ \mathbf{x}(0) &= (S_0, Y_{1,0}, \dots, Y_{n,0})^T, \quad Y_n(T) = \varepsilon, \\ \dot{\boldsymbol{\lambda}} &= -\frac{\partial H(\mathbf{x}, u(t), \boldsymbol{\lambda})}{\partial \mathbf{x}}, \\ \lambda_S(T) &= \lambda_{Y_1}(T) = \dots = \lambda_{Y_{n-1}}(T) = 0, \\ 0 &= H(\mathbf{x}, u(t), \boldsymbol{\lambda}), \end{aligned}$$

and the control horizon  $T$  is defined by (4).

Since  $H$  is linear respect to  $u(t)$ , the value of  $u^*(t)$  is determined by the sign of the so-called *switching function*:

$$\psi(\mathbf{x}, \boldsymbol{\lambda}) = \frac{\partial H}{\partial u} = A + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}).$$

In particular,

$$u^*(t) = \begin{cases} 0 & \text{if } \psi(\mathbf{x}^*, \boldsymbol{\lambda}) > 0 \\ u_s(t) & \text{if } \psi(\mathbf{x}^*, \boldsymbol{\lambda}) = 0 \\ u_{max} & \text{if } \psi(\mathbf{x}^*, \boldsymbol{\lambda}) < 0 \end{cases} \tag{12}$$

with  $u_s(t) \in U$ .

If  $\psi(\mathbf{x}^*, \boldsymbol{\lambda})$  vanishes on a set of isolated points, the corresponding control function is a *bang–bang* control, as well known, and its value at these points is not defined. However, for convenience of numerical simulations, we set it equal to  $u_{max}$ . The isolated time instants, internal to  $[0, T]$ , in which  $\psi$  vanishes are called *switching times*.

Unless otherwise specified, we restrict the use of the superscript  $*$  only to the optimal control function  $u^*(t)$ , in order to simplify the notation.

### 4 Admissible optimal solutions

The main result is that the admissible optimal controls for the OC problem under investigation are bang–bang with at most two switching times.

**Theorem 2** *The optimal solution  $u^*(t)$  to the OC problem is a bang–bang control. When  $u^*(t) \neq 0$ , it is of the form*

$$u^*(t) = \begin{cases} 0 & t \in [0, \tau_1) \\ u_{max} & t \in [\tau_1, \tau_2] \\ 0 & t \in (\tau_2, T] \end{cases} \tag{13}$$

with  $\tau_1, \tau_2 \in [0, T]$ , and  $\tau_1 \leq \tau_2$ .



**Proof** Throughout the proof we omit the superscript \* for the optimal quantities, in order to simplify the notation.

First we compute the Hamiltonian (9) along the optimal solution:

$$\begin{aligned}
 H &= Au(t) + \beta SY_n \left( 1 - \lambda_S + \sum_{k=1}^n \lambda_{Y_k} \right) - (u(t) + \mu) \sum_{k=1}^n \lambda_{Y_k} Y_k + \mu \sum_{k=1}^{n-1} \lambda_{Y_{k+1}} Y_k \\
 &= Au(t) - \sum_{k=1}^n \dot{\lambda}_{Y_k} Y_k \equiv 0,
 \end{aligned}
 \tag{14}$$

and explicitly write the adjoint system (10):

$$\begin{aligned}
 \dot{\lambda}_S &= -\beta Y_n \left( 1 - \lambda_S + \sum_{k=1}^n \lambda_{Y_k} \right), \\
 \dot{\lambda}_{Y_k} &= (u(t) + \mu) \lambda_{Y_k} - \mu \lambda_{Y_{k+1}}, \quad k = 1, \dots, n - 1 \\
 \dot{\lambda}_{Y_n} &= -\beta S \left( 1 - \lambda_S + \sum_{k=1}^n \lambda_{Y_k} \right) + (u(t) + \mu) \lambda_{Y_n}.
 \end{aligned}
 \tag{15}$$

The switching function with its first and second time derivatives read, respectively:

$$\psi = A - \sum_{k=1}^n \lambda_{Y_k} Y_k,
 \tag{16}$$

$$\dot{\psi} = \beta SY_n (1 - \lambda_S),$$

$$\ddot{\psi} = \beta S \dot{Y}_n (1 - \lambda_S) + \beta^2 SY_n^2 \sum_{k=1}^n \lambda_{Y_k}.
 \tag{17}$$

Let us prove that  $\psi$  can vanish only in isolated points. By contradiction, suppose that  $\psi$  vanishes on some interval  $\mathcal{I}$ . Then, also all its time derivatives vanish there; in particular,  $\psi = \dot{\psi} = \ddot{\psi} = 0$  on  $\mathcal{I}$  yields:

$$\sum_{k=1}^n \lambda_{Y_k} Y_k \Big|_{\mathcal{I}} = A,
 \tag{18}$$

$$\lambda_S \Big|_{\mathcal{I}} = 1,
 \tag{19}$$

$$\sum_{k=1}^n \dot{\lambda}_{Y_k} \Big|_{\mathcal{I}} = 0,
 \tag{20}$$

implying

$$\sum_{k=1}^n \dot{\lambda}_{Y_k} \Big|_{\mathcal{I}} = 0.
 \tag{21}$$

By explicitly writing the addenda of (21), one obtains  $\sum_{k=2}^n \lambda_{Y_k} = 0$  on  $\mathcal{I}$ , that, along with (20), gives  $\lambda_{Y_1} = 0$  on  $\mathcal{I}$ . Then, also  $\dot{\lambda}_{Y_1} = 0$  on  $\mathcal{I}$  and, by proceeding in series, we have

$$\lambda_{Y_k}|_{\mathcal{I}} = 0, \quad \forall k = 1, \dots, n \tag{22}$$

(see the corresponding adjoint equations in (15)). Two cases must be considered:

- $A > 0$ . Then, equalities (22) are in contrast with (18).
- $A = 0$ . From (18)–(22) it follows that, once  $\psi$  vanishes, it must remain null, namely  $T \in \mathcal{I}$ . However, the equality (19) is in contrast with the transversality condition  $\lambda_S(T) = 0$ .

In both cases a contradiction arises. As a consequence,  $\psi$  can vanish only in isolated points, namely  $u(t)$  is a bang–bang control: it can assume only the extreme values  $u(t) = 0$  and  $u(t) = u_{max}$ .

Let us now prove that  $u(t)$  changes its value at most two times and, in such a case, the two switches are from  $u(t) = 0$  to  $u(t) = u_{max}$  and again back to  $u(t) = 0$ .

As first step, we consider the value assumed by the quantities involved in the OC problem at the extinction time  $T$ . Taking into account the transversality conditions (11), we have  $H(T) = Au(T) + \beta S(T)Y_n(T) + \lambda_{Y_n}(T)\dot{Y}_n(T) = 0$ . Since it is  $\dot{Y}_n(T) < 0$  by the definition of  $T$ , it follows that

$$\lambda_{Y_n}(T) > 0. \tag{23}$$

From the adjoint equations (15), one yields

$$\dot{\lambda}_S(T) < 0, \tag{24}$$

$$\dot{\lambda}_{Y_k}(T) = 0, \quad k = 1, \dots, n - 2, \tag{25}$$

$$\dot{\lambda}_{Y_{n-1}}(T) < 0. \tag{26}$$

Substituting (25) into the formula of the Hamiltonian (14) provides  $\dot{\lambda}_{Y_{n-1}}(T)Y_{n-1}(T) + \dot{\lambda}_{Y_n}(T)Y_n(T) = Au(T) \geq 0$ , which, in turn, implies  $\dot{\lambda}_{Y_n}(T) > 0$ .

Let us also verify that  $\lambda_{Y_k}(T^-) > 0 > \dot{\lambda}_{Y_k}(T^-), k = 1, \dots, n - 2$ . Take  $k = n - 2$  and assume, by contradiction, that  $\lambda_{Y_{n-2}}(T^-) < 0 < \dot{\lambda}_{Y_{n-2}}(T^-)$ . By handling the corresponding adjoint equation (see (15)), one yields

$$\mu\lambda_{Y_{n-1}}(T^-) = (u(T^-) + \mu)\lambda_{Y_{n-2}}(T^-) - \dot{\lambda}_{Y_{n-2}}(T^-) < 0,$$

that is not possible because  $\lambda_{Y_{n-1}}(T) = 0$  and  $\dot{\lambda}_{Y_{n-1}}(T) < 0$ . Thus, a contradiction arises and it must be  $\lambda_{Y_{n-2}}(T^-) > 0 > \dot{\lambda}_{Y_{n-2}}(T^-)$ . By applying the same argument, one can proceed in series for  $1 \leq k < n - 2$ .

As a further step, we show that  $\lambda_{Y_k} \geq 0$  in  $[0, T], \forall k = 1, \dots, n$ . To this end, let us define  $\bar{t} = \inf \mathcal{A}$ , with

$$\mathcal{A} = \{t \in [0, T] : \lambda_{Y_k} \geq 0 \text{ in } [t, T], \forall k = 1, \dots, n\}$$

and prove that  $\bar{t} = 0$ . From the above considerations it follows that  $T \in \mathcal{A} \neq \emptyset$  and  $\bar{t} < T$ . By contradiction suppose  $\bar{t} > 0$ , namely there exists  $\bar{k} \in \{1, \dots, n\}$  such that  $\lambda_{Y_{\bar{k}}}(\bar{t}) = 0 < \dot{\lambda}_{Y_{\bar{k}}}(\bar{t})$ . Two cases must be considered:

- $\bar{k} < n$ . Then,  $\lambda_{Y_{\bar{k}+1}}(\bar{t}) = -\dot{\lambda}_{Y_{\bar{k}}}(\bar{t})/\mu < 0$  (see (15)), which is in contrast with the definition of  $\bar{t}$ .
- $\bar{k} = n$ . Then,  $\text{sgn}(\dot{\lambda}_S(\bar{t})) = \text{sgn}(\dot{\lambda}_{Y_n}(\bar{t})) > 0$ . Since  $\dot{\lambda}_S(T) < 0$  as given in (24), there exists  $\hat{t} \in \mathcal{A}$ ,  $\bar{t} < \hat{t} < T$ , such that

$$\dot{\lambda}_S(\hat{t}) = 0. \tag{27}$$

By substituting (27) into the expression of  $H$  and rearranging, we obtain

$$H(\hat{t}) = u(\hat{t})\psi(\hat{t}) - \mu \sum_{k=1}^{n-1} \lambda_{Y_{k+1}}(\hat{t})(Y_{k+1}(\hat{t}) - Y_k(\hat{t})) - \mu\lambda_1(\hat{t})Y_1(\hat{t}) < 0,$$

which is contrast with the identity  $H \equiv 0$ .

In both cases a contradiction arises. As a consequence,  $\bar{t} = 0$ , namely

$$\lambda_{Y_k} \geq 0 \text{ in } [0, T], \forall k = 1, \dots, n. \tag{28}$$

Note that we have also proved that

$$\dot{\lambda}_S < 0 \text{ in } [0, T]. \tag{29}$$

As last step, let us assume, by contradiction, that  $u(t)$  admits two consecutive switches, being the first one from  $u_{max}$  to 0 and vice versa the second one. Let  $\tau_{s1}, \tau_{s2} \in (0, T)$ , with  $\tau_{s1} < \tau_{s2}$ , be the corresponding switching times, respectively, from  $u(t) = u_{max}$  to  $u(t) = 0$  and from  $u(t) = 0$  to  $u(t) = u_{max}$ . From (12) it follows that  $\dot{\psi}(\tau_{s1}) > 0 > \dot{\psi}(\tau_{s2})$ . Since  $\dot{\psi}(T) > 0$  (see (11)–(17)),  $\dot{\psi}$  vanishes at least two times in  $(0, T)$ , say them  $t_M \in (\tau_{s1}, \tau_{s2})$  and  $t_m \in (\tau_{s2}, T)$ , that are, respectively, relative maximum and minimum points for  $\psi$ . From (17) it follows that  $\lambda_S(t_M) = 1 = \lambda_S(t_m)$ , which, however, contradicts the condition (29). As a consequence, the presence of three or more switching times is definitively excluded. Also, if  $u(t)$  changes its value two times, it is necessarily from 0 to  $u_{max}$  and again back to 0. This concludes the proof. □

In Theorem 2, the time instants  $\tau_1 \in [0, T]$  and  $\tau_2 \in [\tau_1, T]$  are introduced to extend the concept of switching time. Specifically,  $\tau_1$  [resp.  $\tau_2$ ] is the *starting* [resp. *ending*] *intervention time*, i.e. the first [resp. last] time instant at which the control  $u^*(t)$  assumes the value  $u_{max}$ . The knowledge of such time instants allows to immediately obtain the control time-profile. Indeed, except for the case  $u^*(t) \equiv 0$  when the best strategy is to ‘do nothing’, from Theorem 2 it follows that the following four possible control profiles are allowed:

- (i)  $\tau_1 = 0$  and  $\tau_2 = T$ , i.e.  $u^*(t) \equiv u_{max}$ ,

- (ii)  $\tau_1 > 0$  and  $\tau_2 = T$ , i.e.  $\tau_1$  is the only switching time and  $u^*(t)$  changes its value from 0 to  $u_{max}$  (say, *delayed* control),
- (iii)  $\tau_1 = 0$  and  $\tau_2 < T$ , i.e.  $\tau_2$  is the only switching time and  $u^*(t)$  changes its value from  $u_{max}$  to 0 (say, *reactive* control),
- (iv)  $\tau_1, \tau_2 \in (0, T)$  are two switching times and  $u^*(t)$  changes its value from 0 to  $u_{max}$  and again back to 0.

We also prove that the control profile (i) is the only admissible if one wants to minimize just the epidemic size and not also the isolation efforts, namely  $A = 0$  in the objective functional (7).

**Corollary 3** *If  $A = 0$  in the objective functional (7), then the optimal solution  $u^*(t)$  to the OC problem is constantly equal to  $u_{max}$ .*

**Proof** Let us resume the proof of Theorem 2 and consider the case  $A = 0$ . Then, the switching function reads

$$\psi = - \sum_{k=1}^n \lambda_{Y_k} Y_k.$$

At the extinction time  $T$ ,  $\psi(T) = -\lambda_{Y_n}(T)Y_n(T) < 0$  (see (11)–(23)). From (12), it follows that  $u(T) = u_{max}$ .

By contradiction, let  $\tau_s$  be a switching time for  $u(t)$  from  $u(t) = 0$  to  $u(t) = u_{max}$ . By definition,  $\psi(\tau_s) = - \sum_{k=1}^n \lambda_{Y_k}(\tau_s)Y_k(\tau_s) = 0$ , implying that there exist at least two indexes  $k_1, k_2 \in \{1, \dots, n\}$  such that  $\lambda_{Y_{k_1}}(\tau_s) < 0 < \lambda_{Y_{k_2}}(\tau_s)$ . However, this in contrast with the sign of the adjoint variables (28). As a consequence, it must be  $u(t) \equiv u_{max}$ , which is the assert. □

### 5 Numerical results

As far as the numerical simulations are concerned, the shape of admissible optimal controls allows us to adopt a simple *ad hoc* numerical scheme, which generalizes those implemented by Bolzoni et al. (2017, 2019). Such a scheme is based on the idea of identifying each bang–bang solution  $u^*(t)$  with two real parameters: the starting and ending intervention times. Therefore, the functional  $J(u)$  to be minimized can be seen as the function that links the pair of characteristic intervention times to the objective integral (7) of the OC problem:

$$J : (\tau_1, \tau_2) \rightarrow Au_{max}(\tau_2 - \tau_1) + \int_0^T \beta SY_n dt.$$

An optimal control  $u^*(t)$  will be identified by the pair  $(\tau_1, \tau_2)$  such that  $(\tau_1, \tau_2) = \operatorname{argmin} J$ . The numerical solution will be computed by evaluating the function  $J(\tau_1, \tau_2)$  over a suitable interval and looking for its minimum value. Of course, if the minimum value of  $J$  is greater than the corresponding value in absence of control (i.e.  $\int_0^T \beta SY_n dt$  as predicted by system (2) with  $u(t) \equiv 0$ ), then the method

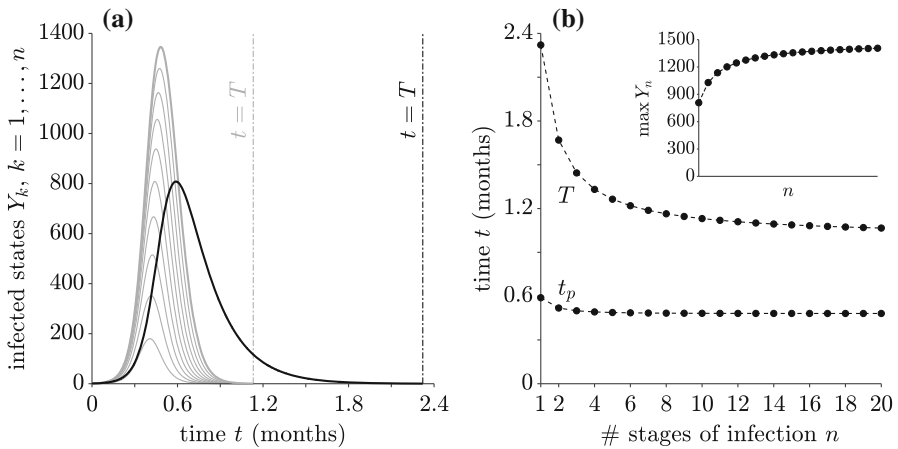
selects for  $u^*(t) \equiv 0$  as the optimal isolation policy. For the numerical integrations of the Cauchy problem (2)–(5) the fourth-order Runge–Kutta scheme with uniform time steps is used. For a detailed discussion of the implemented numerical method see Bolzoni et al. (2017).

## 5.1 Parametrization

Under the assumption of different shapes in the distribution of the infectious period in model (2), we numerically investigate the optimal isolation strategies focusing on highly contagious livestock diseases, such as foot and mouth disease (FMD) and highly-pathogenic avian influenza (HPAI). However, thanks to the generality of the proposed framework, the obtained results can be qualitatively applied to different contexts, like that of non-pharmaceutical interventions (such as self-isolation) in human epidemics. As in Bolzoni et al. (2019), we assume the farm as the unit of observation, since the control measures in livestock diseases are generally implemented at farm level (Probert et al. 2016). Moreover, our choice to work at farm level is strengthened by the finding obtained by Ferguson et al. (2001), who estimated that the infectious period of a farm infected by FMD can be described through an Erlang distribution with  $n \gg 1$ .

The initial conditions for model (2) variables are set to a single infected unit ( $Y_{k,0} = 1, \forall k = 1, \dots, n$ ) in a totally susceptible population ( $S_0 = 2000$ ). By setting  $Y_{k,0} = 1, \forall k = 1, \dots, n$ , we assume that the infection is introduced in the population through a newly infected farm, i.e.  $I_{1,0} = 1$  and  $I_{k,0} = 0, k = 2, \dots, n$ . While this assumption seems obvious when the epidemiological units are represented by farms, other assumptions on the initial conditions can be reasonable when the epidemiological units are represented by individuals (e.g. in human diseases). Thus, we also test the effects on the optimal control of different assumptions on the initial conditions. Specifically, we derive the optimal solutions by varying the stage of infection of the first infected unit (say,  $k_0$ ), which corresponds to assume  $Y_{k,0} = 0, 1 \leq k < k_0$  and  $Y_{k,0} = 1, k_0 \leq k \leq n$ .

We set the infectious period, corresponding to the average time a unit stays infected before recovering, to 1/5 months. Thus, the recovery rate ( $\gamma = \mu/n$ ) in model (2) is set to  $\gamma = 5 \text{ month}^{-1}$ . The chosen value of infectious period falls into the range of those obtained both for FMD and HPAI (Boender et al. 2007; Bouma et al. 2003; Haydon et al. 1997; Rossi et al. 2017; Sharkey et al. 2008). We set the basic reproduction number ( $\mathcal{R}_0$ ), which represents the average number of secondary infections produced by a single infected in a completely susceptible population (Hethcote 2000), to  $\mathcal{R}_0 = 4$  (Bouma et al. 2003; Haydon et al. 1997; Sharkey et al. 2008; Stegeman et al. 2004). Since  $\mathcal{R}_0$  in model (2) can be written as  $\mathcal{R}_0 = \beta S_0/\gamma$  (Hethcote 2000), we set  $\beta = 0.01 \text{ farm}^{-1} \text{ month}^{-1}$ . As in Bolzoni et al. (2019), we set the maximum isolation effort to  $u_{max} = 1 \text{ month}^{-1}$  and  $\varepsilon = 0.5$ . Hansen and Day (2011) pointed out that any value of  $\varepsilon$  smaller than one fits the purpose of the analysis, while Bolzoni et al. (2019) showed in a similar framework that the shape of the optimal control does not significantly depend on the chosen value of  $\varepsilon$ .



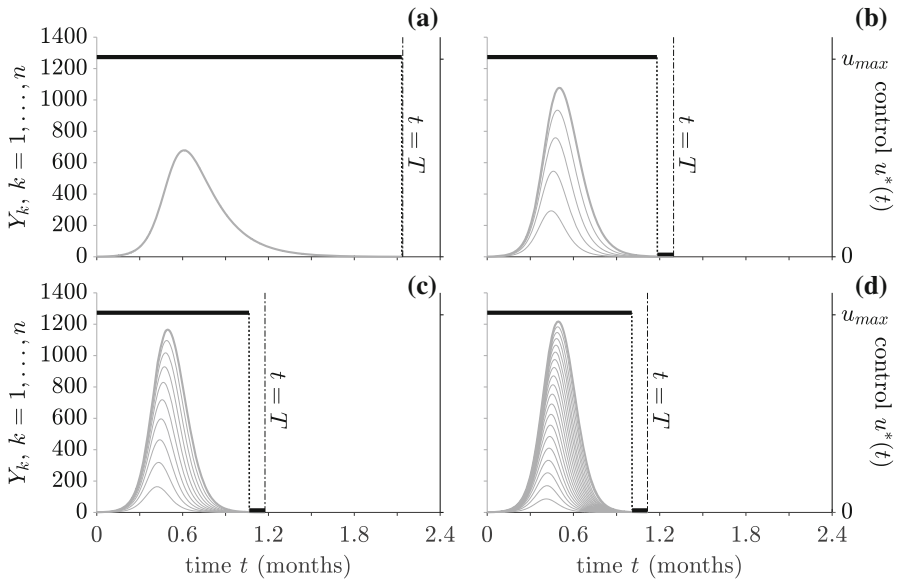
**Fig. 1** Numerical simulations of model (2) with no control ( $u(t) \equiv 0$ ). **a** Infected units in the first  $k$  stages of infection  $Y_k, k = 1, \dots, n$ , by assuming  $n = 1$  (black solid line) and  $n = 10$  (gray solid lines); dash-dotted vertical lines denote the corresponding extinction time  $T$ . **b** Extinction time  $T$  and time of peak  $t_p$  ( $Y_n(t_p) = \max(Y_n)$ ) as functions of  $n \in \{1, \dots, 20\}$ . In box: maximum number of infected units,  $\max(Y_n)$ , as functions of  $n \in \{1, \dots, 20\}$ . Parameters  $Y_{k,0} = 1, \forall k = 1, \dots, n$ ; unspecified parameters as in Sect. 5.1

In our simulations, we also test the sensitivity of the optimal control solutions to different combinations of the number of stages of infection  $n$  and the control parameter  $A$  (which represents the ratio between the costs per time unit of the isolation effort and the costs of a single new infection).

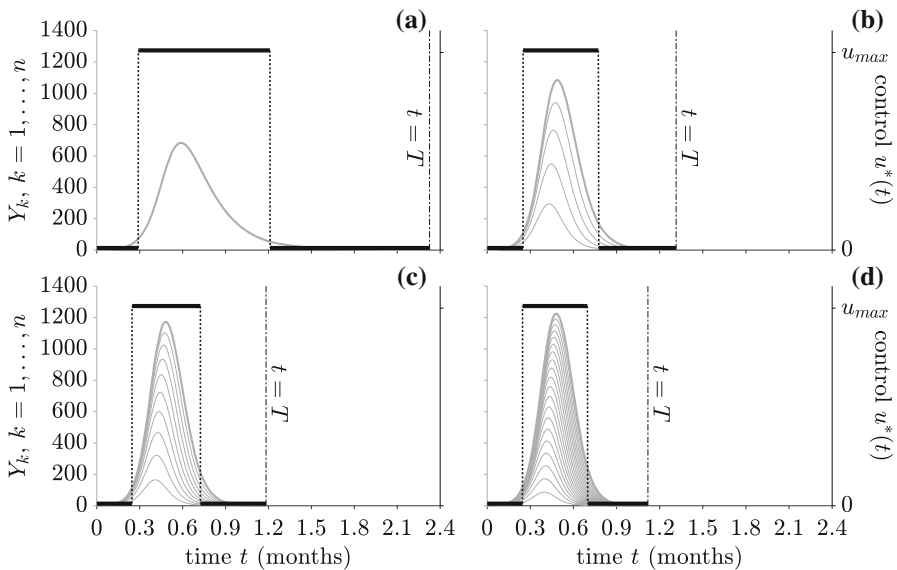
### 5.2 Simulations

Hethcote and Tudor (1980) observed that changing the distribution of the infectious period does not lead to significant differences in the asymptotic behavior of SIR-like epidemic models. However, it can crucially affect the quantitative predictions of the epidemic model on disease burden, epidemic duration, and effectiveness of disease control strategies, as shown by Wearing et al. (2005).

In Fig. 1, we display the numerical simulations of model (2) infected dynamics in the absence of control (i.e.  $u(t) \equiv 0$ ). Figure 1a shows the dynamics of the infected state variables  $Y_k, k = 1, \dots, n$ , as a function of time ( $t$ ), for  $n = 1$  (black curve) and  $n = 10$  (gray curves). (Thick curves represent the total number of infected,  $Y_n$ .) In Fig. 1b, we display the values of the extinction time ( $T$ ), the time at which the epidemic peak occurs ( $t_p = \operatorname{argmax}(Y_n)$ ), and the total number of infected at the peak ( $\max(Y_n)$ , see the box) obtained in the simulations as functions of the number of stages of infection  $n$ . Similarly to previous results obtained with SIR-like epidemic models (Wearing et al. 2005), our numerical analyses on model (2) in the absence of control confirm that the shape of the infectious period distribution quantitatively affects the model predictions. Specifically, an Erlang-distributed model ( $n > 1$ ) predicts a faster initial increase of the infected curve, a larger peak of infection, and a faster extinction time with respect to the exponential-distributed one ( $n = 1$ ). From  $n = 1$  to  $n = 20$ ,



**Fig. 2** Numerical simulations of OC problem. Infected units in the first  $k$  stages of infection  $Y_k, k = 1, \dots, n$  (gray solid lines), and optimal control function  $u^*(t)$  (black solid lines), by assuming  $n = 1$  (a),  $n = 5$  (b),  $n = 10$  (c),  $n = 20$  (d). Dash-dotted vertical lines denote the corresponding extinction time  $T$ . Parameters  $A = 0.1, Y_{k,0} = 1, \forall k = 1, \dots, n$ ; unspecified parameters as in Sect. 5.1



**Fig. 3** Numerical simulations of OC problem. Infected units in the first  $k$  stages of infection  $Y_k, k = 1, \dots, n$  (gray solid lines), and optimal control function  $u^*(t)$  (black solid line), by assuming  $n = 1$  (a),  $n = 5$  (b),  $n = 10$  (c),  $n = 20$  (d). Dash-dotted vertical lines denote the corresponding extinction time  $T$ . Parameters  $A = 10, Y_{k,0} = 1, \forall k = 1, \dots, n$ ; unspecified parameters as in Sect. 5.1

the peak of infection increases by about 74% (808 vs. 1407 infected units), while the disease extinction time decreases by about 54% (2.321 vs. 1.066 months).

In Fig. 2 (resp. Fig. 3), we show the numerical solutions of the OC problem for four values of the number of stages of infection:  $n \in \{1, 5, 10, 20\}$ , by setting the relative cost in the objective functional (7) equal to  $A = 0.1$  [resp.  $A = 10$ ]. Gray curves represent the temporal dynamics of the infected state variables  $Y_k$ ,  $k = 1, \dots, n$ , black curves those of the optimal isolation strategies  $u^*(t)$ . Figures 2 and 3 show that, analogously to that observed in the uncontrolled model, different assumptions on the shape of the infectious period distribution quantitatively affect the optimal control solutions. Specifically, OC models assuming exponentially distributed infectious periods tend to overestimate the costs associated to control (we estimate an increase of about 50% in the case of  $n = 1$  with respect to  $n = 10$ , see Figs. 2, 3).

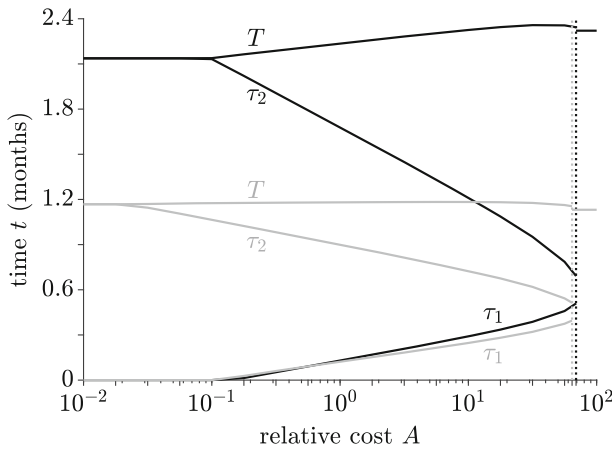
In addition to the quantitative differences described above, our numerical analyses show that different assumptions on the distribution of the infectious period can affect qualitatively the solutions of the OC problem. In Fig. 4, we display the values of the starting ( $\tau_1$ ) and ending ( $\tau_2$ ) intervention times and the extinction time ( $T$ ) for the optimal solutions as functions of the relative cost  $A$ , by setting  $n = 1$  (black curves) and  $n = 10$  (gray curves). From Fig. 4, we notice that for a wide range of  $A$  values (approximately from  $A = 2 \cdot 10^{-2}$  to  $A = 10^{-1}$ ) the optimal control problem selects for a constant control  $u^*(t) \equiv u_{max}$  (i.e.  $\tau_1 = 0$  and  $\tau_2 = T$ ) when  $n = 1$ , while the optimal control admits a one-switch reactive solution (i.e.  $\tau_1 = 0$  and  $\tau_2 < T$ ) when  $n = 10$ . Furthermore, for more limited ranges of  $A$ , we show that the optimal control problem can: (i) admit a one-switch solution when  $n = 1$ , while it admits a two-switches solution when  $n = 10$  (see  $A \approx 10^{-1}$  in Fig. 4); (ii) admit a two-switches solution when  $n = 1$ , while it selects for  $u^*(t) \equiv 0$  when  $n = 10$  (see the black and gray dash-dotted lines in Fig. 4). Interestingly, for both  $n = 1$  and  $n = 10$ , the starting and ending intervention times are monotone (increasing and decreasing, respectively) functions of  $A$ , while the disease extinction time  $T$  is increasing for a wide range of  $A$  and then starts to decrease when  $A$  approaching the value above which  $u^*(t) \equiv 0$ .

Finally, we also consider variations with respect to the initial conditions of the OC problem. Specifically, we set the relative cost  $A = 10$ , the number of stages of infection  $n = 10$ , and vary the stage entered by the first infected unit ( $k_0$ ). Namely,  $Y_{k,0} = 0$ ,  $1 \leq k < k_0$ ,  $Y_{k,0} = 1$ ,  $k_0 \leq k \leq n$ . The values of the starting ( $\tau_1$ ) and ending ( $\tau_2$ ) intervention times and the extinction time ( $T$ ) for the optimal solutions are displayed in Fig. 5. We notice that different assumptions on the initial class  $k_0$  of the first infected do not substantially change the shape of the optimal solution. Both the intervention times and the extinction time slightly increase with  $k_0$ : from  $k_0 = 1$  to  $k_0 = 10$  the growth is approximately 0.065 months.

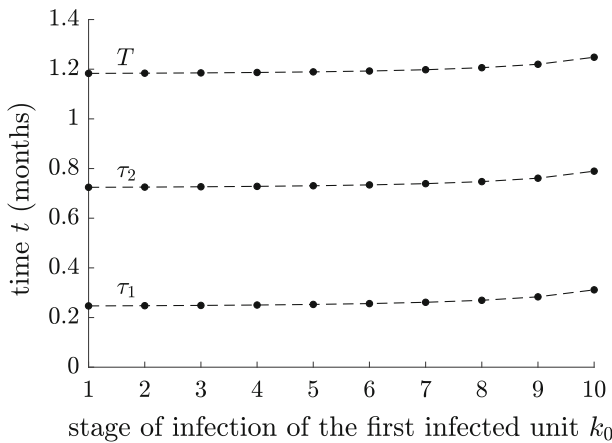
## 6 Discussion and conclusion

In this work, we investigate optimal isolation strategies in an SIR model with Erlang distribution of the infectious period. We adopt the *method of stages* (Cox and Miller 1965), by splitting the infected compartment in  $n$  fictitious stages in series ( $n \in \mathbb{N}_+$ ); infected units at any stage are isolated with time-variant bounded rate  $u(t)$  (here, units





**Fig. 4** Numerical simulations of OC problem. Starting intervention time  $\tau_1$ , ending intervention time  $\tau_2$  and extinction time  $T$  as functions of  $A$  in semi-logarithmic scale. Black solid lines: simulations by assuming  $n = 1$ . Gray solid lines: simulations by assuming  $n = 10$ . Dotted vertical lines denote the  $A$  values for which  $u^*(t)$  passes from a two-switches control to  $u^*(t) \equiv 0$ . Parameters  $Y_{k,0} = 1, \forall k = 1, \dots, n$ ; unspecified parameters as in Sect. 5.1



**Fig. 5** Numerical simulations of OC problem. Starting intervention time  $\tau_1$ , ending intervention time  $\tau_2$  and extinction time  $T$  as functions of  $k_0$ , the stage entered by the first infected unit, with  $k_0 \in \{1, \dots, 10\}$ . Parameters  $n = 10, A = 10, Y_{k,0} = 0, 1 \leq k < k_0, Y_{k,0} = 1, k_0 \leq k \leq n$ ; unspecified parameters as in Sect. 5.1

represent individuals or farms). The aim of the optimal control (OC) problem is to jointly minimize the costs associated to the control efforts and the epidemic size.

By characterizing optimal control shapes according to Pontryagin’s minimum principle (Pontryagin et al. 1962), we prove that, for both  $n = 1$  (exponential distribution of the infectious period) and  $n > 1$  (non-exponential distribution of the infectious period), the OC problem admits only bang–bang solutions with at most two switches. When the optimal control  $u^*(t)$  is constant for the whole time horizon  $[0, T]$ , it can be

either null or equal to its maximum value  $u_{max}$ . When only one switch is admitted, the OC problem can select for both reactive isolation (characterized by  $u^*(0) = u_{max}$  and  $u^*(T) = 0$ ) and delayed isolation (characterized by  $u^*(0) = 0$  and  $u^*(T) = u_{max}$ ). When the optimal solution admits two switches, then the OC problem selects for controls where isolation is implemented within a temporal window between times  $\tau_1$  and  $\tau_2$  (with  $\tau_1 < \tau_2$  and  $\tau_1, \tau_2 \in (0, T)$ ).

Numerical analyses show that increasing in the number of stages of infection  $n$  is positively associated to faster, but with higher peaks, epidemic outbreaks both in absence and in presence of control. When optimal isolation strategies are implemented, even if, *a priori*, the spectrum of admissible control profiles is the same for any  $n$ , the optimal solutions vary both quantitatively and qualitatively. Specifically, the OC problem in the case  $n = 1$  can select for solutions with a different number of switches (if any) with respect to the case  $n > 1$ . In addition, we observe that the control costs associated to the epidemic significantly decrease with  $n$ , as a consequence of the reduced epidemic lengths.

We also test the sensitivity of the optimal control solutions to the weight parameter  $A$  in the objective functional, that is the ratio between the costs per time unit of the isolation effort and the costs of a single new infection. From numerical simulations (see Fig. 4), we observe that, by increasing  $A$ ,  $u^*(t)$  passes from a control constantly equal to  $u_{max}$  (that is the unique admissible solution in the special case  $A = 0$ , see Corollary 3), to a reactive control up to a two-switches control. Lastly, when  $A$  is sufficiently large, the best strategy is to ‘do nothing’ (i.e.  $u^*(t) \equiv 0$ ). While it seems unrealistic to expect that stakeholders and policy-makers can express their preferences on specific values of parameter  $A$ , the outputs of the optimization problem can identify the parameter regions where the benefits of a control campaign outweigh the costs and those where the costs outweigh the benefits, thus helping stakeholders and policy-makers in their evaluations.

To the best of our knowledge, our work is the first attempt to use an SIR model with non-exponential distribution of the infectious period in an optimal control framework. As is widely known, for many infectious diseases the infectious period distributions are not exponential. Hence, models with exponential distributions may lead to biased outcomes in both the understanding of behaviors of disease dynamics and the evaluations of disease control programs. The Erlang distribution proved to be a valuable compromise between realism and analytical tractability, and was largely used to estimate infectious period distributions by fitting with empirical data (see, just to mention few papers, Eichner and Dietz 2003; Ferguson et al. 2001; Wearing et al. 2005).

This work serves as a proof-of-principle verification of the impact on optimal isolation strategies of different assumptions about the distribution of the infectious period. Our analytical and numerical results stress the conclusion that simple epidemic models, such as the ones assuming exponential distribution on the infectious period or homogeneous mixing in the host population, can not provide reliable solutions to real-world cases of disease control from the quantitative point of view. However, the theoretical findings provided by these models can be used to inform more complex simulation models developed for specific epidemiological scenarios where more realistic descriptions of the biological, ecological and epidemiological processes are included.

We investigate the specific control problem where the costs of control are proportional to isolation efforts instead of to the overall number of units isolated. This assumption fits the scenarios where the costs linked to the activities of surveillance and testing aimed at detecting the units to isolate (which depend on the effort and not on the target population) are larger than the costs of isolating the infected units. We find that the optimal solutions obtained under this assumption substantially differ from those obtained for optimal control problems where the control costs are proportional to the number of infected units (Morton and Wickwire 1974; Wickwire 1975).

Roughly speaking, there are two alternative formulations for the cost of the control campaign in the objective functional: (i) control costs proportional to the overall control activities, that is the control rate multiplied by the units to whom it is addressed (e.g., in our notations,  $u(t)S$  in case of vaccination,  $u(t)I$  in case of isolation or treatment, see e.g. Behncke 2000; Morton and Wickwire 1974; Wickwire 1975); (ii) control costs proportional to control efforts (that is the control rate  $u(t)$  as considered here, see e.g. Behncke 2000; Bolzoni et al. 2014; Zhou et al. 2014). The assumption (i) was largely adopted in OC applications to the SIR model under different control strategies. In such a case, the problems only support the trivial solutions of applying the maximal effort for the entire epidemic ( $u^*(t) \equiv u_{max}$ ) (Wickwire 1975) or having at most one switch from  $u^*(t) = u_{max}$  to  $u^*(t) = 0$  (reactive control) (Behncke 2000; Morton and Wickwire 1974). These are also the only admissible solutions for optimal vaccination policies when the assumption (ii) is adopted (Zhou et al. 2014). Conversely, we prove here that optimal isolation policies under the assumption (ii) are not necessarily active since the initial time of the epidemic. Namely, the optimal starting intervention time could be postponed, which yields delayed controls or even two-switches controls (depending on if the ending intervention time coincides or not with the disease extinction time).

The biological explanation for the optimality of delayed [resp. two-switches] control when the costs are proportional to the control efforts relies on the remark that, when the maximum effort admissible (i.e.  $u_{max}$ ) is not large enough to significantly flatten the epidemic curve, employing resources to find infected units at the beginning [resp. the beginning and the end] of the epidemic may not be efficient, since several susceptible units would be screened/tested in the face of few infected ones. The described scenario is similar to the case where the disease control is applied only when the number of infected (or new infections) exceeds a given threshold, which can be mathematically investigated through sliding models (Bolzoni et al. 2020; Qin et al. 2016).

**Acknowledgements** We are thankful to the anonymous reviewer for the useful comments. The work of LB was supported by the Italian Ministry of Health (Grant IZSLER PRC2014003). The work of MG was supported by the Italian Ministry of Education, University and Research (Prin 2017YBKNCE). Support by INdAM-GNFM and by University of Parma is also gratefully acknowledged by RDM and MG. This research benefited from the HPC (High Performance Computing) facility of the University of Parma.

**Funding** The work of LB was supported by the Italian Ministry of Health (Grant IZSLER PRC2014003). The work of MG was supported by the Italian Ministry of Education, University and Research (Prin 2017YBKNCE).

## Declaration

**Conflict of interest** The authors declare they have no competing interests.

**Availability of data and material** Not applicable.

**Code availability** Not applicable.

## References

- Agrachev AA, Sachkov YL (2004) Control theory from the geometric viewpoint, vol 87. Encyclopaedia of mathematical sciences. Springer, Berlin
- Anderson RM, May RM (1992) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford
- Bailey NT (1954) A statistical method of estimating the periods of incubation and infection of an infectious disease. *Nature* 174(4420):139–140
- Behncke H (2000) Optimal control of deterministic epidemics. *Optim Control Appl Methods* 21:269–285
- Boender GJ, Hagenaars TJ, Bouma A, Nodelijk G, Elbers ARW, de Jong MCM, van Boven M (2007) Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Comput Biol* 3:e71
- Bolzoni L, Tesson V, Groppi M, De Leo GA (2014) React or wait: which optimal culling strategy to control infectious diseases in wildlife. *J Math Biol* 69:1001–1025
- Bolzoni L, Bonacini E, Soresina C, Groppi M (2017) Time-optimal control strategies in SIR epidemic models. *Math Biosci* 292:86–96
- Bolzoni L, Bonacini E, Della Marca R, Groppi M (2019) Optimal control of epidemic size and duration with limited resources. *Math Biosci* 315:108232
- Bolzoni L, Della Marca R, Groppi M, Gragnani A (2020) Dynamics of a metapopulation epidemic model with localized culling. *Discrete Contin Dyn Syst B* 25:2307–2330
- Bouma A, Elbers ARW, Dekker A, de Koeijer A, Bartels C, Vellema P, van der Wal P, van Rooij EMA, Pluimers FH, de Jong MCM (2003) The foot-and-mouth disease epidemic in The Netherlands in 2001. *Prev Vet Med* 57:155–166
- Chachuat B (2007) Nonlinear and dynamic optimization: from theory to practice. Technical report, Automatic Control Laboratory, EPFL, Switzerland
- Cox DR, Miller HD (1965) The theory of stochastic processes. Chapman and Hall, London
- Cunniffe NJ, Stutt ROJH, van den Bosch F, Gilligan CA (2012) Time-dependent infectivity and flexible latent and infectious periods in compartmental models of plant disease. *Phytopathology* 102(4):365–380
- Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife—threats to biodiversity and human health. *Science* 287:443–449
- DEFRA/DCMS (2002) Economic cost of foot and mouth disease in the UK. Technical report, Joint Working Paper of the Department for the Environment, Food and Rural Affairs and the Department of Culture, Media and Sport, UK
- Eichner M, Dietz K (2003) Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am J Epidemiol* 158:110–117
- Evans NHM, Peacock B (1993) Statistical distributions, 2nd edn. Wiley, New York
- Feng Z, Xu D, Zhao H (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bull Math Biol* 69:1511–1536
- Feng Z, Zheng Y, Hernandez-Ceron N, Zhao H, Glasser JW, Hill AN (2016) Mathematical models of Ebola—consequences of underlying assumptions. *Math Biosci* 277:89–107
- Ferguson NM, Donnelly CA, Anderson RM (2001) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* 292:1155–1160
- Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A (2020) Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc Natl Acad Sci USA* 117:10484–10491
- Gough K (1977) The estimation of latent and infectious periods. *Biometrika* 64(3):559–565

- Gumel AB, Ruan SG, Day T, Watmough J, Brauer F, van den Driessche P, Gabrielson D, Bowman C, Alexander ME, Ardal S, Wu JH, Sahai B (2004) Modelling strategies for controlling SARS outbreaks. *Proc R Soc B* 271:2223–2232
- Hansen E, Day T (2011) Optimal control of epidemics with limited resources. *J Math Biol* 62:423–451
- Haydon DT, Woolhouse MEJ, Kitching RP (1997) An analysis of foot-and-mouth-disease epidemics in the UK. *Math Med Biol J IMA* 14:1–9
- Hethcote HW (2000) The mathematics of infectious diseases. *SIAM Rev* 42:599–653
- Hethcote HW, Tudor DW (1980) Integral equation models for endemic infectious diseases. *J Math Biol* 9(1):37–47
- Keeling MJ, Grenfell BT (1997) Disease extinction and community size: modeling the persistence of measles. *Science* 275(5296):65–67
- Krylova O, Earn DJD (2013) Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J R Soc Interface* 10:20130098
- Lloyd AL (2001a) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc R Soc B* 268(1470):985–993
- Lloyd AL (2001b) Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* 60(1):59–71
- Ma J, Earn DJD (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull Math Biol* 68(3):679–702
- McCallum H, Dobson A (1995) Detecting disease and parasite threats to endangered species and ecosystems. *Trends Ecol Evol* 10:190–194
- Morton R, Wickwire K (1974) On the optimal control of a deterministic epidemic. *Adv Appl Probab* 6:622–635
- Pontryagin L, Boltyanskii V, Gamkrelidze R, Mishchenko E (1962) The mathematical theory of optimal processes. International series of monographs in pure and applied mathematics. Interscience Publishers, Los Angeles
- Probert WJM, Shea K, Fonnesebeck CJ, Runge MC, Carpenter TE, Dürr S, Garner MG, Harvey N, Stevenson MA, Webb CT, Werkman M, Tildesley MJ, Ferrari MJ (2016) Decision-making for foot-and-mouth disease control: objectives matter. *Epidemics* 15:10–19
- Qin W, Tang S, Xiang C, Yang Y (2016) Effects of limited medical resource on a Filippov infectious disease model induced by selection pressure. *Appl Math Comput* 283:339–354
- Rossi G, Smith RL, Pongolini S, Bolzoni L (2017) Modelling farm-to-farm disease transmission through personnel movements: from visits to contacts, and back. *Sci Rep* 7:2375
- Schmitendorf WE (1976) Pontryagin's principle for problems with isoperimetric constraints and for problems with inequality terminal constraints. *J Optim Theory Appl* 18(4):561–567
- Sharkey KJ, Bowers RG, Morgan KL, Robinson SE, Christley RM (2008) Epidemiological consequences of an incursion of highly pathogenic H5N1 avian influenza into the British poultry flock. *Proc R Soc B* 275:19–28
- Sherborne N, Blyuss KB, Kiss IZ (2015) Dynamics of multi-stage infections on networks. *Bull Math Biol* 77(10):1909–1933
- Simpson REH (1952) Infectiousness of communicable diseases in the household (measles, chickenpox, and mumps). *Lancet* 260:549–554
- Stegeman A, Bouma A, Elbers ARW, de Jong MCM, Nodelijk G, de Klerk F, Koch G, van Boven M (2004) Avian influenza A virus (H7N7) epidemic in The Netherlands in 2003: course of the epidemic and effectiveness of control measures. *J Infect Dis* 190:2088–2095
- Wang X, Shi Y, Feng Z (2017) Evaluations of interventions using mathematical models with exponential and non-exponential distributions for disease stages: the case of Ebola. *Bull Math Biol* 79:2149–2173
- Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLoS Med* 2:e174
- Wickwire K (1975) Optimal isolation policies for deterministic and stochastic epidemics. *Math Biosci* 26:325–346
- Zhou Y, Yang K, Zhou K, Liang Y (2014) Optimal vaccination policies for an SIR model with limited resources. *Acta Biotheor* 62(2):171–181