CRM

CENTRE DE RECERCA MATEMÀTICA

Preprint núm. 1219
February 2016

Optimal controls for an SEIR model of endemically persisting infectious diseases with nonlinear incidence rate

A. Korobeinikov, E.V. Grigorieva, E.N. Khailov

CRM Preprint Series number 1219

OPTIMAL CONTROLS FOR AN SEIR MODEL OF ENDEMICALLY PERSISTING INFECTIOUS DISEASES WITH NONLINEAR INCIDENCE RATE

ANDREI KOROBEINIKOV, ELLINA V. GRIGORIEVA, AND EVGENII N. KHAILOV

Abstract. The objective of this paper is to explore possible impacts of nonlinearity of the functional responses and the number of compartments in an infection disease model on principal qualitative properties of the optimal controls for the model. In order to address this issue, we consider optimal controls for a SEIR model of an endemically persisting infectious disease. In this model, we assume that the incidence rate is given by an unspecified nonlinear function constrained by a few biologically motivated conditions. For this model, we considered five controls (which comprise all controls possible for this model) with a possibility of simultaneous acting of all of these. For this model, we found principal qualitative types of all possible controls. A comparison with a similar SIR model is provided.

1. Introduction

The need of optimizing controls in problems such as pest control, antiviral and anticancer therapies and processes in bioengineering frequently arise in biological and biomedical applications. Control of infectious diseases is probably the most important of all these problems. Naturally, such a diversity of control problems attracted attention of mathematicians working in the optimal control theory, and for the last two decades a substantial progress was made in application of this theory to control problems in biology and medicine. The most notable progress was done for the problems where the objective functionals are formulated as an explicit function of the control. The most commonly used form of such functionals includes a sum of the weighted squares of the controls [1, 2, 5, 6, 9, 10, 21, 23, 24, 32, 35]. For such objective functionals, it is possible, by the virtue of the Pontryagin maximum principle, to obtain the optimal controls as explicit functions of the phase and the adjoint variables. As a result, the optimal control problem can be reformulated as a boundary value problem, which then is solved numerically. The optimal controls provided by the Pontryagin maximum principle are usually continuous functions of the phase and adjoint variables, and hence the

2010 *Mathematics Subject Classification.* 49J15, 58E25, 92D30.

Key words and phrases. SEIR model, nonlinear incidence, infectious disease control, nonlinear control system, Pontryagin maximum principle, non-autonomous quadratic differential system, generalized Rolle's theorem.

resulting boundary value problem is well posed and very convenient for numerical solving.

While the described approach is mathematically convenient and transparent, it also has a number of drawbacks. The most obvious and probably most important of these is that in the majority of biomedical applications the actual forms of the dependence of the objective functionals on the controls are uncertain. The most commonly used form of this dependence, namely the sum of weighted squares of the controls, has its origin in engineering applications, where it has a clear meaning of the energy spent on the control action. Hence the motivation for such type of objective functionals in engineering is minimizing this energy. In biomedical applications, this type of the objective functionals also implies minimizing of certain costs or side-effects. However, in the biological and biomedical applications the financial costs of controls most likely grow linearly with the magnitude of the corresponding control action (with a possible deviation from the linear growth in both directions). The same is most likely true for possible non-financial side-effects of large-scale epidemiological actions, such as massive vaccination. For antiviral or anticancer therapies, possible side-effects are most likely nonlinearly increasing with an increase of a dose. However, in this case an actual form of the increment depends on patient's tolerance, which is highly individual and varies for patients as well as for therapies. Besides, the optimal controls are extremely sensitive to a specific form of the dependence of an objective functional on the controls. For the medical and biological problems, this sensitivity, combined with the above mentioned uncertainty of this dependence, leads to a controversy.

For biological and medical problems, the relative values of the costs and side effects of control policy on one hand, and the outcomes achieved by the policy on another hand, and hence the relative weights of the controls and the outcomes in the objective functionals, are also debatable. In a real-life situation, the costs of a massive vaccination or a quarantine are usually considerably smaller than the potential loses that can be inflicted by an epidemics or an epizootic. The financial costs and side-effects of a therapy are also usually small compared with the expected consequences of an untreated disease. Besides, for a real-life control policy, the magnitudes of controls are usually restricted by external pre-determined factors such as the allocated funding or the health system capacities. For a therapy, the maximal doses of drugs or other treatments are also pre-determined.

Furthermore, the above mentioned approach does not guaranty the optimality of the solution obtained. Instead, it ensures only necessary conditions of the optimality and local extremum of the functional. Finally, for the functionals dependent on the controls, the boundary value problems obtained by the virtue of the Pontryagin maximum principle is usually solved numerically. The use of a numerical procedure implies a specific parametrization of all functional responses, making any generalization or extension to other parametrization hardly possible. There are also a number of other issues of mathematical nature which are related to a form of the objective functional. A comprehensive discussion of mathematical aspects of this problem can be found in recently published book by H. Schättler and U. Ledzewicz [39].

The above mentioned considerations, combined with the mentioned sensitivity of the optimal controls to a particular form of the functions, motivate the formulation of optimal control problems where the objective functionals are independent of the controls and the controls are bounded. However, for such formulations an explicit form of the optimal controls cannot be directly found, and the problem itself is usually more difficult to handle. Thus, singular regimes are common in this sort of problems. Despite a significant advance in dealing with the singularities, in particular for the biomedical problems, which was recently done by H. Schättler and U. Ledzewicz (see [33, 38, 39] and bibliography therein), detecting the singular regimes and establishing their optimality remains to be a highly nontrivial mathematical problem.

Biological systems are extremely complex and their behavior is influenced by a very large number of factors. The majority of these factors are neither studied, nor even understood with a sufficient degree of certainty. Under these circumstances, one hardly can expect that the available information on a bio-system is, or will be, sufficient. Among other complications, this implies that any (mathematical or of any other kind) model of a biological process involves a significant number of assumptions and hypotheses, many of which are implicitly assumed. This, in turn, leads to a question about the reliability and relevance of the results obtained employing these models formulated with specific form of functional responses and parametrization. This problem was realized fairly early. One way of dealing with this problem was suggested in 1930s by A.N. Kolmogorov, who suggested to "reverse" the problem [25]: instead of studying properties of a biological system using a specifically defined parametrization, Kolmogorov suggested to formulate a model where the functional responses are given by unspecified functions and then establish the properties (such as positivity, monotonicity or concavity/convexity, etc.) which these functional responses should have in order to ensure that a considered system possessing certain properties.

This concept was successfully used in the global analysis, in particularly in combination with the direct Lyapunov method [26, 27, 28, 29, 30, 31]. It is obvious, however, that this concept hardly can be employed in a combination with numerical analysis, as the latter requires a specific parametrization of functional responses. This makes the use of Kolmogorov's concept in combination with the traditionally formulated optimal control problems with the control-dependent objective functionals rather problematic.

Nevertheless, the problem was recognized in optimal control theory as well. Horst Behncke was probably the first who made use of the Kolmogorov's concept in combination with an optimal control problem in epidemic models [3]. He considered the classic SIR Kermack-McKendrick model and SEIR model for an isolated epidemic (that is, the demographic processes were neglected) assuming that the incidence rate (which is believed to be the most important of all functional responses in epidemic models) was given by an unspecified function constrained by a few conditions. For these models, Behncke considered three controls, namely vaccination (or removal) of the susceptible individuals, isolation of the infectious individuals and education (that is, indirect measures leading to a reduction of the incidence rate), which apparently comprise all controls possible for such a simple model. Considering arguments as above, that is taking into consideration the disproportionality of the costs and financial losses inflicted by a full-scale epidemic compared with the costs of the implementation of a control policy, Behncke employed objective functionals independent of the controls. In this particular case he considered the minimization of the cumulative number of the infected individuals for a given time interval as the objective. For this objective and assuming that the controls are bounded and that only one of the controls is employed in every case (that is, there is no interference of the controls), he found qualitative forms of the optimal controls for each of these three controls.

The control problem considered in [3] was to large extent simplified by the simplicity of the system dynamics and the phase space structure: it is known that the phase trajectories of this system form a one-parametric family of arches in the phase space (and hence the system has a first integral). The starting and ending points of these arches belong to the coordinate axis where the infection level is equal to 0, which is composed of fixed points. Along these trajectories, the number (or concentration) of the susceptible individuals monotonically decreases, whereas the level of infection (the number, or concentration of the infected individuals) initially grows and then, after a critical point is reached, strictly decreases to 0. An SIR (or an SEIR) model where demographic processes are incorporated has a considerably more complex dynamics: it the simplest case of an SIR model with the birth and death of the individuals, the system has a positive fixed point (a node or a focus), and the phase trajectories wind on this point. The analysis of the optimal controls even for this simplest case presents a considerable challenge. The optimal controls for an SIR model with demographic processes and with the incidence rate was given by an unspecified function were considered in [19]. In this paper, four controls, which, in contrast to [3], were assumed acting simultaneously, were considered, and their principal qualitative types were found.

The principal objective of this paper is to address a question whether and how the number of compartments of a model affects the optimal controls. The practical importance of this issue is clear: any mathematical model comprises a number of hypothesis and assumptions, and division of a population into classes or compartments is one of the most important of these. Obviously, all models of a process must represent, within certain degree of accuracy, the underlying biology of the system. Therefore, if the corresponding optimal controls exhibit sensitivity to a particular scheme of such a division, then employing these optimal controls in real-life practice is unbearable.

To address this issue, in this paper we consider optimal controls for a SEIR model of an endemically persisting infection, which, compared to a SIR model, has an extra compartment of the "exposed" individuals. As in [3] and [19], in this model the incidence rate is assumed to be given by an unspecified function constrained by a few biologically feasible conditions (such as positivity and monotonicity). The model incorporates up to five bounded controls, which cover all possible for this model controls and which are assumed to be able to act simultaneously. To analyze the optimal controls for this model, the authors use an approach that was initially developed to study the optimal control for a biochemical process [11, 12, 13, 17], and later extended to controls for endemically persisting infectious diseases (such as childhood infection, e.g., measles) [15, 19] and antiretroviral (HIV) therapy [14, 16]. A major feature of this approach is that the objective functional is assumed to be independent of the (bounded) controls. For such problems, the optimal controls are likely to be of the bang-bang type. (This hypothesis has to be proved in each particular case.) For the bang-bang controls, a number of switchings for each control can be analytically estimated from above and then the values of controls on each of the subintervals can be found.

For a real-life problem, having the number of switching found, one can immediately reduce the optimal control problem to a considerably simpler problem of the finite-dimensional optimization. Then the exact moments of switching can be found numerically using any of the available solvers. A crucial for our objectives feature of this approach is that the estimations for the number of switchings and the values of the controls on the subintervals can be found analytically. This implies that there is a significant flexibility with respect to particular forms of the functional responses of the model and hence a possibility to use this approach in combination with the Kilmogorov's "reversion of a problem" concept.

The paper is structured as following: in next Section we formulate the problem. In Section 3, using the Pontryagin maximum principle, we find differential equations for the adjoint variables and switching functions and show that the optimal controls are bang-bang controls. Section 4 contains the main mathematical results; in this Section we establish estimations for the numbers of switchings of the optimal controls. Having the estimations, in Section 5 we describe the corresponding optimal controls. The last Section contains some discussion of the results. In Appendix A we show that for objective functionals depending on the controls the optimal controls depend on a particular form of non-linearity of functional responses in the model. In Appendix B we show that one of the constraints on the model parameters, which we imposed for the sake of mathematical convenience only, can be omitted. We show how the corresponding proofs can be done in the case when this constraint is not valid.

2. Model and problem formulation

The spread of an infectious disease with a latency in a population can be described by a SEIR epidemic model. In this paper, we consider the following SEIR epidemic model with a nonlinear incidence rate:

$$\begin{cases} \dot{S}(t) = \mu - f(S(t), I(t)) - \mu S(t), \\ \dot{E}(t) = f(S(t), I(t)) - (\theta + \mu)E(t), \\ \dot{I}(t) = \theta E(t) - (\sigma + \mu)I(t), \\ \dot{R}(t) = \sigma I(t) - \mu R(t). \end{cases} \tag{2.1}$$

In this model, $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are, respectively, the fractions of susceptible, exposed (in the latent state), infectious and removed (recovered and immune) individuals; that is

$$S(t) + E(t) + I(t) + R(t) = 1. \tag{2.2}$$

We assume that the population size is constant. The constant population size assumption is typical in mathematical epidemiology and is based on the hypothesis that the time scale of the epidemic process is considerably faster than that of demographic one. This assumption and equality (2.2) enable us to omit the fourth equation and reduce the system to three equations for $S(t)$, $E(t)$ and $I(t)$.

The model assumes that births and natural (that is, not associated with the infection) deaths occur at a per capita rate μ. (The birth and death rates equality is just a matter of convenience and is not of principle importance for further analysis.) All the newborns are susceptible and enter the susceptible compartment S. The susceptible individuals can be infected by the infectious with rate $f(S, I)$. After an instance of infection, the individuals immediately enter the exposed compartment E. The exposed individuals (those in the latent state) are not infectious and eventually move to the infectious compartment I at per capita rate θ (that is, an average duration of the latency is $1/\theta$). The infectious individuals can infect the susceptibles and are removed (recover or die of the infection) at per capita rate σ.

In this paper we assume that the rate of infection is given by function $f(S, I)$ which satisfies the following set of assumptions:

- (A1) $f(S, I)$ is a continuous differentiable function for all $S, I \geq 0$;
- (A2) $f(S, I) > 0$ holds for all $S, I > 0$, and $f(0, I) = f(S, 0) = 0$ for all $S, I \geq 0$;
- (A3) partial derivatives $f'_S(S, I)$, $f'_I(S, I)$ are positive for all $S, I > 0$ and non-negative for all $S \geq 0, I = 0$ and $S = 0, I \geq 0$;
- (A4) there are constants: $M_S = \max_{(S,I)\in\Omega} f'_S(S, I)$, $M_I = \max_{(S,I)\in\Omega} f'_I(S, I)$, where $\Omega = \{(S, I) : S \geq 0,\ I \geq 0,\ S + I \leq 1\}$;
- (A5) the inequality $4M_S^2 - 3M_I^2 + 1.5\mu^2 + 3(\theta + \mu)^2 + 0.5(\sigma + \mu)^2 - \theta^2 > 0$ holds.

It is noteworthy that these assumptions are considerably less restrictive than those that are usually imposed on the incidence rate. Thus, an usual set of assumptions for incidence rate $f(S,I)$ includes the positiveness of both partial derivatives (as we mentioned earlier, in our case this positiveness is a consequence of other assumptions), the monotonicity of the function (which usually required to ensure the uniqueness of a positive equilibrium state in the model) and the concavity of function $f(S,I)$ with respect to I (together with the monotonicity, this is sufficient to ensure the global asymptotic stability of the unique equilibrium state of the model [27, 28, 30]). In this set of assumptions, (A5) is the only one which biological interpretation is not transparent. However, we have to stress that (A5) is not of principal importance and is introduced for convenience of the analysis: we need this assumption to simplify the reduction of an auxiliary system to a triangular form which is convenient for application of the generalized Rolle's theorem [8]. This makes finding estimations for the maximum number of zeros of switching functions a reasonably straightforward procedure. However, the same or a slightly weaker result can be obtained by other, more demanding techniques. Besides, (A5) holds for any realistic incidence rate. Thus, it can be immediately seen that (A5) holds for any symmetric in both arguments function $f(S,I)$. For instance, in the case of the standard bi-linear incidence rate αSI, $M_S = M_I = \alpha$, and $4M_S^2 - 3M_I^2 = \alpha^2$. The difference $4M_S^2 - 3M_I^2$ can be negative for an incidence rate which has an extremely high increase of incidence at low levels of infection (e.g., very high basic reproduction number R_0, see [27, 28, 30]) and then quickly reaches nearly constant saturation level for slightly higher infection levels. Such incidence rates can be given by Michaelis-Menten type kinetics $\alpha SI(1+\beta I)^{-1}$. For this incidence rate, $M_S = \alpha(1+\beta)^{-1}$ and $M_I = \alpha$, and hence $4M_S^2 - 3M_I^2 < 0$ if $\beta > \alpha/3$. However, the authors are unable to provide an example of a real-life infection that possesses such properties. Moreover, even if there might be found an example of an infection for which the difference $4M_S^2 - 3M_I^2$ is negative, there still remain a positive factor $1.5\mu^2 + 3(\theta+\mu)^2 + 0.5(\sigma+\mu)^2 - \theta^2$.

Model (2.1) admits up to five controls, namely: (*i*, *ii*) vaccinations of the new borns and the susceptibles, with per capita rates $\widetilde{u}(t)$ and $\widetilde{w}(t)$, respectively; (*iii*, *iv*) isolation of the exposed and the infectious individuals, with per capita rates $\widetilde{y}(t)$ and $\widetilde{z}(t)$, respectively, and (*v*) indirect measures, such as education or an imperfect quarantine, aimed at a reduction of the disease transmission, $\widetilde{v}(t)$. The action of control $\widetilde{v}(t)$ results in a reduction of transmission, $f(S,I)/(1+\widetilde{v}(t))$.

We assume that the controls are bounded; that is, relationships $\widetilde{u}(t) \in [0, \widetilde{u}_{\max}]$, $\widetilde{v}(t) \in [0, \widetilde{v}_{\max}]$, $\widetilde{w}(t) \in [0, \widetilde{w}_{\max}]$, $\widetilde{y}(t) \in [0, \widetilde{y}_{\max}]$ and $\widetilde{z}(t) \in [0, \widetilde{z}_{\max}]$ hold. Naturally, for all controls the lower bounds are the absence of a control, that is zero. The upper bounds that corresponds to the maximal intensity of a corresponding control are predetermined on the basis of available funding or the health system capability. Furthermore, for the vaccination of newborns, $\widetilde{u}_{\max} \leq 1$ holds. We also assume that

$$\widetilde{w}_{\max} \neq \sigma, \quad , \widetilde{w}_{\max} \neq \sigma + \widetilde{z}_{\max}, \quad \widetilde{w}_{\max} \neq \theta + \widetilde{y}_{\max}. \tag{2.3}$$

These constraints are introduced only for convenience of the analysis.

Introducing these controls into model (2.1), we obtain a control model

$$\begin{cases} \dot{S}(t) = \mu(1 - \widetilde{u}(t)) - (1 + \widetilde{v}(t))^{-1} f(S(t), I(t)) - (\mu + \widetilde{w}(t))S(t), \\ \dot{E}(t) = (1 + \widetilde{v}(t))^{-1} f(S(t), I(t)) - (\theta + \mu + \widetilde{y}(t))E(t), \\ \dot{I}(t) = \theta E(t) - (\sigma + \mu + \widetilde{z}(t))I(t), \\ \dot{R}(t) = \mu\widetilde{u}(t) + \widetilde{w}(t)S(t) + \widetilde{y}(t)E(t) + (\sigma + \widetilde{z}(t))I(t) - \mu R(t). \end{cases} \tag{2.4}$$

As we mentioned above, equality (2.2) enables us to omit one equation reducing the system to three equations. Therefore, further we consider the reduced system for $S(t)$, $E(t)$ and $I(t)$. Denoting

$$\begin{aligned} &u(t) = \mu(1 - \widetilde{u}(t)), \quad v(t) = (1 + \widetilde{v}(t))^{-1}, \quad w(t) = \mu + \widetilde{w}(t), \\ &y(t) = \theta + \mu + \widetilde{y}(t), \quad z(t) = \sigma + \mu + \widetilde{z}(t), \end{aligned} \tag{2.5}$$

we obtain control system

$$\begin{cases} \dot{S}(t) = u(t) - v(t)f(S(t), I(t)) - w(t)S(t), \\ \dot{E}(t) = v(t)f(S(t), I(t)) - y(t)E(t), \\ \dot{I}(t) = \theta E(t) - z(t)I(t). \end{cases} \tag{2.6}$$

Here the controls are subjects of constraints

$$\begin{aligned} &u(t) \in [u_{\min}, \mu], \quad v(t) \in [v_{\min}, 1], \quad w(t) \in [w_{\min}, w_{\max}], \\ &y(t) \in [y_{\min}, y_{\max}], \quad z(t) \in [z_{\min}, z_{\max}], \end{aligned} \tag{2.7}$$

where

$$\begin{aligned} &u_{\min} = \mu(1 - \widetilde{u}_{\max}), \quad v_{\min} = (1 + \widetilde{v}_{\max})^{-1}, \quad w_{\min} = \mu, \\ &w_{\max} = \mu + \widetilde{w}_{\max}, \quad y_{\min} = \theta + \mu, \quad y_{\max} = \theta + \mu + \widetilde{y}_{\max}, \\ &z_{\min} = \sigma + \mu, \quad z_{\max} = \sigma + \mu + \tilde{z}_{\max}. \end{aligned} \tag{2.8}$$

Control system (2.6) is defined on a finite time interval $[0, T]$ and should be complemented by initial conditions

$$S(0) = S_0, \;\; E(0) = E_0, \;\; I(0) = I_0, \;\; S_0, \; E_0, \; I_0 > 0, \;\; S_0 + E_0 + I_0 < 1.$$

Set

$$\Lambda = \Big\{ (S, E, I) : S > 0, \; E > 0, \; I > 0, \; S + E + I < 1 \Big\}$$

is the feasible region of the reduced system (2.6).

Let set $D_0(T)$ be the set of admissible controls; that is, $D_0(T)$ is the set of all possible Lebesgue measurable functions $(u(t), v(t), w(t), y(t), z(t))$, which for almost all $t \in [0, T]$ satisfy constraints (2.7). The following Lemma ensures the boundedness, positiveness and continuity of solutions for system (2.6):

Lemma 1. *Let* $(S_0, E_0, I_0) \in \Lambda$ *be held. Then for any admissible controls* $(u(t), v(t), w(t), y(t), z(t))$ *the corresponding solutions* $(S(t), E(t), I(t))$ *to system* (2.6) *are defined on the entire interval* $[0, T]$*, and inclusion*

$$(S(t), E(t), I(t)) \in \Lambda$$

holds for all $t \in [0, T]$*.* (*That is, set* Λ *is a positively invariant set of system* (2.6).)

The proof is fairly straightforward and we omit it. Proofs of such statements are given for example in [10, 33].

For control system (2.6) and for the set of admissible controls $D_0(T)$, we consider the problem of minimizing the level of infection at the end of time interval $[0, T]$. That is, we consider this system together with the objective functional

$$J_0(u(\cdot), v(\cdot), w(\cdot), y(\cdot), z(\cdot)) = \alpha E(T) + \beta I(T). \tag{2.9}$$

In this functional, α and β are positive constant. For the majority of practically relevant cases, $\alpha = \beta$. However, for the sake of generality, we prefer to allow nonequal coefficients as well. Lemma 1 and Theorem 4 in [34] (Chapter 4) ensure the existence of optimal controls $(u_*(t), v_*(t), w_*(t), y_*(t), z_*(t))$ and corresponding optimal solutions $(S_*(t), E_*(t), I_*(t))$ for this problem.

3. Pontryagin maximum principle

To analyze optimal control problem (2.6), (2.9), we apply the Pontryagin maximum principle [37]. Let us define Hamiltonian

$$\begin{aligned} &H(S, E, I, u, v, w, y, z, \psi_1, \psi_2, \psi_3) = \\ &\qquad (u - vf(S, I) - wS)\psi_1 + (vf(S, I) - yE)\psi_2 + (\theta E - zI)\psi_3, \end{aligned} \tag{3.1}$$

where ψ_1, ψ_2 and ψ_3 are adjoint variables. The Hamiltonian satisfies

$$\begin{aligned} H'_S &= vf'_S(S, I)(\psi_2 - \psi_1) - w\psi_1, \\ H'_E &= -y\psi_2 + \theta\psi_3, \\ H'_I &= vf'_I(S, I)(\psi_2 - \psi_1) - z\psi_3, \end{aligned}$$

and

$$\begin{aligned} &H'_u = \psi_1, \qquad H'_v = f(S, I)(\psi_2 - \psi_1), \\ &H'_w = -S\psi_1, \quad H'_y = -E\psi_2, \quad H'_z = -I\psi_3. \end{aligned} \tag{3.2}$$

Arguments of these partial derivatives are the same as that of Hamiltonian (3.1), and therefore we omitted them.

By the Pontryagin maximum principle, for the optimal controls

$$(u_*(t), v_*(t), w_*(t), y_*(t), z_*(t))$$

and the corresponding optimal solutions $(S_*(t), E_*(t), I_*(t))$, on interval $[0, T]$ the adjoint system

$$\begin{cases} \dot{\psi}_1^*(t) = w_*(t)\psi_1^*(t) - v_*(t)f'_S(S_*(t), I_*(t))(\psi_2^*(t) - \psi_1^*(t)), \\ \dot{\psi}_2^*(t) = y_*(t)\psi_2^*(t) - \theta\psi_3^*(t), \\ \dot{\psi}_3^*(t) = z_*(t)\psi_3^*(t) - v_*(t)f'_I(S_*(t), I_*(t))(\psi_2^*(t) - \psi_1^*(t)), \\ \psi_1^*(T) = 0, \quad \psi_2^*(T) = -\alpha, \quad \psi_3^*(T) = -\beta, \end{cases} \tag{3.3}$$

has a non-trivial solution $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$ such that optimal controls $(u_*(t), v_*(t), w_*(t), y_*(t), z_*(t))$ maximize Hamiltonian

$$H(S_*(t), E_*(t), I_*(t), u, v, w, y, z, \psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$$

with respect to variables $u \in [u_{\min}, \mu]$, $v \in [v_{\min}, 1]$, $w \in [w_{\min}, w_{\max}]$, $y \in [y_{\min}, y_{\max}]$ and $z \in [z_{\min}, z_{\max}]$ for almost all $t \in [0, T]$ and satisfy conditions

$$u_*(t) = \begin{cases} \mu & \text{, if } \quad L_u(t) > 0, \\ \forall u \in [u_{\min}, \mu] & \text{, if } \quad L_u(t) = 0, \\ u_{\min} & \text{, if } \quad L_u(t) < 0; \end{cases} \tag{3.4}$$

$$v_*(t) = \begin{cases} 1 & \text{, if } \quad L_v(t) > 0, \\ \forall v \in [v_{\min}, 1] & \text{, if } \quad L_v(t) = 0, \\ v_{\min} & \text{, if } \quad L_v(t) < 0; \end{cases} \tag{3.5}$$

$$w_*(t) = \begin{cases} w_{\max} & \text{, if } \quad L_w(t) > 0, \\ \forall w \in [w_{\min}, w_{\max}] & \text{, if } \quad L_w(t) = 0, \\ w_{\min} & \text{, if } \quad L_w(t) < 0; \end{cases} \tag{3.6}$$

$$y_*(t) = \begin{cases} y_{\max} & \text{, if } \quad L_y(t) > 0, \\ \forall y \in [y_{\min}, y_{\max}] & \text{, if } \quad L_y(t) = 0, \\ y_{\min} & \text{, if } \quad L_y(t) < 0; \end{cases} \tag{3.7}$$

$$z_*(t) = \begin{cases} z_{\max} & \text{, if } \quad L_z(t) > 0, \\ \forall z \in [z_{\min}, z_{\max}] & \text{, if } \quad L_z(t) = 0, \\ z_{\min} & \text{, if } \quad L_z(t) < 0. \end{cases} \tag{3.8}$$

Here, by (3.2), (A2) and Lemma 1, the switching functions

$$\begin{array}{lll} L_u(t) = \psi_1^*(t), & L_v(t) = \psi_2^*(t) - \psi_1^*(t), & L_w(t) = -\psi_1^*(t), \\ L_y(t) = -\psi_2^*(t), & L_z(t) = -\psi_3^*(t) & \end{array} \tag{3.9}$$

entirely determine the corresponding optimal controls.

Please note that in (3.9)

$$L_u(t) = -L_w(t), \tag{3.10}$$

and hence it suffices to establish properties of one of these two functions.

The following differential equations for functions $L_v(t)$, $L_w(t)$, $L_y(t)$ and $L_z(t)$ immediately follow from system (3.3):

$$\left\{\begin{array}{l}\dot{L}_w(t) = w_*(t)L_w(t) + v_*(t)f'_S(S_*(t), I_*(t))L_v(t),\\ \dot{L}_y(t) = y_*(t)L_y(t) - \theta L_z(t),\\ \dot{L}_z(t) = z_*(t)L_z(t) + v_*(t)f'_I(S_*(t), I_*(t))L_v(t),\\ \dot{L}_v(t) = \Big(v_*(t)f'_S(S_*(t), I_*(t)) + w_*(t)\Big)L_v(t)\\ \qquad +(w_*(t) - y_*(t))L_y(t) + \theta L_z(t),\\ L_w(T) = 0,\ L_y(T) = \alpha,\ L_z(T) = \beta,\ L_v(T) = -\alpha.\end{array}\right. \tag{3.11}$$

Moreover, function $L_u(t)$ is to be found from equality (3.10).

The following Lemma is of principal importance for problem (2.6), (2.9):

Lemma 2. *Adjoint functions $\psi_1^*(t)$, $\psi_2^*(t)$, $\psi_3^*(t)$ are not equal identically to zero on any finite subinterval $\Delta \subset [0,T]$.*

Proof. Let us assume, by contradiction, that there is a subinterval $\Delta_1 \subset [0,T]$ such that $\psi_1^*(t) = 0$ for all $t \in \Delta_1$. Then $\dot{\psi}_1^*(t) = 0$ must hold almost everywhere on Δ_1 as well. Substituting these two equalities into the first equation of system (3.3), we obtain that $\psi_2^*(t) - \psi_1^*(t) = 0$, and hence $\psi_2^*(t) = 0$ holds for all $t \in \Delta_1$, and $\dot{\psi}_2^*(t) = 0$ holds almost everywhere on Δ_1. Then the second equation of system (3.3) yields $\psi_3^*(t) = 0$ for all $t \in \Delta_1$. At the same time, the third equation of this system is also satisfied. Hence, we have $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t)) = 0$ on Δ_1. However, system (3.3) is a system of linear non-autonomous homogeneous differential equations. For such a system, this equality implies that $\psi_*(t) = 0$ holds for all $t \in [0,T]$. This contradicts the existence of a non-trivial solution $\psi_*(t)$. Hence the hypothesis is incorrect, and adjoint function $\psi_1^*(t)$ is not equal to zero on any subinterval of $[0,T]$.

Repeating the same arguments for function $\psi_2^*(t)$, we again obtain a contradiction. Therefore, adjoint function $\psi_2^*(t)$ is not equal to zero on any subinterval of $[0,T]$ either.

Arguments for $\psi_3^*(t)$ require more care. Let us assume, as above, that there is a subinterval $\Delta_3 \subset [0,T]$ such that $\psi_3^*(t) = 0$ for all $t \in \Delta_3$, and hence that $\dot{\psi}_3^*(t) = 0$ holds almost everywhere on Δ_3 as well. Then, from the third equation of system (3.3), it follows that $\psi_2^*(t) - \psi_1^*(t) = 0$ holds on Δ_3, and hence equality $\dot{\psi}_1^*(t) = \dot{\psi}_2^*(t)$ holds almost everywhere on Δ_3 as well. Furthermore, substituting $\psi_3^*(t) = 0$ into the first two equations of system (3.3), we get

$$\dot{\psi}_1^*(t) = w_*(t)\psi_1^*(t), \quad \dot{\psi}_2^*(t) = y_*(t)\psi_2^*(t), \quad t \in \Delta_3. \tag{3.12}$$

These equations imply that functions $\psi_1^*(t)$ and $\psi_2^*(t)$ cannot be equal to zero in Δ_3, as otherwise they must be equal to zero on the entire subinterval Δ_3. Hence, for all $t \in \Delta_3$, both these functions are simultaneously either positive, or negative. By (3.6) and (3.7), this means that controls $w_*(t)$ and $y_*(t)$ take either values $w_{\min}$ and $y_{\min}$, or $w_{\max}$ and $y_{\max}$, respectively. Furthermore, by (3.12), the equality of the functions and their derivatives also means that either $w_{\min} = y_{\min}$, or $w_{\max} = y_{\max}$ must hold. However, by (2.3) and (2.8), neither of these equalities is possible, and hence we have a contradiction again. This completes the proof. □

Lemma 2 combined with relationships (3.6)–(3.9) immediately leads to the following Corollary:

Corollary 3. *Optimal controls $w_*(t)$, $y_*(t)$ and $z_*(t)$ are bang-bang controls, which take only values $\{w_{\min}, w_{\max}\}$, $\{y_{\min}, y_{\max}\}$ and $\{z_{\min}, z_{\max}\}$, respectively.*

An important property which we would like to stress is that these controls have no singularities (see [38] for details), and they are entirely determined by the corresponding switching functions $L_w(t)$, $L_y(t)$ and $L_z(t)$.

To establish the same properties for function $L_v(t)$ we have to prove the following Lemma:

Lemma 4. *Switching function $L_v(t)$ is not equal identically to zero on any finite subinterval of interval $[0,T]$.*

Proof. By contradiction, let us assume that there is a subinterval $\Delta_0 \subset [0,T]$ where $L_v(t) = \psi_2^*(t) - \psi_1^*(t) = 0$. Then, as in the proof of Lemma 2, equalities

$$\psi_1^*(t) = \psi_2^*(t), \quad \dot{\psi}_1^*(t) = \dot{\psi}_2^*(t) \tag{3.13}$$

necessary hold. Moreover, on subinterval Δ_0 the equations of system (3.3) are:

$$\begin{cases} \dot{\psi}_1^*(t) = w_*(t)\psi_1^*(t), \\ \dot{\psi}_2^*(t) = y_*(t)\psi_2^*(t) - \theta\psi_3^*(t), \\ \dot{\psi}_3^*(t) = z_*(t)\psi_3^*(t). \end{cases} \tag{3.14}$$

If $\psi_1^*(t) = 0$ at any point of the subinterval Δ_0, then it follows from the first equation that this function is equal to zero for all $t \in \Delta_0$. Moreover, by (3.13), in this case $\psi_2(t)$ and $\dot{\psi}_2^*(t)$ are equal to zero on the interval Δ_0 as well. (Derivative $\dot{\psi}_2^*(t)$ is equal to zero almost everywhere on Δ_0.) Then the second equation yields $\psi_3^*(t) = 0$ for all $t \in \Delta_0$, and hence, by the same arguments as these in the proof of Lemma 2, we come to a contradiction. Therefore, $\psi_1^*(t) \neq 0$ on Δ_0, and, by (3.13), functions $\psi_1^*(t)$ and $\psi_2^*(t)$ are of the same sign. Then, by (3.6), (3.7), on subinterval Δ_0 controls $w_*(t)$ and $y_*(t)$ takes values either $\{w_{\min}, y_{\min}\}$, or $\{w_{\max}, y_{\max}\}$, respectively.

Function $\psi_3^*(t)$ cannot be equal to zero on subinterval Δ_0. In opposite case we have $\psi_3^*(t) = 0$ for all $t \in \Delta_0$. Then, applying the arguments from the proof of Lemma 2, we come to a contradiction. Therefore, function $\psi_3^*(t)$ is either positive, or negative definite on the subinterval Δ_0. Hence, by (3.8), control $z_*(t)$ takes value either $z_{\min}$, or $z_{\max}$.

Finally, equalities (3.13) allows us to combine the first and second equations of (3.14) and yields equality

$$(y_*(t) - w_*(t))\psi_2^*(t) = \theta\psi_3^*(t), \quad t \in \Delta_0. \tag{3.15}$$

Differentiating this equality, and then substituting the derivatives from (3.14), we obtain

$$y_*(t)(y_*(t) - w_*(t))\psi_2^*(t) = \theta(z_*(t) + y_*(t) - w_*(t))\psi_3^*(t), \quad t \in \Delta_0.$$

Combining this with (3.15), we get $(z_*(t) - w_*(t))\psi_3^*(t) = 0$, and hence $w_*(t) = z_*(t)$ holds for all $t \in \Delta_0$. This implies that either $w_{\min} = z_{\min}$, or $w_{\max} = z_{\max}$, or $w_{\min} = z_{\max}$, or $w_{\max} = z_{\min}$ must hold. However, by (2.3) and (2.8), neither of these equalities is possible, and hence we came to a contradiction. Therefore, the initial assumption is not correct, and function $L_v(t)$ is not equal to zero on any finite subinterval of the interval $[0, T]$. This completes the proof. □

Lemma 4 and (3.5) immediately lead to the Corollary:

Corollary 5. *Optimal control $v_*(t)$ is a bang-bang control, which only take values $\{v_{min}, 1\}$ on interval $[0, T]$. The control has no singularities (see [38]) and is entirely determined by switching functions $L_v(t)$.*

4. Estimation for the number of switchings

In previous Section, we show that in problem (2.6), (2.9) the optimal controls $(v_*(t), w_*(t), y_*(t), z_*(t))$ are bang-bang controls with values $\{v_{\min}, 1\}$, $\{w_{\min}, w_{\max}\}$, $\{y_{\min}, y_{\max}\}$ and $\{z_{\min}, z_{\max}\}$, respectively. If we would be able to estimate the maximum possible number of switchings of the controls, then we can immediately reduce the optimal control problem to a considerably simpler problem of the finite dimensional optimization, which then can be solved numerically. Formulas (3.4)–(3.8) imply that for this we have to estimate the number of zeros of the corresponding switching functions $L_v(t)$, $L_w(t)$, $L_y(t)$, $L_z(t)$ defined by system (3.11). In this system, equations for $L_v(t)$, $L_y(t)$ and $L_z(t)$ are independent from $L_w(t)$, and hence we will firstly estimate the number of zeros for function $L_v(t)$, then for $L_y(t)$ and $L_z(t)$, and finally we will estimate the number of switchings of function $L_w(t)$.

Function $L_v(t)$, $L_y(t)$ and $L_z(t)$ satisfy differential equations

$$\begin{cases} \dot{L}_v(t) = a(t)L_v(t) + (w_*(t) - y_*(t))L_y(t) + \theta L_z(t), \\ \dot{L}_y(t) = y_*(t)L_y(t) - \theta L_z(t), \\ \dot{L}_z(t) = z_*(t)L_z(t) + b(t)L_v(t). \end{cases} \tag{4.1}$$

Here $a(t) = v_*(t)f_S'(S_*(t), I_*(t)) + w_*(t)$ and $b(t) = v_*(t)f_I'(S_*(t), I_*(t))$. Please note that, by (A4), (2.8) and Corollary 5, $a(t)$ and $b(t)$ are positive on the interval $[0, T]$.

In order to simplify equations of system (4.1), let us denote

$$G(t) = \theta^{-1}(w_*(t) - y_*(t))L_y(t) + L_z(t). \tag{4.2}$$

Then

$$\dot{L}_v(t) = a(t)L_v(t) + \theta G(t).$$

To obtain the differential equation for $G(t)$, we have to be sure that $G(t)$ is differentiable almost everywhere on $[0,T]$. For the differentiability of $G(t)$, it suffices if $w_*(t)$ and $y_*(t)$ are piecewise constant functions; that is they must have a finite number of switchings on this interval. In turn, this implies that the corresponding switching functions $L_w(t)$ and $L_y(t)$ have finite numbers of zeros on $[0,T]$. From (3.11) it follows that this is possible if $L_v(t)$ has a finite number of zeros. Hence, we assume that the following Condition holds:

Condition 6. *Let $L_v(t)$ has a finite number of zeros on interval $[0,T]$.*

Further we will demonstrate that this Condition is correct.

Condition 6 ensures that function $G(t)$ is differentiable almost everywhere on $[0,T]$, and hence functions $L_v(t)$, $G(t)$ and $L_z(t)$ satisfy differential equations

$$\begin{cases} \dot{L}_v(t) = a(t)L_v(t) + \theta G(t), \\ \dot{G}(t) = b(t)L_v(t) + y_*(t)G(t) - (w_*(t) - z_*(t))L_z(t), \\ \dot{L}_z(t) = b(t)L_v(t) + z_*(t)L_z(t). \end{cases} \tag{4.3}$$

Please note that solutions $L_v(t)$, $G(t)$, $L_z(t)$ of system (4.3) and functions $L_v(t)$, $L_y(t)$ and $L_z(t)$ satisfying system (4.1) are absolutely continuous functions on the interval $[0,T]$. On the other hand, the auxiliary function $G(t)$ defined by formula (4.2) is only a piecewise absolutely continuous function, which satisfies almost everywhere the same differential equation as the function $G(t)$ in system (4.3). It is easy to show that under such changing the function $G(t)$ the switching function $L_v(t)$ is not changed.

For function $L_v(t)$ the following Lemma holds:

Lemma 7. *Switching function $L_v(t)$ has at most two distinct zeros on interval $[0,T)$.*

Proof. System (4.3) is a system of linear non-autonomous homogeneous differential equations. The matrix of this system can be brought on the whole interval $[0,T]$ to the upper-triangle form

$$\begin{pmatrix} a(t) & \theta & 0 \\ b(t) & y_*(t) & -(w_*(t) - z_*(t)) \\ b(t) & 0 & z_*(t) \end{pmatrix} \longrightarrow \begin{pmatrix} * & * & 0 \\ 0 & * & * \\ 0 & 0 & * \end{pmatrix},$$

which is considerably simpler for the analysis. Indeed, substitution

$$\bar{L}_v(t) = L_v(t), \;\; \bar{G}(t) = G(t), \;\; \bar{L}_z(t) = L_z(t) + q_1(t)L_v(t) + q_2(t)G(t),$$

where functions $q_1(t)$ and $q_2(t)$ are to be defined later, yields

$$(4.4)\qquad \begin{cases} \dot{\bar{L}}_v(t) = a(t)\bar{L}_v(t) + \theta\bar{G}(t), \\ \dot{\bar{G}}(t) \ = \Big(b(t) + (w_*(t) - z_*(t))q_1(t)\Big)\bar{L}_v(t) \\ \qquad +\Big(y_*(t) + (w_*(t) - z_*(t))q_2(t)\Big)\bar{G}(t) \\ \qquad -(w_*(t) - z_*(t))\bar{L}_z(t), \\ \dot{\bar{L}}_z(t) = \Big[\dot{q}_1(t) + b(t) + (a(t) - z_*(t))q_1(t) \\ \qquad +b(t)q_2(t) + (w_*(t) - z_*(t))q_1(t)q_2(t)\Big]\bar{L}_v(t) \\ \qquad +\Big[\dot{q}_2(t) + \theta q_1(t) + (y_*(t) - z_*(t))q_2(t) \\ \qquad +(w_*(t) - z_*(t))q_2^2(t)\Big]\bar{G}(t) \\ \qquad +\Big(z_*(t) - (w_*(t) - z_*(t))q_2(t)\Big)\bar{L}_z(t). \end{cases}$$

To simplify the second and third equations of system (4.4), we define $q_1(t)$ and $q_2(t)$ as following:

$$(4.5)\qquad \begin{cases} \dot{q}_1(t) = -b(t) - (a(t) - z_*(t))q_1(t) - b(t)q_2(t) \\ \qquad -(w_*(t) - z_*(t))q_1(t)q_2(t), \\ \dot{q}_2(t) = -\theta q_1(t) - (y_*(t) - z_*(t))q_2(t) - (w_*(t) - z_*(t))q_2^2(t). \end{cases}$$

Then

$$\begin{cases} \dot{\bar{L}}_v(t) = a(t)\bar{L}_v(t) + \theta\bar{G}(t), \\ \dot{\bar{G}}(t) \ = \Big(b(t) + (w_*(t) - z_*(t))q_1(t)\Big)\bar{L}_v(t) \\ \qquad +\Big(y_*(t) + (w_*(t) - z_*(t))q_2(t)\Big)\bar{G}(t) \\ \qquad -(w_*(t) - z_*(t))\bar{L}_z(t), \\ \dot{\bar{L}}_z(t) = \Big(z_*(t) - (w_*(t) - z_*(t))q_2(t)\Big)\bar{L}_z(t). \end{cases}$$

Substitution

$$\widetilde{L}_v(t) = \bar{L}_v(t), \quad \widetilde{G}(t) = \bar{G}(t) + q_3(t)\bar{L}_v(t), \quad \widetilde{L}_z(t) = \bar{L}_z(t),$$

where function $q_3(t)$ is to be defined later, yields

$$\begin{cases} \dot{\widetilde{L}}_v(t) = (a(t) - \theta q_3(t))\widetilde{L}_v(t) + \theta\widetilde{G}(t), \\ \dot{\widetilde{G}}(t) \ = \Big[\dot{q}_3(t) + b(t) + (w_*(t) - z_*(t))(q_1(t) - q_2(t)q_3(t)) \\ \qquad +(a(t) - y_*(t))q_3(t) - \theta q_3^2(t)\Big]\widetilde{L}_v(t) \\ \qquad +\Big(y_*(t) + (w_*(t) - z_*(t))q_2(t) + \theta q_3(t)\Big)\widetilde{G}(t) \\ \qquad -(w_*(t) - z_*(t))\widetilde{L}_z(t), \\ \dot{\widetilde{L}}_z(t) = \Big(z_*(t) - (w_*(t) - z_*(t))q_2(t)\Big)\widetilde{L}_z(t). \end{cases}$$

To simplify the second equation of this system, we define $q_3(t)$ by equation

$$\begin{aligned}\dot{q}_3(t) = -b(t) - (w_*(t) - z_*(t))(q_1(t) - q_2(t)q_3(t)) \\ -(a(t) - y_*(t))q_3(t) + \theta q_3^2(t).\end{aligned} \tag{4.6}$$

Then

$$\begin{cases} \dot{\widetilde{L}}_v(t) = (a(t) - \theta q_3(t))\widetilde{L}_v(t) + \theta \widetilde{G}(t), \\ \dot{\widetilde{G}}(t) = \Big(y_*(t) + (w_*(t) - z_*(t))q_2(t) + \theta q_3(t)\Big)\widetilde{G}(t) \\ \qquad -(w_*(t) - z_*(t))\widetilde{L}_z(t), \\ \dot{\widetilde{L}}_z(t) = \Big(z_*(t) - (w_*(t) - z_*(t))q_2(t)\Big)\widetilde{L}_z(t). \end{cases} \tag{4.7}$$

We have to prove that functions $q_1(t)$, $q_2(t)$ and $q_3(t)$ exist on the entire interval $[0, T]$. In order to do this, by (4.5) and (4.6), we write the system of quadratic non-autonomous differential equations for $q_1(t)$, $q_2(t)$ and $q_3(t)$ in matrix form:

$$\begin{cases} \dot{q}_1(t) = q^\top(t)Q_1(t)q(t) + d_1^\top(t)q(t) + r_1(t), \\ \dot{q}_2(t) = q^\top(t)Q_2(t)q(t) + d_2^\top(t)q(t) + r_2(t), \\ \dot{q}_3(t) = q^\top(t)Q_3(t)q(t) + d_3^\top(t)q(t) + r_3(t). \end{cases} \tag{4.8}$$

Here

$$Q_1(t) = \begin{pmatrix} 0 & -0.5(w_*(t) - z_*(t)) & 0 \\ -0.5(w_*(t) - z_*(t)) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$Q_2(t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -(w_*(t) - z_*(t)) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$Q_3(t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0.5(w_*(t) - z_*(t)) \\ 0 & 0.5(w_*(t) - z_*(t)) & \theta \end{pmatrix}$$

are symmetric matrices;

$$d_1(t) = \begin{pmatrix} -(a(t) - z_*(t)) \\ -b(t) \\ 0 \end{pmatrix}, \quad d_2(t) = \begin{pmatrix} -\theta \\ -(y_*(t) - z_*(t)) \\ 0 \end{pmatrix},$$

$$d_3(t) = \begin{pmatrix} -(w_*(t) - z_*(t)) \\ 0 \\ -(a(t) - y_*(t)) \end{pmatrix}$$

are vector functions, and

$$r_1(t) = -b(t), \quad r_2(t) = 0, \quad r_3(t) = -b(t)$$

are scalar functions; sign $^\top$ means a transpose. The system should be complemented by initial conditions $q_i(0) = q_i^0$, $i = 1, 2, 3$. This Cauchy problem has a solution $q(t) = (q_1(t), q_2(t), q_3(t))^\top$ on an interval $[0, t_{\max})$, where $t_{\max}$ can be either infinity or a finite number [20, 34]. However, it may occur that $t_{\max} \leq T$, and hence we have to prove the existence of the solution of this system on the entire interval $[0, T]$. This ensures that system (4.7), which has the required upper triangular matrix, is well defined and also has a solution on this interval.

By contradiction, let assume that for any solution of system (4.8) the maximal interval of the existence is $[0, t_1)$, where $t_1 \leq T$. Then, by Lemma in [7] (Paragraph 14, Chapter 4), we have

$$\lim_{t \to t_1 - 0} \|q(t)\| = +\infty, \tag{4.9}$$

where $\|\cdot\|$ is the usual Euclidean norm of a vector. Hence, there are $\rho > 0$ and $t_0 \in [0, t_1)$, which are to be defined later, such that $\|q(t)\| \geq \rho$ for all $t \in [t_0, t_1)$. Function $\|q(t)\|$ satisfies

$$\begin{aligned} \frac{d}{dt}\Big(\|q(t)\|\Big) &= \|q(t)\|^{-1} \\ &\times \Big[\Big(q_1(t)q^\top(t)Q_1(t)q(t) + q_2(t)q^\top(t)Q_2(t)q(t) + q_3(t)q^\top(t)Q_3(t)q(t)\Big) \\ &+ \Big(q_1(t)d_1^\top(t)q(t) + q_2(t)d_2^\top(t)q(t) + q_3(t)d_3^\top(t)q(t)\Big) \\ &+ \Big(r_1(t)q_1(t) + r_2(t)q_2(t) + r_3(t)q_3(t)\Big)\Big]. \end{aligned}$$

Recalling definition of $a(t)$ and $b(t)$, we obtain estimations

$$r_1(t)q_1(t) + r_2(t)q_2(t) + r_3(t)q_3(t) \leq C \cdot \|q(t)\|$$

and

$$q_1(t)d_1^\top(t)q(t) + q_2(t)d_2^\top(t)q(t) + q_3(t)d_3^\top(t)q(t) \leq B \cdot \|q(t)\|^2,$$

where $C = \sqrt{2M_I^2}$ and $B = \sqrt{4M_S^2 + M_I^2 + 5w_{\max}^2 + 3y_{\max}^2 + 4z_{\max}^2 + \theta^2}$.

Finally,

$$\begin{aligned} &q_1(t)q^\top(t)Q_1(t)q(t) + q_2(t)q^\top(t)Q_2(t)q(t) + q_3(t)q^\top(t)Q_3(t)q(t) \\ &\leq |q_1(t)|\,\|Q_1(t)q(t)\|\,\|q(t)\| + |q_2(t)|\,\|Q_2(t)q(t)\|\,\|q(t)\| + |q_3(t)|\,\left|q^\top(t)Q_3(t)q(t)\right|. \end{aligned}$$

Here,

$$\|Q_1(t)q(t)\| \leq 0.5\sqrt{w_{\max}^2 + z_{\max}^2}\,\|q(t)\|,\ \|Q_2(t)q(t)\| \leq \sqrt{w_{\max}^2 + z_{\max}^2}\,\|q(t)\|$$

and

$$\left|q^\top(t)Q_3(t)q(t)\right| \leq 0.5\Big(\theta + \sqrt{\theta^2 + w_{\max}^2 + z_{\max}^2}\Big)\|q(t)\|^2,$$

and hence

$$q_1(t)q^\top(t)Q_1(t)q(t) + q_2(t)q^\top(t)Q_2(t)q(t) + q_3(t)q^\top(t)Q_3(t)q(t) \leq A \cdot \|q(t)\|^3,$$

where $A = \sqrt{1.75 \cdot (z_{\max}^2 + w_{\max}^2) + \theta^2}$. Therefore,

$$\frac{d}{dt}\Big(\|q(t)\|\Big) \leq A \cdot \|q(t)\|^2 + B \cdot \|q(t)\| + C, \quad t \in [t_0, t_1). \tag{4.10}$$

Consider the quadratic equation

$$AK^2 - BK + C = 0.$$

By (A5), (2.8) and inequality $-mn \geq -0.5m^2 - 0.5n^2$, the determinant of this equation satisfies

$$\begin{aligned} B^2 - 4AC &= 4M_S^2 + M_I^2 + 5w_{max}^2 + 3y_{max}^2 + 4z_{max}^2 + \theta^2 \\ &\quad - \sqrt{8M_I^2} \cdot \sqrt{7(w_{max}^2 + z_{max}^2) + 4\theta^2} \\ &\geq 4M_S^2 - 3M_I^2 + 1.5w_{max}^2 + 3y_{max}^2 + 0.5z_{max}^2 - \theta^2 > 0, \end{aligned}$$

and hence the quadratic equation has two real roots. The larger root satisfies

$$K_0 = \frac{B + \sqrt{B^2 - 4AC}}{2A}.$$

For all vectors q such that $\|q\| \geq \rho$, let define function $V(q) = \|q\| + K_0$. By (4.10) and using equality $AK_0^2 - BK_0 + C = 0$, it is easy to see that

$$\frac{dV(q(t))}{dt} \leq AV^2(q(t)) - (2AK_0 - B)V(q(t)), \quad t \in [t_0, t_1).$$

Consider auxiliary Cauchy problem

$$\begin{cases} \dot{h}(t) = Ah^2(t) - (2AK_0 - B)h(t), & t \in [t_0, t_1] \\ h(t_0) = h_0 \geq K_0 + \rho. \end{cases}$$

This Bernoulli equation has solution

$$h(t) = \left(\frac{A}{2AK_0 - B} + \left[\frac{1}{h_0} - \frac{A}{2AK_0 - B} \right] e^{(2AK_0 - B)(t - t_0)} \right)^{-1}, \quad t \in [t_0, t_1].$$

We assume that values ρ and t_0 are such that the value in brackets is not equal to zero on interval $[t_0, t_1]$. For instance, for a given ρ it would be sufficient to take t_0 such that $(t_1 - t_0)$ were sufficiently small. This ensures that function $h(t)$ is defined and finite on this interval. Since $h_0 > (2AK_0 - B)A^{-1}$, the value in the square brackets is negative, and hence for all $t \in [t_0, t_1]$ function $h(t)$ is a finite positive definite and monotonically increasing function such that inequality $h(t) < h(t_1)$ holds for for all $t \in [t_0, t_1)$. Therefore, by the Chaplygin theorem (see [40]) and taking into consideration the initial condition $h_0 = V(q(t_0)) = \|q(t_0)\| + K_0$, the inequality

$$\|q(t)\| < h(t) - K_0 < h(t_1) - K_0$$

holds for all $t \in (t_0, t_1)$ as well. However, this inequality contradicts to (4.9), and hence there is a solution $\widehat{q}(t)$ of system (4.8) defined on the entire interval $[0, T]$. This ensure that system (4.7) is defined on this interval as well.

The found solution $\widehat{q}(t) = (\widehat{q}_1(t), \widehat{q}_2(t), \widehat{q}_3(t))^\top$ enables us to rewrite system (4.7) as

$$
(4.11) \qquad \begin{cases} \dot{\widetilde{L}}_v(t) = (a(t) - \theta\widehat{q}_3(t))\widetilde{L}_v(t) + \theta\widetilde{G}(t), \\ \dot{\widetilde{G}}(t) = \Big(y_*(t) + (w_*(t) - z_*(t))\widehat{q}_2(t) + \theta\widehat{q}_3(t)\Big)\widetilde{G}(t) \\ \qquad\quad -(w_*(t) - z_*(t))\widetilde{L}_z(t), \\ \dot{\widetilde{L}}_z(t) = \Big(z_*(t) - (w_*(t) - z_*(t))\widehat{q}_2(t)\Big)\widetilde{L}_z(t). \end{cases}
$$

Depending on the sign of $(\widetilde{w}_{\max} - \sigma)$, we have to consider two possible cases.

Case 1. Let inequality $\widetilde{w}_{\max} - \sigma < 0$ be held. Then, by (2.8), $w_{\max} < z_{\min}$, and hence almost everywhere on the interval $[0, T]$ the inequality $w_*(t) - z_*(t) < 0$ holds. Taking into consideration this inequality and initial condition $\widetilde{L}_v(T) = L_v(T) = -\alpha < 0$, we can show that function $L_v(t) = \widetilde{L}_v(t)$ has no more than two different zeros on interval $[0, T)$.

Indeed, assume that $\widetilde{L}_v(t)$ has at least three different zeros τ_i, $i = 1, 2, 3$, such that $0 \le \tau_1 < \tau_2 < \tau_3 < T$. Then, by the generalized Rolle's theorem [8] applied to the first equation of (4.11), function $\widetilde{G}(t)$ has on interval $(0, T)$ at least two zeros η_i, $i = 1, 2$, such that $0 < \eta_1 < \eta_2 < T$ hold. The same theorem applied to the second equation yields that $\widetilde{L}_z(t)$ has at least one zero $\xi \in (0, T)$. However, the third equation of (4.11), which this function satisfies, is a linear homogeneous differential equation, and hence in this case $\widetilde{L}_z(t) = 0$ must hold everywhere on $[0, T]$. Substituting $\widetilde{L}_z(t) = 0$ and applying the same argument to the second equation, we have the same conclusion, namely that $\widetilde{G}(t) = 0$ must hold on $[0, T]$. Then the same arguments yields the same conclusion for function $\widetilde{L}_v(t)$, that is under the hypothesis $\widetilde{L}_v(t) = 0$ on $[0, T]$. However, this contradicts to the initial condition for $\widetilde{L}_v(t)$, and hence the hypothesis is not correct and function $L_v(t)$ has no more than two different zeros on the interval $[0, T)$.

Case 2. Let inequality $\widetilde{w}_{\max} - \sigma > 0$ be held. Then, by (2.8), $w_{\max} > z_{\min}$, and hence almost everywhere on the interval $[0, T]$ the difference $(w_*(t) - z_*(t))$ takes both positive and negative values. Therefore, we can not apply the generalized Rolle's theorem to system (4.11). This is rather inconvenient, and to overpass this inconvenience we introduce auxiliary function $\widetilde{\Phi}(t) = (w_*(t) - z_*(t))\widetilde{L}_z(t)$. The corresponding differential equation which this function satisfies almost everywhere on interval $[0, T]$, is

$$
\dot{\widetilde{\Phi}}(t) = [z_*(t) - (w_*(t) - z_*(t))\widehat{q}_2(t)]\widetilde{\Phi}(t),
$$

and now system (4.11) can be rewritten as

$$
(4.12) \qquad \begin{cases} \dot{\widetilde{L}}_v(t) = (a(t) - \theta\widehat{q}_3(t))\widetilde{L}_v(t) + \theta\widetilde{G}(t), \\ \dot{\widetilde{G}}(t) \;=\; \Big(y_*(t) + (w_*(t) - z_*(t))\widehat{q}_2(t) + \theta\widehat{q}_3(t)\Big)\widetilde{G}(t) - \widetilde{\Phi}(t), \\ \dot{\widetilde{\Phi}}(t) \;=\; \Big(z_*(t) - (w_*(t) - z_*(t))\widehat{q}_2(t)\Big)\widetilde{\Phi}(t). \end{cases}
$$

Please note that the correspondence between systems (4.11) and (4.12) is the same as that between systems (4.1) and (4.3).

Applying the same arguments as in Case 1 to system (4.12), we came to conclusion that in this case function $L_v(t) = \widetilde{L}_v(t)$ has no more than two different zeros on the interval $[0, T)$ as well. This completes the proof. □

Thus, Lemma 7 confirms that Condition 6 is correct.

5. Optimal controls

Having the estimates for the maximum number of switchings for function $L_v(t)$ found, we can, analyzing system (3.11), find the maximal possible number of switchings for other switching functions and then reduce the optimal control problem to a considerably simpler problem of the finite-dimensional optimization, which can be solved numerically. However, to apply a numerical procedure we need to define a specific parametrization for the incidence rate $f(S, I)$. Therefore, since our objective is to consider impact of the general non-linearity on the controls, we will not do this in this paper. Instead, we consider system (3.11) and describe the possible variants of optimal controls, which can arise in this problem.

According to the number of possible zeros of function $L_v(t)$, that is (i) no zeros, (ii) one zero, or (iii) two zeros, three basic Scenarios are possible for this problem. These scenarios are given below.

Scenario 1 (S1). Function $L_v(t)$ has no zeros on the interval $(0, T]$. Then, we have $L_v(t) < 0$ for all $t \in (0, T]$, and hence, by (3.5), the optimal control $v_*(t)$ has the following type

$$
v_*(t) = v_{\min}, \quad t \in [0, T].
$$

Since $L_v(t) < 0$ holds for all $t \in (0, T]$, then, by (3.11), the inequalities $L_z(t) > 0$, $L_y(t) > 0$ hold for all $t \in [0, T]$ as well. Furthermore, $L_w(t) > 0$ and, by (3.10), $L_u(t) < 0$ hold for $t \in [0, T)$. Hence, by (3.4) and (3.6)–(3.8), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are:

$$
u_*(t) = u_{\min}, \;\; w_*(t) = w_{\max}, \;\; y_*(t) = y_{\max}, \;\; z_*(t) = z_{\max}, \;\; t \in [0, T].
$$

Scenario 2 (S2). Function $L_v(t)$ has one zero on the interval $(0,T)$. Then

$$L_v(t)\begin{cases} >0 & \text{, if } \quad 0\le t<\tau_1^*, \\ =0 & \text{, if } \quad t=\tau_1^*, \\ <0 & \text{, if } \quad \tau_1^*<t\le T, \end{cases} \tag{5.1}$$

where $\tau_1^* \in (0,T)$ is the zero of function $L_v(t)$. Hence, by (3.5), the optimal control $v_*(t)$ is

$$v_*(t)=\begin{cases} 1 & \text{, if } \quad 0\le t\le \tau_1^*, \\ v_{\min} & \text{, if } \quad \tau_1^*<t\le T. \end{cases}$$

By (3.11), (5.1) and the equality

$$L_v(t)=L_w(t)-L_y(t), \quad t\in[0,T], \tag{5.2}$$

that follows from (3.9), it is easy to see that the behavior of switching functions $L_w(t)$, $L_y(t)$ and $L_z(t)$ is determined by values $L_w(0)$, $L_y(0)$ and $L_z(0)$, respectively. Moreover, the analysis of these functions depends on the sign of $L_w(0)$. Therefore, there are two possible cases.

Case 2.1. Let $L_w(0)\ge 0$. Then, by (5.1) and the first equation of system (3.11), $L_w(t)>0$ for all $t\in(0,T)$. Depending on the sign of $L_z(0)$, there are two subcases.

Subcase 2.1.1. If $L_z(0)\ge 0$, then, by (5.1) and the third equation of (3.11), $L_z(t)>0$ holds for all $t\in(0,T]$. From the second equation of (3.11), $L_y(t)>0$ holds for all $t\in[0,T]$, and hence $L_y(0)>0$. By (5.1), (5.2), inequality $L_w(0)>L_y(0)$ holds. In this case this inequality is contradictory only for $L_w(0)=0$, whereas for $L_w(0)>0$, by (3.4), (3.6)–(3.8) and (3.10), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ are as in Scenario 1.

Subcase 2.1.2. If $L_z(0)<0$, then, by (5.1) and the third equation of (3.11),

$$L_z(t)\begin{cases} <0 & \text{, if } \quad 0\le t<\eta_1^*, \\ =0 & \text{, if } \quad t=\eta_1^*, \\ >0 & \text{, if } \quad \eta_1^*<t\le T. \end{cases} \tag{5.3}$$

Here $\eta_1^*\in(0,T)$ is the zero of function $L_z(t)$ such that $\eta_1^*<\tau_1^*$ holds. Therefore, from the second equation of system (3.11) it immediately follows that function $L_y(t)$ has no more than one zero. That is there are two possibilities: either

$$L_y(t)>0, \quad t\in(0,T],$$

or

$$L_y(t)\begin{cases} <0 & \text{, if } \quad 0\le t<\chi_1^*, \\ =0 & \text{, if } \quad t=\chi_1^*, \\ >0 & \text{, if } \quad \chi_1^*<t\le T, \end{cases} \tag{5.4}$$

where $\chi_1^* \in (0,T)$ is the zero of function $L_y(t)$ such that $\chi_1^* < \eta_1^*$. Hence, by (3.4), (3.6)–(3.8) and (3.10), in this case the following types of the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ are possible: either

$$u_*(t) = u_{\min}, \quad w_*(t) = w_{\max}, \quad y_*(t) = y_{\max}, \quad t \in [0,T];$$

$$z_*(t) = \begin{cases} z_{\min} & , \text{ if } \; 0 \le t \le \eta_1^*, \\ z_{\max} & , \text{ if } \; \eta_1^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^*$; or

$$u_*(t) = u_{\min}, \quad w_*(t) = w_{\max}, \quad t \in [0,T];$$

$$y_*(t) = \begin{cases} y_{\min} & , \text{ if } \; 0 \le t \le \chi_1^*, \\ y_{\max} & , \text{ if } \; \chi_1^* < t \le T, \end{cases} \quad z_*(t) = \begin{cases} z_{\min} & , \text{ if } \; 0 \le t \le \eta_1^*, \\ z_{\max} & , \text{ if } \; \eta_1^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^*$ and $\chi_1^* < \eta_1^*$.

Case 2.2. Let $L_w(0) < 0$. Then, from (5.1) and the first equation of system (3.11) it follows that function $L_w(t)$ has one zero $\xi_1^* \in (0,T)$, such that $\xi_1^* < \tau_1^*$ holds, and hence

$$L_w(t) \begin{cases} < 0 & , \text{ if } \; 0 \le t < \xi_1^*, \\ = 0 & , \text{ if } \; t = \xi_1^*, \; t = T, \\ > 0 & , \text{ if } \; \xi_1^* < t < T. \end{cases} \tag{5.5}$$

From (5.1), (5.2) and (5.5) it follows that

$$L_y(0) < 0. \tag{5.6}$$

Now, depending on the sign of $L_z(0)$, there are two subcases.

Subcase 2.2.1. If $L_z(0) \ge 0$, then from (5.1) and the third equation of system (3.11), it follows that $L_z(t) > 0$ for all $t \in (0,T]$. From the second equation of this system $L_y(t) > 0$ for all $t \in [0,T]$, and in particular $L_y(0) > 0$. However, this contradicts (5.6). Hence, this subcase is impossible.

Subcase 2.2.2. If $L_z(0) < 0$, then, from (5.1) and the third equation of system (3.11) it follows that function $L_z(t)$ satisfies (5.3). By this, inequality (5.6) and the second equation of system (3.11), (5.4) is valid for function $L_y(t)$. Then, from (3.4), (3.6)–(3.8) and (3.10), it follows that the following types of the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ are possible in this case:

$$u_*(t) = \begin{cases} \mu & , \text{ if } \; 0 \le t \le \xi_1^*, \\ u_{\min} & , \text{ if } \; \xi_1^* < t \le T, \end{cases} \quad w_*(t) = \begin{cases} w_{\min} & , \text{ if } \; 0 \le t \le \xi_1^*, \\ w_{\max} & , \text{ if } \; \xi_1^* < t \le T, \end{cases}$$

$$y_*(t) = \begin{cases} y_{\min} & , \text{ if } \; 0 \le t \le \chi_1^*, \\ y_{\max} & , \text{ if } \; \chi_1^* < t \le T, \end{cases} \quad z_*(t) = \begin{cases} z_{\min} & , \text{ if } \; 0 \le t \le \eta_1^*, \\ z_{\max} & , \text{ if } \; \eta_1^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^*$, $\xi_1^* < \tau_1^*$ and $\chi_1^* < \eta_1^*$.

The possible types of physical controls $\widetilde{u}_*(t)$, $\widetilde{v}_*(t)$, $\widetilde{w}_*(t)$, $\widetilde{y}_*(t)$ and $\widetilde{z}_*(t)$ corresponding to the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ for Scenarios 1 and 2 are summarized in Table 1.

TABLE 1. Variants of physical controls for Scenarios 1 and 2.

#	$\widetilde{u}_*(t)$	$\widetilde{v}_*(t)$	$\widetilde{w}_*(t)$	$\widetilde{y}_*(t)$	$\widetilde{z}_*(t)$
S1	$\widetilde{u}_{\max}$	$\widetilde{v}_{\max}$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$\widetilde{z}_{\max}$
S2	$\widetilde{u}_{\max}$	$0 \quad, [0,\tau_1^*]$ $\widetilde{v}_{\max}, (\tau_1^*, T]$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$\widetilde{z}_{\max}$
	$\widetilde{u}_{\max}$	$0 \quad, [0,\tau_1^*]$ $\widetilde{v}_{\max}, (\tau_1^*, T]$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$0 \quad, [0,\eta_1^*]$ $\widetilde{z}_{\max}, (\eta_1^*, T]$ $\eta_1^* < \tau_1^*$
	$\widetilde{u}_{\max}$	$0 \quad, [0,\tau_1^*]$ $\widetilde{v}_{\max}, (\tau_1^*, T]$	$\widetilde{w}_{\max}$	$0 \quad, [0,\chi_1^*]$ $\widetilde{y}_{\max}, (\chi_1^*, T]$ $\chi_1^* < \eta_1^*$	$0 \quad, [0,\eta_1^*]$ $\widetilde{z}_{\max}, (\eta_1^*, T]$ $\eta_1^* < \tau_1^*$
	$0 \quad, [0,\xi_1^*]$ $\widetilde{u}_{\max}, (\xi_1^*, T]$ $\xi_1^* < \tau_1^*$	$0 \quad, [0,\tau_1^*]$ $\widetilde{v}_{\max}, (\tau_1^*, T]$	$0 \quad, [0,\xi_1^*]$ $\widetilde{w}_{\max}, (\xi_1^*, T]$ $\xi_1^* < \tau_1^*$	$0 \quad, [0,\chi_1^*]$ $\widetilde{y}_{\max}, (\chi_1^*, T]$ $\chi_1^* < \eta_1^*$	$0 \quad, [0,\eta_1^*]$ $\widetilde{z}_{\max}, (\eta_1^*, T]$ $\eta_1^* < \tau_1^*$

Scenario 3 (S3). When function $L_v(t)$ has two zeros on the interval $(0, T)$, then

$$L_v(t) \begin{cases} < 0 & \text{, if} \quad 0 \le t < \tau_1^*, \\ = 0 & \text{, if} \quad t = \tau_1^*, \\ > 0 & \text{, if} \quad \tau_1^* < t < \tau_2^*, \\ = 0 & \text{, if} \quad t = \tau_2^*, \\ < 0 & \text{, if} \quad \tau_2^* < t \le T, \end{cases} \tag{5.7}$$

where $\tau_1^*, \tau_2^* \in (0, T)$, $\tau_1^* < \tau_2^*$, are zeros of function $L_v(t)$. Hence, by (3.5), the optimal control $v_*(t)$ is of the following type

$$v_*(t) = \begin{cases} v_{\min} & \text{, if} \quad 0 \le t \le \tau_1^*, \\ 1 & \text{, if} \quad \tau_1^* < t \le \tau_2^*, \\ v_{\min} & \text{, if} \quad \tau_2^* < t \le T. \end{cases}$$

From (5.2), (5.7) and (3.11), it can be seen that the behavior of switching functions $L_w(t)$, $L_y(t)$, $L_z(t)$ is determined as by the values of $L_w(0)$, $L_y(0)$, $L_z(0)$, as the values of $L_w(\tau_1^*)$, $L_y(\tau_1^*)$, $L_z(\tau_1^*)$. Likewise Scenario 2, we consider firstly possible cases of the behavior of function $L_w(t)$.

Case 3.1. If $L_w(\tau_1^*) \geq 0$, then, by (5.7) and the first equation of system (3.11), $L_w(t) > 0$ for all $t \in [0, \tau_1^*) \cup (\tau_1^*, T)$, and hence the inequality

$$L_w(t) > 0, \quad t \in [0, \tau_1^*) \tag{5.8}$$

holds as well. Furthermore, by (5.2), (5.7),

$$L_w(t) < L_y(t), \quad t \in [0, \tau_1^*). \tag{5.9}$$

Inequalities (5.8), (5.9) imply that $L_y(t) > 0$ for all $t \in [0, \tau_1^*)$, or, what is the same,

$$\Psi(t) = L_y(t) e^{\int_t^T y_*(r)dr} > 0, \quad t \in [0, \tau_1^*). \tag{5.10}$$

From the second equation of system (3.11) it follows that

$$\dot{\Psi}(t) = -\theta e^{\int_t^T y_*(r)dr} L_z(t). \tag{5.11}$$

Now, using (5.11), we can analyze the function $L_y(t)$. Depending on the sign of $L_z(\tau_1^*)$ there are three possible subcases.

Subcase 3.1.1. Let $L_z(\tau_1^*) \geq 0$. Then, from (5.7) and the third equation of system (3.11) it follows that $L_z(t) > 0$ for all $t \in [0, \tau_1^*) \cup (\tau_1^*, T]$. Hence, by (5.11), function $\Psi(t)$ decreases on interval $[0, T]$ from the value $\Psi(0)$ to the positive value $\Psi(T) = \alpha$. Therefore, for all $t \in [0, T]$, inequality $\Psi(t) > 0$ holds, and hence for inequality $L_y(t) > 0$ holds as well. Consequently, by (3.4), (3.6)–(3.8) and (3.10), in this subcase the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are of the same types as these in Scenario 1.

Subcase 3.1.2. Let $L_z(\tau_1^*) < 0$ and $L_z(0) \leq 0$. Then from (5.7) and the third equation of system (3.11) it follows that function $L_z(t)$ has one zero $\eta_1^* \in (0, T)$, such that $\tau_1^* < \eta_1^* < \tau_2^*$, and

$$L_z(t) \begin{cases} < 0 & , \text{if} \quad 0 < t < \eta_1^*, \\ = 0 & , \text{if} \quad t = \eta_1^*, \\ > 0 & , \text{if} \quad \eta_1^* < t \leq T. \end{cases} \tag{5.12}$$

From (5.11), (5.12) it follows that function $\Psi(t)$ increases from the value $\Psi(0)$ to maximum value $\Psi(\eta_1^*)$ and then decreases to the positive value $\Psi(T) = \alpha$. From inequality (5.10) it follows that $\Psi(0) > 0$, and hence $\Psi(t) > 0$ holds for all $t \in [0, T]$. It implies that $L_y(t) > 0$ for all $t \in [0, T]$ as well. Therefore, by (3.4), (3.6)–(3.8) and (3.10), we conclude that the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are of the following types:

$$u_*(t) = u_{\min}, \quad w_*(t) = w_{\max}, \quad y_*(t) = y_{\max}, \quad t \in [0, T];$$

$$z_*(t) = \begin{cases} z_{\min} & , \text{if} \quad 0 \leq t \leq \eta_1^*, \\ z_{\max} & , \text{if} \quad \eta_1^* < t \leq T, \end{cases}$$

where $\tau_1^* < \eta_1^* < \tau_2^*$.

CRM Preprint Series number 1219

Subcase 3.1.3. Let $L_z(\tau_1^*) < 0$ and $L_z(0) > 0$. Then, by (5.7) and the third equation of system (3.11), function $L_z(t)$ has two zeros $\eta_1^*, \eta_1^* \in (0,T)$, $\eta_1^* < \eta_2^*$, such that $\eta_1^* < \tau_1^* < \eta_2^* < \tau_2^*$, and

$$L_z(t) \begin{cases} > 0 & \text{, if} \quad 0 \le t < \eta_1^*, \\ = 0 & \text{, if} \quad t = \eta_1^*, \\ < 0 & \text{, if} \quad \eta_1^* < t < \eta_2^*, \\ = 0 & \text{, if} \quad t = \eta_2^*, \\ > 0 & \text{, if} \quad \eta_2^* < t \le T. \end{cases} \tag{5.13}$$

Relationships (5.11), (5.13) imply that function $\Psi(t)$ decreases from the value $\Psi(0)$ to the minimum value $\Psi(\eta_1^*)$ and then increases from $\Psi(\eta_1^*)$ to maximum value $\Psi(\eta_2^*)$. On subinterval $t \in (\eta_2^*, T)$ function $\Psi(t)$ again decreases from the value $\Psi(\eta_2^*)$ to the positive value $\Psi(T) = \alpha$. From (5.10) it is easy to see that $\Psi(\eta_1^*) > 0$, and hence $\Psi(t) > 0$ holds for all $t \in [0,T]$ ensuring that $L_y(t) > 0$ holds for the same values of t as well. Therefore, by (3.4), (3.6)–(3.8) and (3.10), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are of the following types:

$$u_*(t) = u_{\min}, \quad w_*(t) = w_{\max}, \quad y_*(t) = y_{\max}, \quad t \in [0,T];$$

$$z_*(t) = \begin{cases} z_{\max} & \text{, if} \quad 0 \le t \le \eta_1^*, \\ z_{\min} & \text{, if} \quad \eta_1^* < t \le \eta_2^*, \\ z_{\max} & \text{, if} \quad \eta_2^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^* < \eta_2^* < \tau_2^*$.

Case 3.2. If $L_w(\tau_1^*) < 0$ and $L_w(0) \le 0$, then, by (5.7) and the first equation of system (3.11), function $L_w(t)$ has one zero $\xi_1^* \in (0,T)$, such that $\tau_1^* < \xi_1^* < \tau_2^*$, and hence $L_w(t)$ satisfies (5.5). Furthermore, by (5.2), (5.7),

$$L_y(\tau_1^*) < 0. \tag{5.14}$$

Now, using the auxiliary function $\Psi(t)$ and equation (5.11), we can analyze the behavior of function $L_y(t)$. As above, there are three possible subcases.

Subcase 3.2.1. Let $L_z(\tau_1^*) \ge 0$. Then, as in Subcase 3.1.1, $L_z(t) > 0$ for all $t \in [0, \tau_1^*) \cup (\tau_1^*, T]$, and hence $L_y(t) > 0$ for all $t \in [0,T]$ as well. This inequality contradicts to (5.14), and hence this subcase is impossible.

Subcase 3.2.2. Let $L_z(\tau_1^*) < 0$ and $L_z(0) \le 0$. Then, as in Subcase 3.1.2, we immediately conclude that in this subcase function $L_z(t)$ satisfies (5.12). By this and (5.14), function $L_y(t)$ satisfies (5.4). Therefore, by (3.4), (3.6)–(3.8) and (3.10), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are of the same types as these in Subcase 2.2.2, where in this subcase $\tau_1^* < \eta_1^* < \tau_2^*$, $\tau_1^* < \xi_1^* < \tau_2^*$ and $\chi_1^* < \eta_1^*$.

Subcase 3.2.3. Let $L_z(\tau_1^*) < 0$ and $L_z(0) > 0$. Then, as in Subcase 3.1.3, function $L_z(t)$ satisfies (5.13). Therefore, taking into consideration (5.14) we conclude that function $L_y(t)$ is either of type (5.4), where $\eta_1^* < \chi_1^* < \eta_2^*$, or of type

$$L_y(t) \begin{cases} > 0 & \text{, if } \ 0 \le t < \chi_1^*, \\ = 0 & \text{, if } \ t = \chi_1^*, \\ < 0 & \text{, if } \ \chi_1^* < t < \chi_2^*, \\ = 0 & \text{, if } \ t = \chi_2^*, \\ > 0 & \text{, if } \ \chi_2^* < t \le T. \end{cases} \tag{5.15}$$

Here $\chi_1^*, \chi_2^* \in (0, T)$, $\chi_1^* < \chi_2^*$, are zeros of function $L_y(t)$ such that $\chi_1^* < \eta_1^* < \chi_2^* < \eta_2^*$. Therefore, by (3.4), (3.6)–(3.8) and (3.10), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ belong to one of the two following types: either

$$u_*(t) = \begin{cases} \mu & \text{, if } \ 0 \le t \le \xi_1^*, \\ u_{\min} & \text{, if } \ \xi_1^* < t \le T, \end{cases} \quad w_*(t) = \begin{cases} w_{\min} & \text{, if } \ 0 \le t \le \xi_1^*, \\ w_{\max} & \text{, if } \ \xi_1^* < t \le T, \end{cases}$$

$$y_*(t) = \begin{cases} y_{\min} & \text{, if } \ 0 \le t \le \chi_1^*, \\ y_{\max} & \text{, if } \ \chi_1^* < t \le T, \end{cases} \quad z_*(t) = \begin{cases} z_{\max} & \text{, if } \ 0 \le t \le \eta_1^*, \\ z_{\min} & \text{, if } \ \eta_1^* < t \le \eta_2^*, \\ z_{\max} & \text{, if } \ \eta_2^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^* < \eta_2^* < \tau_2^*$, $\tau_1^* < \xi_1^* < \tau_2^*$ and $\eta_1^* < \chi_1^* < \eta_2^*$; or

$$u_*(t) = \begin{cases} \mu & \text{, if } \ 0 \le t \le \xi_1^*, \\ u_{\min} & \text{, if } \ \xi_1^* < t \le T, \end{cases} \quad w_*(t) = \begin{cases} w_{\min} & \text{, if } \ 0 \le t \le \xi_1^*, \\ w_{\max} & \text{, if } \ \xi_1^* < t \le T, \end{cases}$$

$$y_*(t) = \begin{cases} y_{\max} & \text{, if } \ 0 \le t \le \chi_1^*, \\ y_{\min} & \text{, if } \ \chi_1^* < t \le \chi_2^*, \\ y_{\max} & \text{, if } \ \chi_2^* < t \le T, \end{cases} \quad z_*(t) = \begin{cases} z_{\max} & \text{, if } \ 0 \le t \le \eta_1^*, \\ z_{\min} & \text{, if } \ \eta_1^* < t \le \eta_2^*, \\ z_{\max} & \text{, if } \ \eta_2^* < t \le T, \end{cases}$$

where $\eta_1^* < \tau_1^* < \eta_2^* < \tau_2^*$, $\tau_1^* < \xi_1^* < \tau_2^*$ and $\chi_1^* < \eta_1^* < \chi_2^* < \eta_2^*$.

Case 3.3. If $L_w(\tau_1^*) < 0$ and $L_w(0) > 0$, then, by (5.7) and the first equation of system (3.11), function $L_w(t)$ has two zeros $\xi_1^*, \xi_2^* \in (0, T)$, $\xi_1^* < \xi_2^*$, such that $\xi_1^* < \tau_1^* < \xi_2^* < \tau_2^*$, and

$$L_w(t) \begin{cases} > 0 & \text{, if } \ 0 \le t < \xi_1^*, \\ = 0 & \text{, if } \ t = \xi_1^*, \\ < 0 & \text{, if } \ \xi_1^* < t < \xi_2^*, \\ = 0 & \text{, if } \ t = \xi_2^*, \ t = T, \\ > 0 & \text{, if } \ \xi_2^* < t < T. \end{cases}$$

Furthermore, by (5.2), (5.7), (5.14) holds in this case, and

$$L_y(0) > 0. \tag{5.16}$$

As above, we analyze the behavior of auxiliary function $\Psi(t)$ depending on the sign of function $L_z(\tau_1^*)$.

Subcase 3.3.1. Let $L_z(\tau_1^*) \geq 0$. Then, as in Subcase 3.1.1, we see that $L_z(t) > 0$ holds for all $t \in [0, \tau_1^*) \cup (\tau_1^*, T]$, and hence $L_y(t) > 0$ holds for all $t \in [0, T]$ as well. This inequality contradicts (5.14), and hence this subcase is impossible.

Subcase 3.3.2. Let $L_z(\tau_1^*) < 0$ and $L_z(0) \leq 0$. Then, as in Subcase 3.1.2, function $L_z(t)$ satisfies (5.12), and, by (5.11), function $\Psi(t)$ increases from the value $\Psi(0)$ to the maximum value $\Psi(\eta_1^*)$ and then decreases to the positive value $\Psi(T)$. Inequality (5.14), or, what is the same, $\Psi(\tau_1^*) < 0$, implies that for all $t \in [0, \tau_1^*]$ inequality $\Psi(t) < 0$ holds. This contradicts (5.16), and hence this subcase is impossible as well.

Subcase 3.3.3. Let $L_z(\tau_1^*) < 0$ and $L_z(0) > 0$. Then, as in Subcase 3.1.3, function $L_z(t)$ satisfies (5.13), and hence, taking into consideration (5.14) and (5.16), we immediately see that function $L_y(t)$ satisfies (5.15). Therefore, by (3.4), (3.6)–(3.8) and (3.10), the optimal controls $u_*(t)$, $w_*(t)$, $y_*(t)$, $z_*(t)$ are of the following types:

$$u_*(t) = \begin{cases} u_{\min} & , \text{ if } \ 0 \leq t \leq \xi_1^*, \\ \mu & , \text{ if } \ \xi_1^* \leq t \leq \xi_2^*, \\ u_{\min} & , \text{ if } \ \xi_2^* < t \leq T, \end{cases} \quad w_*(t) = \begin{cases} w_{\max} & , \text{ if } \ 0 \leq t \leq \xi_1^*, \\ w_{\min} & , \text{ if } \ \xi_1^* \leq t \leq \xi_2^*, \\ w_{\max} & , \text{ if } \ \xi_2^* < t \leq T, \end{cases}$$

$$y_*(t) = \begin{cases} y_{\max} & , \text{ if } \ 0 \leq t \leq \chi_1^*, \\ y_{\min} & , \text{ if } \ \chi_1^* < t \leq \chi_2^*, \\ y_{\max} & , \text{ if } \ \chi_2^* < t \leq T, \end{cases} \quad z_*(t) = \begin{cases} z_{\max} & , \text{ if } \ 0 \leq t \leq \eta_1^*, \\ z_{\min} & , \text{ if } \ \eta_1^* < t \leq \eta_2^*, \\ z_{\max} & , \text{ if } \ \eta_2^* < t \leq T, \end{cases}$$

where $\eta_1^* < \tau_1^* < \eta_2^* < \tau_2^*$, $\xi_1^* < \tau_1^* < \xi_2^* < \tau_2^*$ and $\chi_1^* < \eta_1^* < \chi_2^* < \eta_2^*$.

The types of physical controls $\widetilde{u}_*(t)$, $\widetilde{v}_*(t)$, $\widetilde{w}_*(t)$, $\widetilde{y}_*(t)$ and $\widetilde{z}_*(t)$ corresponding the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ by formulas (2.5) for Scenario 3 are summarized in Table 2.

6. Discussion and conclusion

The goal of this paper is to explore how a nonlinearity of functional responses in a control model affects the optimal controls. Having this objective in mind, we considered a SEIR epidemic model where the incidence rate is given by a unspecified nonlinear function $f(S, I)$ constrained by a few biologically feasible conditions. A conclusion that follows from our analysis is that for the considered model with a nonlinear incidence rate satisfying assumptions (A1–A5) principal qualitative properties of the optimal controls, such as the maximal number of switchings of the controls and an order of these switchings, do not depend on a particular form of the nonlinearity. Moreover, the analysis shows that the values of derivatives $f'_S(S, I)$ and $f'_I(S, I)$ have a stronger impact on the qualitative properties of the optimal controls than a form of the nonlinearity.

TABLE 2. Variants of physical controls for Scenario 3.

#	$\widetilde{u}_*(t)$	$\widetilde{v}_*(t)$	$\widetilde{w}_*(t)$	$\widetilde{y}_*(t)$	$\widetilde{z}_*(t)$
S3:	$\widetilde{u}_{\max}$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$\widetilde{z}_{\max}$
	$\widetilde{u}_{\max}$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$0 \quad, [0, \eta_1^*]$ $\widetilde{z}_{\max}, (\eta_1^*, T]$ $\tau_1^* < \eta_1^* < \tau_2^*$
	$\widetilde{u}_{\max}$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$\widetilde{w}_{\max}$	$\widetilde{y}_{\max}$	$\widetilde{z}_{\max}, [0, \eta_1^*]$ $0 \quad, (\eta_1^*, \eta_2^*]$ $\widetilde{z}_{\max}, (\eta_2^*, T]$ $\eta_1^* < \tau_1^*$ $< \eta_2^* < \tau_2^*$
	$0 \quad, [0, \xi_1^*]$ $\widetilde{u}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$0 \quad, [0, \xi_1^*]$ $\widetilde{w}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$0 \quad, [0, \chi_1^*]$ $\widetilde{y}_{\max}, (\chi_1^*, T]$ $\chi_1^* < \eta_1^*$	$0 \quad, [0, \eta_1^*]$ $\widetilde{z}_{\max}, (\eta_1^*, T]$ $\tau_1^* < \eta_1^* < \tau_2^*$
	$0 \quad, [0, \xi_1^*]$ $\widetilde{u}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$0 \quad, [0, \xi_1^*]$ $\widetilde{w}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$0 \quad, [0, \chi_1^*]$ $\widetilde{y}_{\max}, (\chi_1^*, T]$ $\eta_1^* < \chi_1^* < \eta_2^*$	$\widetilde{z}_{\max}, [0, \eta_1^*]$ $0 \quad, (\eta_1^*, \eta_2^*]$ $\widetilde{z}_{\max}, (\eta_2^*, T]$ $\eta_1^* < \tau_1^*$ $< \eta_2^* < \tau_2^*$
	$0 \quad, [0, \xi_1^*]$ $\widetilde{u}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$0 \quad, [0, \xi_1^*]$ $\widetilde{w}_{\max}, (\xi_1^*, T]$ $\tau_1^* < \xi_1^* < \tau_2^*$	$\widetilde{y}_{\max}, [0, \chi_1^*]$ $0 \quad, (\chi_1^*, \chi_2^*]$ $\widetilde{y}_{\max}, (\chi_2^*, T]$ $\chi_1^* < \eta_1^*$ $< \chi_2^* < \eta_2^*$	$\widetilde{z}_{\max}, [0, \eta_1^*]$ $0 \quad, (\eta_1^*, \eta_2^*]$ $\widetilde{z}_{\max}, (\eta_2^*, T]$ $\eta_1^* < \tau_1^*$ $< \eta_2^* < \tau_2^*$
	$\widetilde{u}_{\max}, [0, \xi_1^*]$ $0 \quad, (\xi_1^*, \xi_2^*]$ $\widetilde{u}_{\max}, (\xi_2^*, T]$ $\xi_1^* < \tau_1^*$ $< \xi_2^* < \tau_2^*$	$\widetilde{v}_{\max}, [0, \tau_1^*]$ $0 \quad, (\tau_1^*, \tau_2^*]$ $\widetilde{v}_{\max}, (\tau_2^*, T]$	$\widetilde{w}_{\max}, [0, \xi_1^*]$ $0 \quad, (\xi_1^*, \xi_2^*]$ $\widetilde{w}_{\max}, (\xi_2^*, T]$ $\xi_1^* < \tau_1^*$ $< \xi_2^* < \tau_2^*$	$\widetilde{y}_{\max}, [0, \chi_1^*]$ $0 \quad, (\chi_1^*, \chi_2^*]$ $\widetilde{y}_{\max}, (\chi_2^*, T]$ $\chi_1^* < \eta_1^*$ $< \chi_2^* < \eta_2^*$	$\widetilde{z}_{\max}, [0, \eta_1^*]$ $0 \quad, (\eta_1^*, \eta_2^*]$ $\widetilde{z}_{\max}, (\eta_2^*, T]$ $\eta_1^* < \tau_1^*$ $< \eta_2^* < \tau_2^*$

These results were obtained for the SEIR model and for the objective functional of minimizing the level of infection at the end of a given time interval. However, the analysis in this paper, as well as results of [3] and [19], indicates that the same conclusion likely holds for more complex models and for objective functionals independent from the controls. Moreover, on the basis of the analysis we conject that nonlinearities of other functional responses in the model would not affect the principal properties of the optimal controls either, provided that the signs of the derivatives of these responses are preserved.

We have to stress that the conclusion that the principal properties of the optimal controls do not depend on a form of the nonlinearity of functional responses of a model holds only for the functional which are independent of the controls themselves. For objective functionals which explicitly depend on the controls this conclusion may be incorrect. By virtue of the Pontryagin maximum principle, for the functionals which depend on control nonlinearly the optimal controls can be found as explicit functions of the phase and adjoint variables. Some of the controls can explicitly depend on the functional responses, whereas the others will be affected by the non-linearity of these implicitly via the phase and adjoint variables. To illustrate this idea, in Appendix A we found the form optimal controls for the SEIR control model (2.6) combined with the most common objective functional depending on a sum of the squares of the controls.

For the considered control problems, we found all variants of the optimal controls that can happen in this problem. On the basis of the analysis, we conject that the implementation of a specific variant depends on specific initial conditions $S(0), E(0), I(0)$ and the maximal values of derivatives $f'_S(S,I)$ and $f'_I(S,I)$ in the feasible region. For a specific optimal control problem, knowing a number of switchings, or having accurate estimations for these, allows a straightforward reduction of the problem to a considerably simpler problem of finite-dimensional optimization. The latter problem can be solved numerically using any of existing solvers. We did not do this in this paper, because our goal was to explore the influence of the nonlinearity on the controls rather then consider a particular case.

The constraints (A1–A5), which we imposed in this paper on incidence rate $f(S,I)$, hold for any realistic infectious disease. Moreover, these constraints are considerably less restrictive than, for instance, the conditions for the uniqueness of a positive equilibrium state and the global asymptotic stability of the system [27, 30]. It may be worthy of mentioning that if derivatives $f'_I(S,I)$ or $f'_S(S,I)$ can be equal to zero in the region Ω (that is, if (A3) does not hold), than one or several of the switching functions can be identically equal to zero on a subinterval of $[0,T]$. This, in turn, implies that the corresponding optimal controls cannot be precisely found from (3.4)–(3.8), and singularities can arise on this subinterval [38].

We also use constraints (2.3) in our analysis. The only reason for introducing these constraints is convenience of the analysis. As we mentioned earlier, these inequalities naturally hold in practically relevant cases. Despite this, the proof of the fact that obtained in Section 5 results are valid, when restrictions (2.3) are violated, is given in Appendix B.

Apart from assumptions (A1–A5) and constraints (2.3), in this paper we also use the assumption of the constant population case. While the use of this assumption, allowing to reduce the order of a model by one, is justified by the significant difference of the timescales of demographic and epidemic processes and therefore is common in epidemic modeling, it still might be expected that this could affect the controls. Results in [15] and [18] for SIR control models with and without this assumption indicate that for long-lasting epidemics which inflict significant variation of a population this might be possible.

The objective functional that we consider in this paper, namely a weighted sum of the exposed and infectious individuals at the terminal moment of time T, corresponds to the objective of elimination of the infection in the population. It is implicitly assumed the the level of infection at moment T is below the survival level of this infection. An alternative possibility is a policy aimed at the protection of a population. The latter task is common when the elimination is unfeasible (e.g., as in the case of a influenza epidemic) and corresponds to a control problem of minimizing the cumulative level of infected (that is, the sum of exposed and infectious individuals) over a given time interval. A combination of both objectives is also possible.

A practically relevant conclusion that follows from our analysis is that the vaccination of the newborns and the susceptibles (which in this paper are defined by different controls) should be started, carried out and ended simultaneously. That is, vaccinating these two groups should be considered as one control. It can be conjected that this conclusion is also valid for other epidemic models and allows to simplify formulation and analysis of a problem of infectious disease control for these models. Another practically relevant conclusion that follows from our analysis is that the two major types of control policies, namely vaccination of the susceptible individuals and the screening and isolation of the infected, are connected only via indirect control policy (such as education). Since the latest is the cheapest of these three policies and in the same time has the largest inertia (it is unlikely that educated individuals would change their behavior immediately at the moment when the education is stopped), it can be expected that the effect of this policy would be constant, or nearly constant, throughout an epidemic. In such a case, two other major control policies, namely the vaccination (or other ways of protecting the susceptibles) and the screening/isolation of the infected, are completely independent. This fact also allows to considerably simplify the formulation of a control problem and its analysis.

An intriguing feature of the optimal controls for the SEIR model that we found is that in the SEIR control problem all the considered controls can have by one switching more than the same controls in the SIR model considered in [15, 18, 19]. The reason of this disparity, as well as the conditions for these extra switchings to happen in the SEIR model, are unclear. Our conjecture is that this variant can happen for initial conditions where $E'(0)$ and $I'(0)$ are of different signs.

It may be noteworthy that the arguments that were used in the proof were earlier used for the estimation of number of zeros of switching functions for a SIR control model in [15] and for a model of a biochemical process in [17].

Acknowledgements. E.N. Khailov is supported by the Russian Science Foundation, project 14-11-00539. A. Korobeinikov is supported by the Ministry of Science and Innovation of Spain via Ramón y Cajal Fellowship RYC-2011-08061, by CONACYT (Mexico) via grant N219614, and by AGAUR (Generalitat de Catalunya) via grant 2014SGR-1307.

References

[1] B.M. Adams, H.T. Banks, H.D. Kwon, and H.T. Tran, *Dynamic multidrug therapies for HIV: optimal and STI control approaches.* Math. Biosci. Eng. 1 (2004) no. 2, 223–241.

[2] S. Anita, V. Arnaŭtu, and V. Capasso. Introduction to Optimal Control Problems in Life Sciences and Economics. Birkhäuser, USA, 2011.

[3] H. Behncke, *Optimal control of deterministic epidemics.* Optim. Control Appl. Meth. 21 (2000), 269–285.

[4] M.H.A. Biswas, L.T. Paiva, and M.d.R. de Pinho, *A SEIR model for control of infectious diseases with constraints.* Math. Biosci. Eng. 11 (2014), no. 4, 761–784.

[5] S. Butler, D. Kirschner, and S. Lenhart, Optimal Control of the Chemotherapy Affecting the Infectivity of HIV, in Advances in Mathematical Population Dynamics – Molecules, Cells and Man, 6, (eds. O. Arino, D. Axelrod, and M. Kimmel), World Scientific, Singapore, 1997, 557–569.

[6] R.V. Culshaw, S. Ruan, and R.J. Spiteri, *Optimal HIV treatment by maximising immune response.* J. Math. Biol. 48 (2004), no. 5, 545–562.

[7] B.P. Demidovich, Lectures on Stability Theory, Nauka, Moscow, 1967.

[8] A.V. Dmitruk, *A generalized estimate on the number of zeros for solutions of a class of linear differential equations.* SIAM Journal on Control and Optimization 30 (1992), no. 5, 1087–1091.

[9] K.R. Fister, S. Lenhart, and J.S. McNally, *Optimizing chemotherapy in an HIV model.* Electronic Journal of Differential Equations 1998 (1998), no. 32, 1–12.

[10] H. Gaff and E. Schaefer, *Optimal control applied to vaccination and treatment strategies for various epidemiological models.* Math. Biosci. Eng. 6 (2009), no. 3, 469–492.

[11] E.V. Grigorieva, N.V. Bondarenko, E.N. Khailov, and A. Korobeinikov, *Three-dimensional nonlinear optimal control model of wastewater biotreatment.* Neural, Parallel, and Scientific Computations 20 (2012), no. 1, 23–36.`http://www.dynamicpublishers.com/Neural/npsccontent.htm`

[12] E. Grigorieva, N. Bondarenko, E. Khailov, and A. Korobeinikov, Finite-Dimensional Methods for Optimal Control of Autothermal Thermophilic Aerobic Digestion, in Industrial Waste, (eds. K.-Y. Show and X. Guo), InTech, Croatia, 2012, 91–120. (ISBN:978-953-51-0253-3) `http://www.intechopen.com/articles/show/title/nite-dimensional-methods-for-optimal-controlof-waste-water-process`

[13] E.V. Grigorieva, N.V. Bondarenko, E.N. Khailov, and A. Korobeinikov, *Analysis of optimal control problems for the process of wastewater biological treatment.* Revista de Matemática: Teoría y Aplicaciones 20 (2013), no. 2, 103–118.(ISSN:14092433)

[14] E.V. Grigorieva, N.V. Bondarenko, E.N. Khailov, and A. Korobeinikov, *Modeling and optimal control for antiretroviral therapy*. Journal of Biological Systems 22 (2014), no. 2, 199–217. doi:10.1142/S0218339014400026

[15] E.V. Grigorieva, E.N. Khailov, and A. Korobeinikov, *Optimal control for a SIR infectious disease model*. J. Coupled Syst. Multiscale Dyn. 1 (2013), no. 3, 324–331. doi:10.1166/jcsmd.2013.1022

[16] E.V. Grigorieva, E.N. Khailov, and A. Korobeinikov, *An optimal control problem in HIV treatment*. Discret. Contin. Dyn. Syst., supplement (2013), 311–322. http://aimsciences.org/journals/pdfs.jsp?paperID=9216\&mode=full http://aimsciences.org/journals/contentsListPro.jsp?pubID=640

[17] E.V. Grigorieva, E.N. Khailov, and A. Korobeinikov, *Parametrization of the attainable set for a nonlinear control model of a biochemical process*. Math. Biosci. Eng. 10 (2013), no. 4, 1067–1094. doi:10.3934/mbe.2013.10.1067

[18] E.V. Grigorieva, E.N. Khailov, and A. Korobeinikov, *Optimal control for an epidemic in a population of varying size*. Discret. Contin. Dyn. Syst., supplement (2015), 549–561.

[19] E.V. Grigorieva, E.N. Khailov, and A. Korobeinikov, *Optimal control for a SIR epidemic model with nonlinear incidence rate*. Math. Model. Nat. Phenom. (2015) (to appear).

[20] P. Hartman, Ordinary Differential Equations, John Wiley & Sons, NY, 1964.

[21] H.R. Joshi, *Optimal control of an HIV immunology model*. Optim. Control Appl. Meth. 23 (2002), 199–213.

[22] H.R. Joshi, S. Lenhart, M.Y. Li, and L. Wang, Optimal Control Methods Applied to Desease Models, in AMS Volume on Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challenges, AMS Contemporary Mathematical Series, vol. 410, 2006, 187–207.

[23] D. Kirschner, S. Lenhart, and S. Serbin, *Optimal control of the chemotherapy of HIV*. J. Math. Biol. 23 (1997), 199–213.

[24] D. Kirschner, S. Lenhart, and S. Serbin, *Optimizing chemotherapy of HIV infection: scheduling, ammounts and initiation of treatment*. J. Math. Biol. 35 (1997), 775–792.

[25] A. Kolmogorov, *Sulla teoria di Volterra della lotta per l'esistenza*. Gi. Inst. Ital. Attuari 7 (1936), 74–80.

[26] A. Korobeinikov, *Lyapunov functions and global stability for SIR and SIRS epidemiological models with non-linear transmission*. Bull. Math. Biol. 68 (2006), no. 3, 615–626.

[27] A. Korobeinikov, *Global properties of infectious disease models with non-linear incidence*. Bull. Math. Biol. 69 (2007), no. 6, 1871–1886.

[28] A. Korobeinikov, *Global asymptotic properties of virus dynamics models with dose dependent parasite reproduction and virulence, and nonlinear incidence rate*. Math. Med. Biol. 26 (2009), no. 3, 225–239. http://imammb.oxfordjournals.org/cgi/reprint/dqp006

[29] A. Korobeinikov, *Stability of ecosystem: global properties of a general predator-prey model*. Math. Med. Biol. 26 (2009), 309–321. http://imammb.oxfordjournals.org/cgi/reprint/dqp009

[30] A. Korobeinikov and P.K. Maini, *Nonlinear incidence and stability of infectious disease models*. Math. Med. Biol. 22 (2005), no. 2, 113–128.

[31] A. Korobeinikov and S.V. Petrovskii, Toward a general theory of ecosystem stability: plankton–nutrient interaction as a paradigm, in Aspects of Mathematical Modelling, (eds. R.J. Hosking and E. Venturino), Birkhäuser, Basel, 2008, 27–40.

[32] J.J. Kutch and P. Gurfil, Optimal control of HIV infection with a continuously-mutating viral population, in Proceedings of American Control Conference, Anchorage, Alaska, 2002, 4033–4038.

[33] U. Ledzewicz and H. Schättler, *On optimal singular controls for a general SIR-model with vaccination and treatment*. Discret. Contin. Dyn. Syst., supplement (2011), 981–990.

[34] E.B. Lee and L. Marcus, Foundations of Optimal Control Theory, John Wiley & Sons, New York, 1967.
[35] S. Lenhart and J.T. Workman, Optimal Control Applied to Biological Models, CRC Press, Taylor & Francis, London, 2007.
[36] M.S. Nikol'skii, *On the convergence of optimal controls in some optimization problems.* Moscow University Computational Mathematics and Cybernetics 1 (2004), 24–31.
[37] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Mishchenko, Mathematical Theory of Optimal Processes, John Wiley & Sons, New York, 1962.
[38] H. Schättler and U. Ledzewicz, Geometric Optimal Control: Theory, Methods and Examples, Springer, New York-Heidelberg-Dordrecht-London, 2012.
[39] H. Schättler and U. Ledzewicz, Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods, Springer, New York-Heidelberg-Dordrecht-London, 2015.
[40] A.N. Tikhonov, A.B. Vasil'eva, and A.G. Sveshnikov, Differential Equations, Springer, Berlin, 1985.
[41] F.P. Vasil'ev, Optimization Methods, Factorial Press, Moscow, 2002.

Appendix A

In order to demonstrate that for objective functionals which depend on the controls, the optimal controls depend on a form of the nonlinearity of functional responses, let us consider objective functional

$$\widetilde{J}(u,v,w,y,z) = J(u,v,w,y,z) + 0.5\kappa \int_0^T \Big((1-\mu^{-1}u(t))^2 + v^2(t)$$
$$+((w(t)-\mu)^2 + (y(t)-(\theta+\mu))^2 + (z(t)-(\sigma+\mu))^2\Big)dt$$

with the control model (2.6). Here $J(u,v,w,y,z)$ is the functional defined by (2.9), and under the integral is a sum of the squares of the physical controls $\widetilde{u}(t)$, $\widetilde{v}(t)$, $\widetilde{w}(t)$, $\widetilde{y}(t)$ and $\widetilde{z}(t)$. For this objective functional, the Hamiltonian is

$$\widetilde{H}(S,E,I,u,v,w,y,z,\psi_1,\psi_2,\psi_3) = H(S,E,I,u,v,w,y,z,\psi_1,\psi_2,\psi_3)$$
$$-0.5\kappa\Big((1-\mu^{-1}u)^2 + v^2 + (w-\mu)^2 + (y-(\theta+\mu))^2 + (z-(\sigma+\mu))^2\Big),$$

where $H(S,E,I,u,v,w,y,z,\psi_1,\psi_2,\psi_3)$ is defined by (3.1). The new terms in the Hamiltonian do not include the phase variables, and therefore the adjoint variables for this control problem satisfy system (3.3). The difference is only the derivatives

$$\widetilde{H}'_u = \psi_1 + \kappa\mu^{-1}(1-\mu^{-1}u), \quad \widetilde{H}'_v = f(S,I)(\psi_2-\psi_1) - \kappa v,$$
$$\widetilde{H}'_w = -S\psi_1 - \kappa(w-\mu), \qquad \widetilde{H}'_y = -E\psi_2 - \kappa(y-(\theta+\mu)),$$
$$\widetilde{H}'_z = -I\psi_3 - \kappa(z-(\sigma+\mu)).$$

There is no sense of using switching functions for this control problem.

Hamiltonian $\widetilde{H}(S, E, I, u, v, w, y, z, \psi_1, \psi_2, \psi_3)$ is concave (the inverted parabola) with respect to the controls, and therefore its maximum with respect to controls depends on whether the maximum of the parabola is within the ranges of the controls or is outside of these. Moreover, for this control problem the optimal controls are not bang-bang controls, but are continuous functions satisfying the following equalities:

$$u_*(t) = \begin{cases} u_{\min} & \text{, if} \quad \Gamma_u(t) < u_{\min}, \\ \Gamma_u(t) & \text{, if} \quad \Gamma_u(t) \in [u_{\min}, \mu], \\ \mu & \text{, if} \quad \Gamma_u(t) > \mu, \end{cases}$$

$$v_*(t) = \begin{cases} v_{\min} & \text{, if} \quad \Gamma_v(t) < v_{\min}, \\ \Gamma_v(t) & \text{, if} \quad \Gamma_v(t) \in [v_{\min}, 1], \\ 1 & \text{, if} \quad \Gamma_v(t) > 1, \end{cases}$$

$$w_*(t) = \begin{cases} w_{\min} & \text{, if} \quad \Gamma_w(t) < w_{\min}, \\ \Gamma_w(t) & \text{, if} \quad \Gamma_w(t) \in [w_{\min}, w_{\max}], \\ w_{\max} & \text{, if} \quad \Gamma_w(t) > w_{\max}, \end{cases}$$

$$y_*(t) = \begin{cases} y_{\min} & \text{, if} \quad \Gamma_y(t) < y_{\min}, \\ \Gamma_y(t) & \text{, if} \quad \Gamma_y(t) \in [y_{\min}, y_{\max}], \\ y_{\max} & \text{, if} \quad \Gamma_y(t) > y_{\max}, \end{cases}$$

$$z_*(t) = \begin{cases} z_{\min} & \text{, if} \quad \Gamma_z(t) < z_{\min}, \\ \Gamma_z(t) & \text{, if} \quad \Gamma_z(t) \in [z_{\min}, z_{\max}], \\ z_{\max} & \text{, if} \quad \Gamma_z(t) > z_{\max}, \end{cases}$$

where

$$\Gamma_u(t) = \mu + \mu^2\kappa^{-1}\psi_1(t), \quad \Gamma_v(t) = \kappa^{-1}f(S(t), I(t))(\psi_2(t) - \psi_1(t)),$$
$$\Gamma_w(t) = \mu - \kappa^{-1}S(t)\psi_1(t), \quad \Gamma_y(t) = (\theta + \mu) - \kappa^{-1}E(t)\psi_2(t),$$
$$\Gamma_z(t) = (\sigma + \mu) - \kappa^{-1}I(t)\psi_3(t).$$

(The optimal controls can be also expressed in terms of the minimum and maximum of two functions; see [4, 9, 10, 21, 22].) It can be immediately seen that the optimal control $v_*(t)$ explicitly depends on function $f(S, I)$. Furthermore, the other controls explicitly depend on the phase and adjoint variables. Hence they will be also affected by a particular form of the nonlinearity of function $f(S, I)$, or a non-linearity of other functional responses.

Appendix B

Here we show that it is possible to abandon the inequalities (2.3). At the same time, the results of Section 5 about the types of the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$, and thus the types of physical controls $\widetilde{u}_*(t)$, $\widetilde{v}_*(t)$, $\widetilde{w}_*(t)$, $\widetilde{y}_*(t)$ and $\widetilde{z}_*(t)$, remain valid.

Firstly, using formulas (2.5), (2.8) we rewrite the constraints (2.3) for the controls $w(t)$, $y(t)$, $z(t)$ as

$$w_{\max} \neq z_{\min}, \quad w_{\max} \neq z_{\max}, \quad w_{\max} \neq y_{\max}. \tag{B.1}$$

Now, let us write the system (2.6) in a matrix form. For this, we consider $x = (S, E, I)^\top$ as a vector of the phase variables and $\phi = (u, v, w, y, z)^\top$ as a vector of the controls. Then, we introduce the constant matrix Ω_0 as

$$\Omega_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \theta & 0 \end{pmatrix},$$

and the matrix function $\Pi(x)$ by the following formula:

$$\Pi(x) = \begin{pmatrix} 1 & -f(S,I) & -S & 0 & 0 \\ 0 & f(S,I) & 0 & -E & 0 \\ 0 & 0 & 0 & 0 & -I \end{pmatrix}.$$

Then, the system (2.6) can be rewritten as

$$\begin{cases} \dot{x}(t) = \Omega_0 x(t) + \Pi(x(t))\phi(t), \quad t \in [0, T], \\ x(0) = x_0 = (S_0, E_0, I_0)^\top. \end{cases} \tag{B.2}$$

According to (B.1) we consider three situations that were not able to explore previously. Namely, let one of the following equalities hold:

$$w_{\max} = z_{\min}, \quad w_{\max} = z_{\max}, \quad w_{\max} = y_{\max}, \tag{B.3}$$

and we use the perturbation method. Let $\varepsilon \in [0, \varepsilon_0]$ be a small perturbation parameter, where the positive value ε_0 is chosen such that for all values $\varepsilon \in (0, \varepsilon_0]$ the inequalities:

$$w^\varepsilon_{\max} \neq z_{\min}, \quad w^\varepsilon_{\max} \neq z_{\max}, \quad w^\varepsilon_{\max} \neq y_{\max}$$

hold, and for $\varepsilon = 0$ it would be valid that inequality from (B.1), which we assume to be satisfied. Here, $w^\varepsilon_{\max} = w_{\max} + \varepsilon$. Then, for all $\varepsilon \in (0, \varepsilon_0]$ we are in the situations, described in Section 5, for which the types of the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ are also known.

Thus, the object of the perturbation is the control region $[w_{\min}, w_{\max}]$ of the control $w(t)$. Instead of this region, we consider the perturbed region $[w_{\min}, w^\varepsilon_{\max}]$, and the perturbed set of admissible controls $D_\varepsilon(T)$ occurs. Moreover, for each value of $\varepsilon \in (0, \varepsilon_0]$ the optimal solutions:

$$x_*^\varepsilon(t) = (S_*^\varepsilon(t), E_*^\varepsilon(t), I_*^\varepsilon(t))^\top, \quad \phi_*^\varepsilon(t) = (u_*^\varepsilon(t), v_*^\varepsilon(t), w_*^\varepsilon(t), y_*^\varepsilon(t), z_*^\varepsilon(t))^\top$$

of the corresponding optimal control problem (2.6), (2.9) exist by the same arguments that were used in Section 2 for $\varepsilon = 0$. For the optimal value of the perturbed functional (2.9) we have equalities:

$$\min_{(u(\cdot),v(\cdot),w(\cdot),y(\cdot),z(\cdot))\in D_\varepsilon(T)} J_\varepsilon(u(\cdot), v(\cdot), w(\cdot), y(\cdot), z(\cdot)) = J_\varepsilon(u_*^\varepsilon(\cdot), v_*^\varepsilon(\cdot), w_*^\varepsilon(\cdot), y_*^\varepsilon(\cdot), z_*^\varepsilon(\cdot)) = J_\varepsilon^*.$$

The optimal controls $u_*^\varepsilon(t)$, $v_*^\varepsilon(t)$, $w_*^\varepsilon(t)$, $y_*^\varepsilon(t)$ and $z_*^\varepsilon(t)$ have the following types:

$$u_*^\varepsilon(t), w_*^\varepsilon(t) = \begin{cases} u_{\min}, w_{\max}^\varepsilon & , \text{ if } \quad 0 \le t \le \xi_{1,\varepsilon}^*, \\ \mu \quad\ , w_{\min} & , \text{ if } \quad \xi_{1,\varepsilon}^* \le t \le \xi_{2,\varepsilon}^*, \\ u_{\min}, w_{\max}^\varepsilon & , \text{ if } \quad \xi_{2,\varepsilon}^* < t \le T, \end{cases}$$

(B.4)
$$v_*^\varepsilon(t) = \begin{cases} v_{\min} & , \text{ if } \quad 0 \le t \le \tau_{1,\varepsilon}^*, \\ 1 & , \text{ if } \quad \tau_{1,\varepsilon}^* < t \le \tau_{2,\varepsilon}^*, \\ v_{\min} & , \text{ if } \quad \tau_{2,\varepsilon}^* < t \le T, \end{cases}$$

$$y_*^\varepsilon(t) = \begin{cases} y_{\max} & , \text{ if } \quad 0 \le t \le \chi_{1,\varepsilon}^*, \\ y_{\min} & , \text{ if } \quad \chi_{1,\varepsilon}^* < t \le \chi_{2,\varepsilon}^*, \\ y_{\max} & , \text{ if } \quad \chi_{2,\varepsilon}^* < t \le T, \end{cases} \quad z_*^\varepsilon(t) = \begin{cases} z_{\max} & , \text{ if } \quad 0 \le t \le \eta_{1,\varepsilon}^*, \\ z_{\min} & , \text{ if } \quad \eta_{1,\varepsilon}^* < t \le \eta_{2,\varepsilon}^*, \\ z_{\max} & , \text{ if } \quad \eta_{2,\varepsilon}^* < t \le T, \end{cases}$$

where $\xi_{j,\varepsilon}^*, \tau_{j,\varepsilon}^*, \chi_{j,\varepsilon}^*, \eta_{j,\varepsilon}^* \in [0, T)$, $j = 1, 2$ are the corresponding moments of switching.

In some situations the optimal controls $(u_*^\varepsilon(t), v_*^\varepsilon(t), w_*^\varepsilon(t), y_*^\varepsilon(t), z_*^\varepsilon(t))$ have less than two switchings on the interval $(0, T)$, or do not have them at all. Such situations were described in detail in Section 2. In all such situations, without loss of generality, we assume that on the interval $[0, T)$ the optimal controls $(u_*^\varepsilon(t), v_*^\varepsilon(t), w_*^\varepsilon(t), y_*^\varepsilon(t), z_*^\varepsilon(t))$ have two corresponding switchings $(\xi_{1,\varepsilon}^*, \xi_{2,\varepsilon}^*)$, $(\tau_{1,\varepsilon}^*, \tau_{2,\varepsilon}^*)$, $(\chi_{1,\varepsilon}^*, \chi_{2,\varepsilon}^*)$, $(\eta_{1,\varepsilon}^*, \eta_{2,\varepsilon}^*)$ and some of which are located in $t = 0$. Then, every optimal controls $(u_*^\varepsilon(t), v_*^\varepsilon(t), w_*^\varepsilon(t), y_*^\varepsilon(t), z_*^\varepsilon(t))$ one-to-one correspond to the vector of switchings

$$\Theta_\varepsilon^* = \Big\{(\xi_{1,\varepsilon}^*, \xi_{2,\varepsilon}^*); (\tau_{1,\varepsilon}^*, \tau_{2,\varepsilon}^*); (\chi_{1,\varepsilon}^*, \chi_{2,\varepsilon}^*); (\eta_{1,\varepsilon}^*, \eta_{2,\varepsilon}^*)\Big\} \in [\Phi_2(T)]^4,$$

where

$$[\Phi_2(T)]^4 = \Phi_2(T) \times \Phi_2(T) \times \Phi_2(T) \times \Phi_2(T),$$

$$\Phi_2(T) = \Big\{(\lambda_1, \lambda_2) : 0 \le \lambda_1 \le \lambda_2 \le T\Big\}.$$

Now, let us consider an arbitrary sequence $\{\varepsilon_k\} \to +0$. The set $[\Phi_2(T)]^4$, in which the vectors $\Theta^*_{\varepsilon_k}$ are located, is compact. Hence, from the sequence $\{\Theta^*_{\varepsilon_k}\}$ we can choose a subsequence $\{\Theta^*_{\varepsilon_{k_m}}\}$ converging to vector Θ^*_0 such that

$$\Theta^*_0 = \Big\{(\xi^*_{1,0}, \xi^*_{2,0}); (\tau^*_{1,0}, \tau^*_{2,0}); (\chi^*_{1,0}, \chi^*_{2,0}); (\eta^*_{1,0}, \eta^*_{2,0})\Big\} \in [\Phi_2(T)]^4 .$$

Let this vector correspond to the controls $(u^0_*(t), v^0_*(t), w^0_*(t), y^0_*(t), z^0_*(t))$ by (B.4), which, in turn, correspond to the solutions $(S^0_*(t), E^0_*(t), I^0_*(t))$ as well. Without loss of generality, we assume that the sequence $\{\Theta^*_{\varepsilon_k}\}$ itself converges to the vector Θ^*_0.

Now, we show that the sequence of the controls

$$\phi^{\varepsilon_k}_*(t) = (u^{\varepsilon_k}_*(t), v^{\varepsilon_k}_*(t), w^{\varepsilon_k}_*(t), y^{\varepsilon_k}_*(t), z^{\varepsilon_k}_*(t))^\top$$

converges to the control

$$\phi^0_*(t) = (u^0_*(t), v^0_*(t), w^0_*(t), y^0_*(t), z^0_*(t))^\top$$

in the metric of the space $L_1[0, T]$, where

$$||\phi(\cdot)||_{L_1[0,T]} = \int_0^T \Big\{|u(t)| + |v(t)| + |w(t)| + |y(t)| + |z(t)|\Big\} dt.$$

Indeed, we have the relationships:

$$\begin{aligned}
||\phi^{\varepsilon_k}_*(\cdot) - \phi^0_*(\cdot)||_{L_1[0,T]} &= \int_0^T \Big\{|u^{\varepsilon_k}_*(t) - u^0_*(t)| + |v^{\varepsilon_k}_*(t) - v^0_*(t)| \\
&+|w^{\varepsilon_k}_*(t) - w^0_*(t)| + |y^{\varepsilon_k}_*(t) - y^0_*(t)| + |z^{\varepsilon_k}_*(t) - z^0_*(t)|\Big\} dt \\
&\le K_1 \sum_{j=1}^{2} |\xi^*_{j,\varepsilon_k} - \xi^*_{j,0}| + K_2 \sum_{j=1}^{2} |\tau^*_{j,\varepsilon_k} - \tau^*_{j,0}| + K_3 \sum_{j=1}^{2} |\chi^*_{j,\varepsilon_k} - \chi^*_{j,0}| \\
&+K_4 \sum_{j=1}^{2} |\eta^*_{j,\varepsilon_k} - \eta^*_{j,0}| + \varepsilon_k T \le K_0 ||\Theta^*_{\varepsilon_k} - \Theta^*_0|| + \varepsilon_k T,
\end{aligned}$$

where

$$K_1 = (\mu - u_{\min}) + (w_{\max} - w_{\min}), \quad K_2 = 1 - v_{\min},$$
$$K_3 = y_{\max} - y_{\min}, \quad K_4 = z_{\max} - z_{\min},$$

and the constant K_0 is determined by the values K_j, $j = \overline{1, 4}$. Then, we obtain the inequality

(B.5) $$||\phi^{\varepsilon_k}_*(\cdot) - \phi^0_*(\cdot)||_{L_1[0,T]} \le K_0 ||\Theta^*_{\varepsilon_k} - \Theta^*_0|| + \varepsilon_k T,$$

from which the required fact follows.

Next, let us estimate the difference of the solutions:

$$x^{\varepsilon_k}_*(t) = (S^{\varepsilon_k}_*(t), E^{\varepsilon_k}_*(t), I^{\varepsilon_k}_*(t))^\top, \quad x^0_*(t) = (S^0_*(t), E^0_*(t), I^0_*(t))^\top$$

corresponding to the controls $\phi_*^{\varepsilon_k}(t)$, $\phi_*^0(t)$. Integrating Cauchy problem (B.2) we have the following formulas:

$$x_*^{\varepsilon_k}(t) = \mathrm{e}^{t\Omega_0}x_0 + \int_0^t \mathrm{e}^{(t-s)\Omega_0}\Pi(x_*^{\varepsilon_k}(s))\phi_*^{\varepsilon_k}(s)ds,$$

$$x_*^0(t) = \mathrm{e}^{t\Omega_0}x_0 + \int_0^t \mathrm{e}^{(t-s)\Omega_0}\Pi(x_*^0(s))\phi_*^0(s)ds,$$

where e^M is the exponential of the constant matrix M.

Subtracting the second formula from the first and transforming the obtained expression, the following relationship can be found:

$$x_*^{\varepsilon_k}(t) - x_*^0(t) = \int_0^t \mathrm{e}^{(t-s)\Omega_0}\left(\Pi(x_*^{\varepsilon_k}(s)) - \Pi(x_*^0(s))\right)\phi_*^{\varepsilon_k}(s)ds$$
$$+ \int_0^t \mathrm{e}^{(t-s)\Omega_0}\Pi(x_*^0(s))\left(\phi_*^{\varepsilon_k}(s) - \phi_*^0(s)\right)ds.$$

Using Lemma 1, control constraints (2.7), the finite-increments formula ([41]) for the function $f(S,I)$ and arguments in [36] we see that there exist the positive constants M_1, M_2 independent of the particular choice of the controls $\phi_*^{\varepsilon_k}(t)$, $\phi_*^0(t)$ for which the following inequality holds:

$$||x_*^{\varepsilon_k}(t) - x_*^0(t)|| \le M_1||\phi_*^{\varepsilon_k}(\cdot) - \phi_*^0(\cdot)||_{L_1[0,T]} + M_2\int_0^t ||x_*^{\varepsilon_k}(s) - x_*^0(s)||ds.$$

Applying the Gronwall's inequality ([41]) to this relationship, we find the inequality:

$$||x_*^{\varepsilon_k}(t) - x_*^0(t)|| \le M_0||\phi_*^{\varepsilon_k}(\cdot) - \phi_*^0(\cdot)||_{L_1[0,T]} \tag{B.6}$$

that is valid for all $t \in [0,T]$. Here M_0 is the positive constant determined by the values M_j, $j = 1,2$. Putting $t = T$ in (B.6) and using the obvious relationship:

$$|J_{\varepsilon_k}(u_*^{\varepsilon_k}(\cdot), v_*^{\varepsilon_k}(\cdot), w_*^{\varepsilon_k}(\cdot), y_*^{\varepsilon_k}(\cdot), z_*^{\varepsilon_k}(\cdot)) - J_0(u_*^0(\cdot), v_*^0(\cdot), w_*^0(\cdot), y_*^0(\cdot), z_*^0(\cdot))|$$
$$\le ||x_*^{\varepsilon_k}(T) - x_*^0(T)||,$$

we finally have the inequality:

$$|J_{\varepsilon_k}(u_*^{\varepsilon_k}(\cdot), v_*^{\varepsilon_k}(\cdot), w_*^{\varepsilon_k}(\cdot), y_*^{\varepsilon_k}(\cdot), z_*^{\varepsilon_k}(\cdot)) - J_0(u_*^0(\cdot), v_*^0(\cdot), w_*^0(\cdot), y_*^0(\cdot), z_*^0(\cdot))|$$
$$\le M_0||\phi_*^{\varepsilon_k}(\cdot) - \phi_*^0(\cdot)||_{L_1[0,T]}. \tag{B.7}$$

From inequalities (B.5) and (B.7), the relationship follows

$$\lim_{k\to+\infty}\Big(J_{\varepsilon_k}(u_*^{\varepsilon_k}(\cdot), v_*^{\varepsilon_k}(\cdot), w_*^{\varepsilon_k}(\cdot), y_*^{\varepsilon_k}(\cdot), z_*^{\varepsilon_k}(\cdot))$$
$$- J_0(u_*^0(\cdot), v_*^0(\cdot), w_*^0(\cdot), y_*^0(\cdot), z_*^0(\cdot))\Big) = 0,$$

which implies the validity of the inequality:

$$J_0(u_*^0(\cdot), v_*^0(\cdot), w_*^0(\cdot), y_*^0(\cdot), z_*^0(\cdot)) \le \varliminf_{k\to+\infty} J_{\varepsilon_k}(u_*^{\varepsilon_k}(\cdot), v_*^{\varepsilon_k}(\cdot), w_*^{\varepsilon_k}(\cdot), y_*^{\varepsilon_k}(\cdot), z_*^{\varepsilon_k}(\cdot)).$$

Hence, we have a chain of the inequalities:

$$J_0^* = \min_{(u(\cdot),v(\cdot),w(\cdot),y(\cdot),y(\cdot))\in D_0(T)} J_0(u(\cdot), v(\cdot), w(\cdot), y(\cdot), z(\cdot)) \le J_0(u_*^0(\cdot), v_*^0(\cdot), w_*^0(\cdot), y_*^0(\cdot), z_*^0(\cdot)) \le \varliminf_{k\to+\infty} J_{\varepsilon_k}(u_*^{\varepsilon_k}(\cdot), v_*^{\varepsilon_k}(\cdot), w_*^{\varepsilon_k}(\cdot), y_*^{\varepsilon_k}(\cdot), z_*^{\varepsilon_k}(\cdot)) = \varliminf_{k\to+\infty} J_{\varepsilon_k}^*,$$

which implies the validity of the inequality:

$$J_0^* \le \varliminf_{k\to+\infty} J_{\varepsilon_k}^* \tag{B.8}$$

that means a lower semi-continuity of the optimal value of the perturbed functional from (2.9) at point $\varepsilon = 0$.

Now, let us show that for an arbitrary numerical sequence $\{\varepsilon_k\} \to +0$ the following inequality holds:

$$\varlimsup_{k\to+\infty} J_{\varepsilon_k}^* \le J_0^*, \tag{B.9}$$

which implies an upper semi-continuity of the optimal value of this functional at point $\varepsilon = 0$.

For this, for the functional $J_0(u(\cdot), v(\cdot), w(\cdot), y(\cdot), z(\cdot))$ we choose any minimizing sequence of admissible controls $(u^m(t), v^m(t), w^m(t), y^m(t), z^m(t))$, that is

$$\lim_{m\to+\infty} J_0(u^m(\cdot), v^m(\cdot), w^m(\cdot), y^m(\cdot), z^m(\cdot)) = \min_{(u(\cdot),v(\cdot),w(\cdot),y(\cdot),z(\cdot))\in D_0(T)} J_0(u(\cdot), v(\cdot), w(\cdot), y(\cdot), z(\cdot)) = J_0^*.$$

Then, in the considered minimizing sequence for the value $\varepsilon_k > 0$ there exists the controls $(u^{\varepsilon_k}(t), v^{\varepsilon_k}(t), w^{\varepsilon_k}(t), y^{\varepsilon_k}(t), z^{\varepsilon_k}(t))$ for which the inequality holds:

$$|J_0(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) - J_0^*| < \varepsilon_k$$

that implies the validity of the following relationship:

$$J_0(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) \le J_0^* + \varepsilon_k. \tag{B.10}$$

For any $\varepsilon_k > 0$ the inclusion

$$(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) \in D_{\varepsilon_k}(T)$$

holds. Then, the following equality

$$J_0(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) = J_{\varepsilon_k}(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) \tag{B.11}$$

holds as well.

Combining inequalities (B.10) and (B.11), we find a chain of the relationships:

$$J^*_{\varepsilon_k} \leq J_{\varepsilon_k}(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}, z^{\varepsilon_k}(\cdot))$$
$$= J_0(u^{\varepsilon_k}(\cdot), v^{\varepsilon_k}(\cdot), w^{\varepsilon_k}(\cdot), y^{\varepsilon_k}(\cdot), z^{\varepsilon_k}(\cdot)) \leq J^*_0 + \varepsilon_k,$$

which implies that the following inequality holds:

$$J^*_{\varepsilon_k} \leq J^*_0 + \varepsilon_k$$

that, in turn, means the validity of relationship (B.9).

Finally, by relationships (B.8) and (B.9), we find a chain of the inequalities:

$$J^*_0 \leq \varliminf_{k\to+\infty} J^*_{\varepsilon_k} \leq \varlimsup_{k\to+\infty} J^*_{\varepsilon_k} \leq J^*_0,$$

which implies the equalities:

$$J^*_0 = J_0(u^0_*(\cdot), v^0_*(\cdot), w^0_*(\cdot), y^0_*(\cdot), z^0_*(\cdot)) = \lim_{k\to+\infty} J^*_{\varepsilon_k}. \tag{B.12}$$

Hence, the controls $(u^0_*(t), v^0_*(t), w^0_*(t), y^0_*(t), z^0_*(t))$ in problem (2.9) is optimal.

Thus, we showed that under the validity one of the equalities (B.3) among the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $y_*(t)$ and $z_*(t)$ there exist such optimal controls that are piecewise constant functions with at most two moments of switching.

Andrei Korobeinikov
Centre de Recerca Matemática
Campus de Bellaterra, Edifici C
08193 Bellaterra, Barcelona, Spain
and
Departament de Matemàtiques
Universitat Autònoma de Barcelona
Campus de Bellaterra, Edifici C
08193 Barcelona, Spain
E-mail address: akorobeinikov@crm.cat

Ellina V. Grigorieva
Department of Mathematics and Computer Sciences
Texas Woman's University
Denton, TX 76204, USA
E-mail address: egrigorieva@twu.edu

Evgenii N. Khailov
Department of Computational Mathematics and Cybernetics
Moscow State Lomonosov University
Moscow, 119992, Russia
E-mail address: khailov@cs.msu.su

CRM
CENTRE DE RECERCA MATEMÀTICA
Facultat de Ciències, UAB
08193 Bellaterra
Barcelona, Spain
www.crm.cat