

Research Article

Optimal Controls of the Highly Active Antiretroviral Therapy

Ellina V. Grigorieva ¹, Evgenii N. Khailov ², and Andrei Korobeinikov ^{3,4,5}

¹Department of Mathematics and Computer Sciences, Texas Woman's University, Denton, TX 76204, USA

²Department of Computational Mathematics and Cybernetics, Moscow State Lomonosov University, Moscow 119992, Russia

³Centre de Recerca Matemàtica, Campus de Bellaterra, Edifici C, 08193 Bellaterra, Barcelona, Spain

⁴Departament de Matemàtiques, Universitat Autònoma de Barcelona, Campus de Bellaterra, Edifici C, 08193 Barcelona, Spain

⁵School of Mathematics and Information Science, Shaanxi Normal University, Xi'an, China

Correspondence should be addressed to Andrei Korobeinikov; akorobeinikov@crm.cat

Received 16 April 2019; Accepted 8 August 2019; Published 1 February 2020

Academic Editor: Valery Y. Glizer

Copyright © 2020 Ellina V. Grigorieva et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we study generic properties of the optimal, in a certain sense, highly active antiretroviral therapy (or HAART). To address this problem, we consider a control model based on the 3-dimensional Nowak–May within-host HIV dynamics model. Taking into consideration that precise forms of functional responses are usually unknown, we introduce into this model a nonlinear incidence rate of a rather general form given by an unspecified function of the susceptible cells and free virus particles. We also add a term responsible to the loss of free virions due to infection of the target cells. To mirror the idea of highly active anti-HIV therapy, in this model we assume six controls that can act simultaneously. These six controls affecting different stage of virus life cycle comprise all controls possible for this model and account for all feasible actions of the existing anti-HIV drugs. With this control model, we consider an optimal control problem of minimizing the infection level at the end of a given time interval. Using an analytical mathematical technique, we prove that the optimal controls are bang-bang, find accurate estimates for the maximal possible number of switchings of these controls and establish qualitative types of the optimal controls as well as mutual relationships between them. Having the estimate for the number of switchings found, we can reduce the two-point boundary value problem for Pontryagin Maximum Principle to a considerably simpler problem of the finite-dimensional optimization, which can be solved numerically. Despite this possibility, the obtained theoretical results are illustrated by numerical calculations using BOCOP–2.0.5 software package, and the corresponding conclusions are made.

1. Introduction

1.1. Antiretroviral Drugs and Guidelines. The World Health Organization and the United Nations Programme on HIV/AIDS (UNAIDS) estimated that over 36.9 millions of people were living with HIV/AIDS in 2017; other 1.8 million were newly infected with HIV in the same year [1, 2]. Of these 36.9 millions of the infected, 21.7 millions were living on antiretroviral therapy [1, 2]. These figures mean that the therapy is currently covers less than 60 percent of the infected. One (and probably the major) of the reasons for the insufficient coverage is the high cost of the antiretroviral therapy, which ranges from almost US \$20,000 per patient per year in the United States (where costs of the therapy are the highest) [3] to US \$10,000–15,000 per a person per year in other countries [4] (p. 90). Such costs make the treatment

inaccessible for the majority of patients in developing countries, where more than 95 percent of infections and HIV-associated deaths occurred.

Starting from 1987, when the first antiretroviral drugs, Nucleoside and Nucleotide Reverse Transcriptase Inhibitors, were approved for HIV treatment, over 30 of antiretroviral drugs were introduced in medical practice. A typical action of these drugs is inhibition of one of the several phases of HIV reproduction cycle. Depending on the stage that is inhibited, the drugs can be broadly classified into one of the following six groups [1, 2, 5]:

- (1) Entry or Fusion Inhibitors (Enfuvirtide, Maraviroc) Interfere with Binding, Fusion, and Entry of Virus to the Host Cell by Blocking One of Several Targets.
- (2) Nucleoside reverse transcriptase inhibitors (NRTI) and nucleotide reverse transcriptase inhibitors (NtRTI)

is the oldest and the largest class of antiretroviral drugs that includes drugs such as Zidovudine, Didanosine, Stavudine, Lamivudine, Abacavir, Tenofovir, and combinations of these. The NRTI and NtRTI inhibit reverse transcription by being incorporated into the newly synthesized viral DNA strand as faulty nucleotides.

- (3) Nonnucleoside reverse transcriptase inhibitors, or NNRTI, (Nevirapine, Delavirdine, Efavirenz, Etravirine, Rilpivirine) also inhibit reverse transcriptase, but in a different way, by binding to an allosteric site of the enzyme.
- (4) Integrase inhibitors (Raltegravir, Elvitegravir, Dolutegravir) inhibit the enzyme integrase, which is responsible for integration of the viral DNA into the DNA of the infected cell.
- (5) Protease inhibitors (PI) inhibit the activity of protease, an enzyme that cut nascent proteins into specific pieces for the final assembly of new functioning virions. There are ten approved protease inhibitors, namely: Saquinavir, Indinavir, Ritonavir, Nelfinavir, Amprenavir, Lopinavir/Ritonavir, Atazanavir, Fosamprenavir, Tipranavir, and Darunavir.
- (6) Maturation inhibitors inhibit the last step in gag processing where the viral capsid polyprotein is cleaved, blocking the conversion of the polyprotein into the mature capsid protein. As a result, the virions released have a defective core and are mainly noninfectious.

The guidelines for use of the antiretroviral drugs are a source of controversy within the medical community. Thus, there is diversity of opinion regarding the stage of infection and the CD4+ T cells count (which are the primary target cells for HIV, and which depletion leads to the immune system collapse and development of AIDS) when the therapy should be started. It is generally accepted that the objectives of the treatment are to maximally suppress the viral load and to maintain CD4+ T cells at an acceptable level. Other factors that have to be also taken into consideration are: (i) a possibility of individual intolerance and adverse side-effects; (ii) a possibility of developing drug resistance for a patient, and (iii) a possibility of emergence of drug-resistant strains, which thereby can be passed onto others. Moreover, the cost of the therapy, which is high, should be also taken into consideration. To reduce the possibility of developing drug resistance and of the emergence of drug-resistant strains and to minimize the side effects, a combination of several antiretroviral drugs (typically, of three or four) is usually used. This approach is known as highly active antiretroviral therapy, or HAART. Also, it is generally accepted that, in order to reduce a possibility of emergence of drug resistant strains, antiretroviral therapy, once started, should never be stopped: as current US Guidelines states, "Patients initiating antiretroviral therapy should be willing and able to commit to lifelong treatment" [6].

The above-mentioned controversy in formulating the guidelines implies that some sort of optimization can be applied, and that the problem should be quantified and explored mathematically. Of course, at the present stage, it

would be hardly possible to take into consideration details such as individual sensitivity of a specific patient to a specific drug. However, some common principles that can assist in developing better guidelines can be established. A natural approach for this sort of problems is to use the tools and methods of the classic optimal control theory. The problem of HIV control and the applicability of the optimal control theory to this problem attracted attention of mathematicians fairly early: seminal works by Butler et al. [7] and Kirschner et al. [8] initiated extensive literature in this direction. A mathematical techniques suggested in these publications proved to be successful and was subsequently applied to a variety of mathematical model of HIV (see, e.g., [9–22] and bibliography therein). These results brought important insight into the problem. However, since this advance utilized essentially the same mathematical techniques, all these results also share some drawbacks, that we have to discuss before we proceed to our study.

1.2. Form of the Objective Functional. Starting from [7, 8], all the above-mentioned publication (and, in fact, the vast majority of the literature on the optimal control in mathematical biology and medicine) deal with optimal control problems of maximization of the cumulative (integral) concentration of healthy CD4+ T cells and simultaneous minimization of the cost of the treatment (represented by the weighted squares of the considered controls). Control problems with such functionals are mathematically convenient because they never deal with singular control phenomenon and can be solved numerically as a two-point boundary value problem. A setback of such objective functionals is that the optimal controls, obtained for such objective functionals, are very sensitive to a specific form of the functional and, in particular, to the value of the exponent. That is, the optimal controls obtained for such objective functionals are not robust, and a small variation of the exponent or even the weights leads to significant changes of the optimal controls. Furthermore, the above-mentioned approach does not guarantee the optimality of the obtained solutions. Instead, it ensures only necessary conditions of the optimality and a local extremum of the functional.

At the same time, while the objective functionals that are independent of the controls are free of the mentioned drawbacks, their use involves dealing with considerably more difficult problems, as the control-independent functionals usually lead to bang-bang controls and can have singular arcs. There are also other issues of mathematical nature that 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 Schättler and Ledzewicz [23] (p. 48–50). Nevertheless, despite this complexity, there is a certain progress in application of the control-independent functionals in biomedical control problems. Results that affirmatively demonstrate that the use of such functionals is possible and potentially highly beneficial, include publications of Ledzewicz, Schättler and their coauthors on the optimal control in anticancer therapy (see, e.g., [23, 24] and bibliography therein), and of the authors on the optimal waste water biotreatment [25, 26–28] and on the optimal control of epidemics [29–34].

The objective functionals that include squares of the controls naturally occur in problems originated in mechanics and engineering, where square of a control is usually proportional to the energy required for the control action. A typical motivation for the use of the quadratic objective in HIV treatment (and in biomedical applications in general) is minimizing the “costs” in terms of both financial costs and the adverse side effects. That is, this specific form of objective functional implies that the cost and side effects increase as squares of the drug concentrations in patient’s bloodstream. However, while there is little doubt that the severity of side effects grows faster than linear with the growth of drug concentration, and, hence, an exponential functional response may be a good representation for this, the actual value of the exponent is uncertain. Moreover, it depends on patient’s individual response and varies for different patient and different drugs. This uncertainty combined with the outcome sensitivity to the exponent implies that for clinical practice the value of results, obtained for this particular type of objective functional, is questionable. Furthermore, the quadratic “cost” is hardly able to adequately represent the financial cost of the therapy either.

Taking into consideration the uncertainty of the form of the real-life “cost” function and the sensitivity of the results to this, we consider an optimal control problem for HIV treatment disregarding the “costs”. Instead, we assume that the “costs”, in terms of both the side effects and the financial cost of the therapy, is limited by the maximal acceptable levels of medication. An argument in support of such a formulation is that the actual objectives of the therapy are: (i) prevention of the development of AIDS by a patient, and (ii) control of the epidemic in a society. For an individual patient, these objectives imply reducing virus load to an acceptable safe level. Considering the severity of the possible consequences of the control failure, individual “costs” of the therapy can be disregarded. We also assume that the both, financial cost of the therapy and the side effects are known and accepted prior the start of the therapy.

An initial advance in this direction was made in [35], where the authors considered an optimal control problem for a reasonably simple 3-dimensional antiretroviral therapy model. The model in this paper was based on the 2-dimensional Wodarz model of HIV dynamics [36], where an equation representing the concentration of a drug inhibiting the incidence rate was added. The drug intake was considered as the single control. The optimal control problem for the model with a control-independent objective functional was state and solved, and the principal properties of the optimal controls were established.

1.3. Interaction of Several Simultaneously Acting Controls. The majority of above-mentioned papers on the optimal anti-HIV therapy consider models with a single control, which correspond to the therapies where a single drug (a single action) is used. Very few of existing publications consider a possibility of simultaneous use of two drugs [15, 20, 37–40], whereas works exploring interaction of three or more controls are rather exceptional [41, 42]. On the other hand, the current anti-HIV therapy guidelines necessitate the use of the highly active antiretroviral therapy (HAART) that implies

administering a combination of several (typically, three or four) antiretrovirals. Moreover, over thirty anti-HIV drugs are currently available, which, by their action, can be classified into six groups that affect six different stages of viral life cycle. (NRTI and NNRTI are in different groups but affect the same stage of viral replication). Such a surplus of drugs’ names and actions implies a large number of their possible combinations. Currently, the drug combinations used in HAART are arranged with the aims of (i) minimizing the adverse side effects and (ii) preventing the development of drug resistance. However, the existing surplus of the drugs and their actions presents an opportunity for designing combinations, as well as schedules of their administering, aimed to optimize the HAART efficacy as well. Apparently, this implies the need to explore interaction and interference of the six (that is, equal to the number of the HIV replication cycle stages that the existing antiretroviral drugs can inhibit) simultaneously applied antiretroviral actions.

1.4. Nonlinearity of Functional Responses. The use of a numerical technique in papers [7–10, 12–22] means that in each case results are obtained for a model with a specific parametrization. However, due to intrinsic complexity of biological systems, the actual forms of functional responses and the parametrization are usually unknown in detail, and one hardly can expect that the information available on a bio-system is, or will be, sufficient. This implies that outcome of a study can depend on factors, which are neither studied, nor understood with a sufficient degree of certainty. This, in turn, leads to a question about the reliability and relevance of the results obtained for models with specific forms of functional responses and parametrization. In the framework of the traditional formulation that was used in these papers, this problem cannot be avoided.

This problem was recognized at a very early stage of existence of mathematical biology. One way of dealing with this problem was suggested in 1930s by Kolmogorov [43], who suggested to “reverse” the problem: instead of studying properties of a biological system with specifically defined functional responses and 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.) that these functions should have in order to ensure the possession of certain properties by the considered system. This Kolmogorov’s concept was successfully used in the global analysis, particularly in combination with the direct Lyapunov method [44–48]. However, it is obvious that this concept hardly can be employed in a combination with numerical analysis.

In biomedical applications of the optimal control theory this problem was recognized as well. To the best authors’ knowledge, Horst Behncke was the first who employed the Kolmogorov’s concept to an optimal control problem in mathematical biology [49]. He considered the classic 2-dimensional SIR Kermack–McKendrick epidemic model of an isolated epidemic assuming that the infection rate (which is the most important of all functional responses in epidemic models) is given by an unspecified function constrained by a few conditions. Behncke considered this model under action of one of three controls; namely, vaccination of the susceptible

individuals, isolation of the infectious individuals, and education that reduces a probability of infection: these three controls comprise all controls possible for such a simple model. Considering arguments as above, that is taking into consideration the disparity between the costs of a control policy and the costs and losses inflicted by a full-scale epidemic, Behncke employed objective functionals independent of the controls and considered minimization of the cumulative number of the infected over a given time interval. Assuming that in each case only one control is employed and that the controls are bounded, he found qualitative forms of these optimal controls.

A similar approach was later applied to more complicated problems. Thus, the optimal controls for a SIR model of endemically persistent diseases (i.e., a SIR model with demographic processes) where the incidence rate was given by an unspecified function were considered in [33]. In this control model, four simultaneously acting controls were considered and their principal qualitative types established. It is noteworthy that there are more variants of the optimal controls for this problem compared with the single epidemic models considered by Behncke. Moreover, comparison of these results with the controls found for the same model with the standard bilinear transmission rate in [30, 32] shows that the nonlinearity of incidence rate does not affect the principal properties of the control. A more complicated 3-dimensional SEIR model with an unspecified nonlinear incidence rate and five simultaneously acting controls (which comprise all controls possible for this model) was considered in [34].

1.5. Objectives and Mathematical Technique of this Paper. The objective of this paper is to study and establish properties of the optimal, highly active antiretroviral therapy, at the same time avoiding the above-mentioned drawbacks of the preceding studies. Having this goal in mind, we consider a model of HIV within-host dynamics based on the classical Nowak–May model, which describe interaction of the susceptible and infected target cells and free virus particles [50, 51]. The Nowak–May model postulates that the susceptible target cells can be infected by free virus particles and after the instance of infection move into the infected population; in their turn, the infected cells produce free virus particles that can infect other susceptible cells. Using the Nowak–May model as basis, we assume that the rate of infection is given by a unspecified function of the concentrations of susceptible cells and free virus particles, constrained by a few conditions. Into the model, we also add a term responsible for the loss of free virus due to infection (see [52] for discussion of this term).

In order to mirror the HAART, we introduce six bounded controls that can act simultaneously into this model. These six controls comprise all controls that are possible for this model and correspond to the six stages of the viral replication cycle that the current antiretrovirals can affect. For this model, we consider a control problem of minimizing the infection level and the viral load at the end point of a finite time interval. The objective functional that we consider is independent of the controls.

While for our type of objective functional, the optimal controls are likely to be of the bang-bang type, theoretically, they can contain singular arcs. Thus in this paper, we prove

that our optimal controls are only bang-bang and that no singular arcs are possible. Next, the maximal number of switchings for each control can be analytically estimated and then the specific number of the switching moments for each control can be found numerically as a problem of the finite-dimensional optimization.

To deal with this optimal control problem, we employ a following analytical technique:

- (i) Using Pontryagin Maximum Principle, we found differential equations for the adjoint variables and for the switching functions that entirely determine types of the corresponding optimal controls.
- (ii) Then, analyzing the system of differential equations for switching functions, we prove that the optimal controls are bang-bang.
- (iii) Our next task is to find accurate estimates for the number of zeros of the switching functions which correspond to switching points of controls. In order to do this, we apply the generalized Rolle's theorem [53] to the equations of the system for switching functions. However, a direct application of the theorem to these equations is impossible, and we have to convert the matrix of this system into the upper triangular form. In order to do this, we have to find a suitable transformation, which is usually defined by a system of nonautonomous quadratic differential equations, and to prove that solutions of these equations are defined on the entire time interval. One of the results of this paper is that we managed to reduce a 3-dimensional system of nonautonomous quadratic differential equations to two 2-dimensional systems: the analysis of these two 2-dimensional systems is simpler than that of a single 3-dimensional system.
- (iv) Having the maximal possible number of switchings found, we can immediately find the principal type of the optimal controls.
- (v) Finally, to solve the problem completely, one can find the specific number of switchings and their precise positions, which can be done numerically. Thus, our analytical technique allows to reduce the optimal control problem to a considerably simpler problem of finite-dimensional optimization. Despite this possibility, the obtained theoretical results are illustrated by numerical calculations using BOCOP-2.0.5 software package, and the corresponding conclusions are made.

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 in the model and, hence, a possibility to use this approach in combination with the Kolmogorov's "reversion of a problem" concept; we exploit this possibility in the paper.

An analytical methods used in this paper can be employed to analyze optimal controls for other possible objectives, as

well as for other models of virus dynamics. These findings have the immediate practical relevance providing answers to long-existing questions about the interference of drugs with different actions combinations of which are used in HAART and about the conditions and objectives for which the current therapy guidelines are optimal.

The paper is organized as follows: In Section 2, we formulate the model and state the optimal control problem. In Section 3, we apply Pontryagin Maximum Principle to find the adjoint system and the system of nonautonomous differential equations for the switching functions. In Sections 4, we prove that the optimal controls are bang-bang. In Section 5, we find the estimates for the number of zeros of the switching functions, and thereby, in Section 6, we describe the principal types of the optimal controls that are possible for this problem. In Section 7, we provide results of computations that illustrate our analytical results. Section 8 contains conclusion and discussion of the results. Appendix A describes details of the technique of the reduction of a 3-dimensional matrix of a system of linear nonautonomous homogeneous differential equations for the switching functions to the upper triangular form. Appendix B is devoted to the verification of one of our auxiliary assumptions.

2. Model and the Optimal Control Problem

We consider the following model of the within-host HIV dynamics:

$$\begin{cases} \dot{x}(t) = \lambda - f(x(t), z(t)) - \mu x(t), \\ \dot{y}(t) = f(x(t), z(t)) - \sigma y(t), \\ \dot{z}(t) = ky(t) - f(x(t), z(t)) - \gamma z(t), \\ x(0) = x_0, y(0) = y_0, z(0) = z_0; x_0, y_0, z_0 > 0. \end{cases} \quad (2.1)$$

Here $x(t)$, $y(t)$ and $z(t)$ are the appropriate concentrations of uninfected (susceptible) target cells, infected target cells and free virus particles at time t . The model postulates that the susceptible cells (for HIV these are CD4+ T helper cells) enter into the system (into the patient blood from the thymus, where CD4+ cells mature) at a constant rate λ , that the susceptible cells are infected by the free virus particles at a rate $f(x(t), z(t))$ and the infected cells immediately move to the infected population, that the infected cells produce free virus particles at the rate ky , and that average life-spans of the susceptible cells, infected cells and free virus particles are, respectively, μ^{-1} , σ^{-1} and γ^{-1} (naturally, for HIV $\sigma \geq \mu$). The third equation of the model also includes a removal term representing the loss of virus due to infection of cells. (This term is usually omitted in mathematical models of HIV dynamics as very small compared to the virus death rate $\gamma z(t)$). However, for low viral loads typical for patients under antiviral therapy, the impact of the removal of free virus due to infection can be of importance).

The model (2.1) is nearly classical. In the literature it usually appears with a bilinear incidence rate and without the term representing the loss of free virus due to infection and is referred to as the Nowak–May model [51, 54]. To the best of the authors’ knowledge, a version of the model with the unspecified general nonlinear incidence rate $f(x, z)$ firstly appeared in [46]. The dynamics of this model with both

bilinear and unspecified nonlinear incidence rates is studied in detail both analytically and numerically in [50, 51, 54–56, 44–47]. In particular, and what is of importance for further analysis, it is known that the nonnegative octant of the 3-dimensional space $\{(x, y, z) \in \mathbb{R}^3 : x, y, z \geq 0\}$ is a positive invariant set of system (2.1) and that all solutions of the system initiated in this set are bounded.

In this paper, we assume that the incidence rate $f(x, z)$ satisfies the following conditions.

Assumption 1. Let the function $f(x, z)$ satisfy:

- (i) $f(x, z)$ is a continuous differentiable function for all $x, z \geq 0$;
- (ii) $f(x, z) > 0$ for all $x, z > 0$, and $f(0, z) = f(x, 0) = 0$ for all $x, z \geq 0$;
- (iii) Partial derivatives $f'_x(x, z)$ and $f'_z(x, z)$ are positive for all $x, z > 0$ and nonnegative for all $x \geq 0, z = 0$ and $x = 0, z \geq 0$.

Please note that this set of assumptions is less restrictive than these that are usually used in the literature (see, e.g., [44–47, 56]).

To make system (2.1) controllable, we assume that on a pre-defined time interval $[0, T]$ the system is an object of six independent controls $\tilde{u}(t), \tilde{v}(t), \tilde{w}(t), \tilde{p}(t), \tilde{q}(t)$ and $\tilde{r}(t)$, which have the following biological meanings:

- (i) $\tilde{u}(t)$ is a fraction of immune cells among the new target cells entering the system;
- (ii) $\tilde{w}(t)$ is a per capita rate of “immunization” of the susceptible cells that makes the “immunized” cells entirely unsusceptible to the virus;
- (iii) $\tilde{v}(t)$ is a magnitude of indirect measures that inhibit the incidence rate;
- (iv) $\tilde{p}(t)$ is a magnitude of indirect measures aimed that inhibit the free virus particles production by the infected cells;
- (v) $\tilde{q}(t)$ is a per capita rate of immobilization (or removal) of free virus particles from the system;
- (vi) $\tilde{r}(t)$ is a per capita rate of killing (or of removal by other means) of the infected cells.

We would like to stress that these six controls comprise all interventions that are feasible or potentially possible for model (2.1).

We assume that the controls satisfy the following constraints:

$$\begin{aligned} 0 \leq \tilde{u}(t) \leq \tilde{u}_{\max} < 1, \quad 0 \leq \tilde{v}(t) \leq \tilde{v}_{\max}, \quad 0 \leq \tilde{w}(t) \leq \tilde{w}_{\max}, \\ 0 \leq \tilde{p}(t) \leq \tilde{p}_{\max}, \quad 0 \leq \tilde{q}(t) \leq \tilde{q}_{\max}, \quad 0 \leq \tilde{r}(t) \leq \tilde{r}_{\max}. \end{aligned} \quad (2.2)$$

Here the maximal values of control functions $\tilde{u}_{\max}, \tilde{v}_{\max}, \tilde{w}_{\max}, \tilde{p}_{\max}, \tilde{q}_{\max}$ and \tilde{r}_{\max} are pre-defined positive constants.

Assuming that these six controls are applied to model (2.1) on the time interval $[0, T]$, we obtain the following control model:

$$\begin{cases} \dot{x}(t) = \lambda(1 - \tilde{u}(t)) - (1 + \tilde{v}(t))^{-1}f(x(t), z(t)) \\ \quad - (\mu + \tilde{w}(t))x(t), \\ \dot{y}(t) = (1 + \tilde{v}(t))^{-1}f(x(t), z(t)) - (\sigma + \tilde{r}(t))y(t), \\ \dot{z}(t) = k(1 + \tilde{p}(t))^{-1}y(t) - (1 + \tilde{v}(t))^{-1}f(x(t), z(t)) \\ \quad - (\gamma + \tilde{q}(t))z(t), \\ x(0) = x_0, y(0) = y_0, z(0) = z_0; x_0, y_0, z_0 > 0. \end{cases} \quad (2.3)$$

For convenience of notation, we introduce modified controls:

$$\begin{aligned} u(t) &= \lambda(1 - \tilde{u}(t)), & v(t) &= (1 + \tilde{v}(t))^{-1}, & w(t) &= \mu + \tilde{w}(t), \\ p(t) &= (1 + \tilde{p}(t))^{-1}, & q(t) &= \gamma + \tilde{q}(t), & r(t) &= \sigma + \tilde{r}(t). \end{aligned} \quad (2.4)$$

These modified controls satisfy the following constraints:

$$\begin{aligned} u_{\min} \leq u(t) \leq u_{\max}, & \quad v_{\min} \leq v(t) \leq v_{\max}, & \quad w_{\min} \leq w(t) \leq w_{\max}, \\ p_{\min} \leq p(t) \leq p_{\max}, & \quad q_{\min} \leq q(t) \leq q_{\max}, & \quad r_{\min} \leq r(t) \leq r_{\max}, \end{aligned} \quad (2.5)$$

where

$$\begin{aligned} u_{\min} &= \lambda(1 - \tilde{u}_{\max}), & u_{\max} &= \lambda, & v_{\min} &= (1 + \tilde{v}_{\max})^{-1}, & v_{\max} &= 1, \\ w_{\max} &= \mu + \tilde{w}_{\max}, & w_{\min} &= \mu, & p_{\min} &= k(1 + \tilde{p}_{\max})^{-1}, & p_{\max} &= k, \\ q_{\max} &= \gamma + \tilde{q}_{\max} & q_{\min} &= \gamma, & r_{\max} &= \sigma + \tilde{r}_{\max}, & r_{\min} &= \sigma. \end{aligned} \quad (2.6)$$

In addition to Assumption 1, we make the following assumption.

Assumption 2. Maximal and minimal levels $w_{\min}, w_{\max}, p_{\min}, p_{\max}, q_{\min}, q_{\max}, r_{\min}, r_{\max}$ of the controls satisfy the conditions:

$$w_{\min} \neq \{q_{\min}, q_{\max}, r_{\min}, r_{\max}\}, \quad (2.7)$$

$$w_{\max} \neq \{q_{\min}, q_{\max}, r_{\min}, r_{\max}\}, \quad (2.8)$$

$$r_{\min} \neq \{(p_{\min} + q_{\max}), (p_{\max} + q_{\min})\}, \quad (2.9)$$

$$r_{\max} \neq \{(p_{\min} + q_{\max}), (p_{\max} + q_{\min})\}. \quad (2.10)$$

As a set of admissible controls, $\Omega(T)$, we consider the set of all Lebesgue measurable functions $u(t), v(t), w(t), p(t), q(t)$ and $r(t)$ satisfying inequalities (2.5) for almost all $t \in [0, T]$.

Using modified controls (2.4), we rewrite control system (2.3) as following:

$$\begin{cases} \dot{x}(t) = u(t) - v(t)f(x(t), z(t)) - w(t)x(t), \\ \dot{y}(t) = v(t)f(x(t), z(t)) - r(t)y(t), \\ \dot{z}(t) = p(t)y(t) - v(t)f(x(t), z(t)) - q(t)z(t), \\ x(0) = x_0, y(0) = y_0, z(0) = z_0; x_0, y_0, z_0 > 0. \end{cases} \quad (2.11)$$

The following lemma ensures the boundedness, positiveness, and continuity of solutions for system (2.11).

Lemma 3. For any admissible controls $u(t), v(t), w(t), p(t), q(t)$, and $r(t)$, the corresponding absolutely continuous solution $x(t), y(t), z(t)$ to system (2.11) are defined on the entire interval $[0, T]$, and for its components $x(t), y(t), z(t)$ the inequalities:

$$0 < x(t) < x_{\max}, 0 < y(t) < y_{\max}, 0 < z(t) < z_{\max}, \quad (2.12)$$

where

$$x_{\max}, y_{\max} = \frac{(x_0 + y_0) + \lambda}{\min\{\mu; \sigma\}}, \quad (2.13)$$

$$z_{\max} = \frac{(y_0 + z_0) + ky_{\max}}{\min\{\sigma; \gamma\}}, \quad (2.14)$$

hold for all $t \in [0, T]$.

Proof of this lemma is fairly straightforward and hence we omit it. Proofs of similar statements are given, for example, in [23, 57–59].

On the set of admissible controls $\Omega(T)$ for system (2.11), we consider the problem of minimizing a weighted sum of concentrations of the infected cells and free virus particles at the end of time interval $[0, T]$:

$$J(u, v, w, p, q, r) = \alpha y(T) + \beta z(T) \rightarrow \min_{u(\cdot), v(\cdot), w(\cdot), p(\cdot), q(\cdot), r(\cdot)}. \quad (2.15)$$

Such objective functional usually corresponds to the task of complete elimination of the infection. In (2.15), α and β are positive weights, such that $\alpha \neq \beta$. Please note that the objective functional $J(u, v, w, p, q, r)$ is independent of the controls.

For optimal control problem (2.11), (2.15), inequalities (2.12) in Lemma 3 and Theorem 4 of Chapter 4 in [60] ensure the existence of the optimal controls $u_*(t), v_*(t), w_*(t), p_*(t), q_*(t)$ and $r_*(t)$ and the corresponding optimal solution $(x_*(t), y_*(t), z_*(t))$ of system (2.11). In the following sections we establish the principal properties of these controls.

3. Pontryagin Maximum Principle

To analyze the optimal control problem (2.11), (2.15), we use Pontryagin Maximum Principle [61]. The Hamiltonian of this problem is

$$\begin{aligned} H(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) \\ = (u - vf(x, z) - wx)\psi_1 \\ + (vf(x, z) - ry)\psi_2 + (py - vf(x, z) - qz)\psi_3, \end{aligned} \quad (3.1)$$

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

$$\begin{aligned} \frac{\partial H}{\partial x}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= vf'_x(x, z)(\psi_2 - \psi_1 - \psi_3) - w\psi_1, \\ \frac{\partial H}{\partial y}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= -r\psi_2 + p\psi_3, \\ \frac{\partial H}{\partial z}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= vf'_z(x, z)(\psi_2 - \psi_1 - \psi_3) - q\psi_3, \\ \frac{\partial H}{\partial u}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= \psi_1, \\ \frac{\partial H}{\partial v}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= f(x, z)(\psi_2 - \psi_1 - \psi_3), \\ \frac{\partial H}{\partial w}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= -x\psi_1, \\ \frac{\partial H}{\partial p}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= y\psi_3, \\ \frac{\partial H}{\partial q}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= -z\psi_3, \\ \frac{\partial H}{\partial r}(x, y, z, \psi_1, \psi_2, \psi_3, u, v, w, p, q, r) &= -y\psi_2. \end{aligned} \quad (3.2)$$

Here $f'_x(x, z)$ and $f'_z(x, z)$ are the partial derivatives of function $f(x, z)$ with respect to variables x and z , respectively.

According to Pontryagin Maximum Principle, for the optimal controls $u_*(t), v_*(t), w_*(t), p_*(t), q_*(t), r_*(t)$ and the

corresponding optimal solution $(x_*(t), y_*(t), z_*(t))$ of system (2.11), there exists a vector-function $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$ such that:

- (i) The vector-function $\psi_*(t) = (\psi_1^*(t), \psi_2^*(t), \psi_3^*(t))$ is a nontrivial solution of the adjoint system

$$\begin{cases} \dot{\psi}_1^*(t) = w_*(t)\psi_1^*(t) \\ \quad -v_*(t)f'_x(x_*(t), z_*(t))(\psi_2^*(t) - \psi_1^*(t) - \psi_3^*(t)), \\ \dot{\psi}_2^*(t) = r_*(t)\psi_2^*(t) - p_*(t)\psi_3^*(t), \\ \dot{\psi}_3^*(t) = q_*(t)\psi_3^*(t) \\ \quad -v_*(t)f'_z(x_*(t), z_*(t))(\psi_2^*(t) - \psi_1^*(t) - \psi_3^*(t)), \\ \psi_1^*(T) = 0, \psi_2^*(T) = -\alpha, \psi_3^*(T) = -\beta. \end{cases} \quad (3.3)$$

In this system, it is easy to note that

$$\begin{aligned} \dot{\psi}_1^* &= -\frac{\partial H}{\partial x}(x_*, y_*, z_*, \psi_1^*, \psi_2^*, \psi_3^*, u_*, v_*, \\ &\quad w_*, p_*, q_*, r_*), \\ \dot{\psi}_2^* &= -\frac{\partial H}{\partial y}(x_*, y_*, z_*, \psi_1^*, \psi_2^*, \psi_3^*, u_*, v_*, \\ &\quad w_*, p_*, q_*, r_*), \\ \dot{\psi}_3^* &= -\frac{\partial H}{\partial z}(x_*, y_*, z_*, \psi_1^*, \psi_2^*, \psi_3^*, u_*, v_*, \\ &\quad w_*, p_*, q_*, r_*), \end{aligned} \quad (3.4)$$

and

$$\psi_1^*(T) = -\frac{\partial J}{\partial x(T)}, \psi_2^*(T) = -\frac{\partial J}{\partial y(T)}, \psi_3^*(T) = -\frac{\partial J}{\partial z(T)}. \quad (3.5)$$

- (ii) The controls $u_*(t), v_*(t), w_*(t), p_*(t), q_*(t), r_*(t)$ maximize the Hamiltonian

$$H(x_*(t), y_*(t), z_*(t), \psi_1^*(t), \psi_2^*(t), \psi_3^*(t), u, v, w, p, q, r), \quad (3.6)$$

with respect to variables:

$$\begin{aligned} u \in [u_{\min}, u_{\max}], \quad v \in [v_{\min}, v_{\max}], \quad w \in [w_{\min}, w_{\max}], \\ p \in [p_{\min}, p_{\max}], \quad q \in [q_{\min}, q_{\max}], \quad r \in [r_{\min}, r_{\max}], \end{aligned} \quad (3.7)$$

for almost all $t \in [0, T]$, and, therefore, the following relationships hold:

$$u_*(t) = \begin{cases} u_{\max}, & \text{if } L_u(t) > 0, \\ \forall u \in [u_{\min}, u_{\max}], & \text{if } L_u(t) = 0, \\ u_{\min}, & \text{if } L_u(t) < 0, \end{cases} \quad (3.8)$$

$$v_*(t) = \begin{cases} v_{\max}, & \text{if } L_v(t) > 0, \\ \forall v \in [v_{\min}, v_{\max}], & \text{if } L_v(t) = 0, \\ v_{\min}, & \text{if } L_v(t) < 0, \end{cases} \quad (3.9)$$

$$w_*(t) = \begin{cases} w_{\max}, & \text{if } L_w(t) > 0, \\ \forall w \in [w_{\min}, w_{\max}], & \text{if } L_w(t) = 0, \\ w_{\min}, & \text{if } L_w(t) < 0, \end{cases} \quad (3.10)$$

$$p_*(t) = \begin{cases} p_{\max}, & \text{if } L_p(t) > 0, \\ \forall p \in [p_{\min}, p_{\max}], & \text{if } L_p(t) = 0, \\ p_{\min}, & \text{if } L_p(t) < 0, \end{cases} \quad (3.11)$$

$$q_*(t) = \begin{cases} q_{\max}, & \text{if } L_q(t) > 0, \\ \forall q \in [q_{\min}, q_{\max}], & \text{if } L_q(t) = 0, \\ q_{\min}, & \text{if } L_q(t) < 0, \end{cases} \quad (3.12)$$

$$r_*(t) = \begin{cases} r_{\max}, & \text{if } L_r(t) > 0, \\ \forall r \in [r_{\min}, r_{\max}], & \text{if } L_r(t) = 0, \\ r_{\min}, & \text{if } L_r(t) < 0. \end{cases} \quad (3.13)$$

In these relationships, by Lemma 3, functions

$$\begin{aligned} L_u(t) &= \psi_1^*(t), \quad L_v(t) = \psi_2^*(t) - \psi_1^*(t) - \psi_3^*(t), \\ L_w(t) &= -\psi_1^*(t), \quad L_p(t) = \psi_3^*(t), L_q(t) = -\psi_3^*(t), \\ L_r(t) &= -\psi_2^*(t), \end{aligned} \quad (3.14)$$

are the switching functions, which determine a type of the optimal controls $u_*(t), v_*(t), w_*(t), p_*(t), q_*(t)$ and $r_*(t)$.

Please note that in (3.14)

$$L_u(t) = -L_w(t), L_p(t) = -L_q(t). \quad (3.15)$$

Hence, it suffices to study the properties of only two of these four functions, say $L_w(t)$ and $L_q(t)$. Therefore, further we focus on the analysis of the switching functions $L_v(t), L_w(t), L_q(t)$ and $L_r(t)$. By (3.3) and (3.14), we obtain the following Cauchy problem for the switching functions $L_v(t), L_w(t), L_q(t)$ and $L_r(t)$:

$$\begin{cases} \dot{L}_v(t) = (v_*(t)(f'_x(x_*(t), z_*(t)) + f'_z(x_*(t), z_*(t))) + r_*(t))L_v(t) \\ \quad + (w_*(t) - r_*(t))L_w(t) + (p_*(t) + q_*(t) - r_*(t))L_q(t), \\ \dot{L}_w(t) = w_*(t)L_w(t) + v_*(t)f'_x(x_*(t), z_*(t))L_v(t), \\ \dot{L}_q(t) = q_*(t)L_q(t) + v_*(t)f'_z(x_*(t), z_*(t))L_v(t), \\ \dot{L}_r(t) = r_*(t)L_r(t) - p_*(t)L_q(t), \\ L_v(T) = \beta - \alpha, \quad L_w(T) = 0, \quad L_q(T) = \beta, \quad L_r(T) = \alpha. \end{cases} \quad (3.16)$$

We actively use this Cauchy problem in subsequent arguments.

4. Properties of the Switching Functions

We start the analysis by establishing some properties of switching functions $L_v(t), L_w(t), L_q(t), L_r(t)$.

Lemma 4. *There exist values $t_q, t_r \in [0, T]$ such that*

$$L_q(t) < 0 \quad \text{for } t \in (t_q, T]; \quad L_r(t) < 0 \quad \text{for } t \in (t_r, T]. \quad (4.1)$$

hold.

Proof. The switching functions $L_q(t)$ and $L_r(t)$ are the solutions of the third and fourth equations of system (3.16), respectively, and, hence, these functions are continuous functions. Since $L_q(T) = \beta, L_r(T) = \alpha$ and $\alpha, \beta > 0$, by Theorem on the Stability of the Sign of a Continuous Function (see, e.g., [62]) there are intervals $(t_q, T]$ and $(t_r, T]$ where inequalities (4.1) necessary hold. \square

Lemma 5. *Depending on values of the weighting coefficients α and β , for switching functions $L_v(t)$ and $L_w(t)$, the following statements hold:*

(i) if $\alpha < \beta$, then there exist $t_v, t_w \in (0, T)$, $t_w < t_v$, such that
 $L_v(t) < 0$ for $t \in (t_v, T]$; $L_w(t) < 0$ for $t \in (t_w, T]$;
 (4.2)

(ii) if $\alpha > \beta$, then there exist $t_v, t_w \in (0, T)$, $t_w < t_v$, such that

$$L_v(t) < 0 \text{ for } t \in (t_v, T]; L_w(t) < 0 \text{ for } t \in (t_w, T]. \tag{4.3}$$

Proof. One of the inequalities for function $L_v(t)$ immediately follows from the arguments similar to those presented in the proof of Lemma 4. Indeed, switching function $L_v(t)$ satisfies the first equation of system (3.16), and, hence, is absolutely continuous and, therefore, is also a continuous function. Then, for $\alpha \neq \beta$, by Theorem on the Stability of the Sign of a Continuous Function, there is t_v such that, depending on the sign of $(\alpha - \beta)$, either the first inequality in (4.2), or the first inequality in (4.3), holds.

To get the inequality for $L_w(t)$, we integrate the second equation of system (3.16) on the interval $[0, T]$. The integration yields

$$L_w(t) = - \int_t^T e^{-\int_t^s w_*(\xi) d\xi} v_*(s) f'_x(x_*(s), z_*(s)) L_v(s) ds. \tag{4.4}$$

Then, taking into consideration the constraints imposed on control $v_*(t)$ by (2.5), Assumption 1 and the already established inequality for $L_v(t)$, we immediately see that the corresponding inequality for function $L_w(t)$ in either (4.2) or (4.3) also necessary hold on the interval (t_v, T) . From the absolute continuity of switching function $L_w(t)$ and the Theorem on the Stability of the Sign of a Continuous Function it is immediately follows that there is value $t_w, t_w < t_v$, such that, depending on the sign of $(\alpha - \beta)$, the second inequality either in (4.2) or in (4.3) holds. This completes the proof. \square

Corollary 6. Lemmas 4 and 5 combined with formulas (3.9), (3.10), (3.12), (3.13) and (3.15), immediately lead to the following relationships for the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$:

(i) if $\alpha < \beta$, then

$$\begin{aligned} u_*(t) &= u_{\max}, & w_*(t) &= w_{\min} & \text{for } t \in (t_w, T]; \\ v_*(t) &= v_{\max} & & & \text{for } t \in (t_v, T]; \\ p_*(t) &= p_{\min}, & q_*(t) &= q_{\max} & \text{for } t \in (t_q, T]; \\ r_*(t) &= r_{\max} & & & \text{for } t \in (t_r, T]; \end{aligned} \tag{4.5}$$

(ii) if $\alpha > \beta$, then

$$\begin{aligned} u_*(t) &= u_{\min}, & w_*(t) &= w_{\max} & \text{for } t \in (t_w, T]; \\ v_*(t) &= v_{\min} & & & \text{for } t \in (t_v, T]; \\ p_*(t) &= p_{\min}, & q_*(t) &= q_{\max} & \text{for } t \in (t_q, T]; \\ r_*(t) &= r_{\max} & & & \text{for } t \in (t_r, T]. \end{aligned} \tag{4.6}$$

The following Lemma is of principal importance for further analysis.

Lemma 7. The switching functions $L_v(t)$, $L_w(t)$, $L_q(t)$ and $L_r(t)$ are not identically equal to zero on any subinterval of interval $[0, T]$.

Proof. Firstly, we consider switching function $L_v(t)$. Let us suppose the contrary, that is that there is an interval $\Delta_v = (t_0, t_1) \subset [0, T]$ on which

$$L_v(t) = 0. \tag{4.7}$$

Then

$$\tilde{L}_v(t) = 0, \tag{4.8}$$

almost everywhere on this interval as well and for $t \in \Delta_v$ equations of system (3.16) are

$$\begin{cases} 0 = (w_*(t) - r_*(t))L_w(t) + (p_*(t) + q_*(t) - r_*(t))L_q(t), \\ \dot{L}_w(t) = w_*(t)L_w(t), \\ \dot{L}_q(t) = q_*(t)L_q(t), \\ \dot{L}_r(t) = r_*(t)L_r(t) - p_*(t)L_q(t). \end{cases} \tag{4.9}$$

Hence,

$$L_w(t) = L_w^0 e^{\int_{t_0}^t w_*(\xi) d\xi}, \tag{4.10}$$

and

$$L_q(t) = L_q^0 e^{\int_{t_0}^t q_*(\xi) d\xi}, \tag{4.11}$$

where L_w^0, L_q^0 are some constants.

Now we have to consider the possible cases depending on the particular values of L_w^0 and L_q^0 .

Case 1. If $L_w^0 = 0$ and $L_q^0 = 0$, then, by (4.10) and (4.11),

$$L_w(t) = 0, \quad L_q(t) = 0 \tag{4.12}$$

everywhere on interval Δ_v . Substituting (4.7) and (4.12) into (3.14), we get

$$\psi_1^*(t) = 0, \quad \psi_2^*(t) = 0, \quad \psi_3^*(t) = 0 \tag{4.13}$$

everywhere on interval Δ_v . System (3.3) is linear and homogeneous, and hence, equalities (4.13) necessary hold on the entire interval $[0, T]$. This contradicts to the the boundary conditions at $t = T$ and, hence, this case is impossible.

Case 2. If $L_w^0 = 0$ and $L_q^0 \neq 0$, then, from (4.10) and (4.11),

$$L_w(t) = 0, L_q(t) \neq 0 \tag{4.14}$$

everywhere on interval Δ_v . Hence, by (3.12) and (3.15), on Δ_v the controls $p_*(t)$ and $q_*(t)$ simultaneously take either the values p_{\min} and q_{\max} , or the values p_{\max} and q_{\min} . Moreover, by the inequality from (4.14) and the last equation in (4.9), on the interval Δ_v the switching function $L_r(t)$ has at most one zero. (For analysis, it is sufficient that it is not equal to zero identically.) Hence, by (3.13), $r_* \in \{r_{\min}; r_{\max}\}$ on Δ_v . Substituting (4.14) and the above-found values of $p_*(t)$ and $q_*(t)$ into the first equality in (4.9), we conclude that either

$$p_{\min} + q_{\max} - r_* = 0, \tag{4.15}$$

or

$$p_{\max} + q_{\min} - r_* = 0 \tag{4.16}$$

necessary holds. However, either of these equalities contradicts to the control restrictions (2.9) and (2.10) of Assumption 2. Therefore, this case is also impossible.

Case 3. If $L_w^0 \neq 0$ and $L_q^0 = 0$, then, by (4.10) and (4.11),

$$L_w(t) \neq 0, L_q(t) = 0 \tag{4.17}$$

everywhere on the interval Δ_v . Hence, by (3.10), for all $t \in \Delta_v$, the control $w_*(t)$ takes either value w_{\min} , or value w_{\max} . Equality in (4.17) substituted into the last equation in (4.9) yields

$$L_r(t) = L_r^0 e^{\int_{t_0}^t r_*(\xi) d\xi}, \tag{4.18}$$

where L_r^0 is a constant. Here, the equality $L_r^0 = 0$ is impossible, since, by (3.14), in the case we come to the contradictory equalities (4.13). Therefore, $L_r^0 \neq 0$, and then we have to conclude that on the interval Δ_v the control $r_*(t)$ takes either the value r_{\min} or value r_{\max} . Substituting (4.17) and the above-found values of control $w_*(t)$ into the first equality of (4.9), we immediately conclude that either

$$w_{\min} = r_*, \tag{4.19}$$

or

$$w_{\max} = r_* \tag{4.20}$$

must hold. However, neither of these equalities is possible as contradicting to the control restrictions (2.7) and (2.8) of Assumption 2. Therefore, this case is also impossible.

Case 4. If $L_w^0 \neq 0$ and $L_q^0 \neq 0$, then, by (4.10) and (4.11),

$$L_w(t) \neq 0, L_q(t) \neq 0 \tag{4.21}$$

everywhere on interval Δ_v . Hence, by (3.10) and (3.12), for all $t \in \Delta_v$, control $w_*(t)$ takes either value w_{\min} , or value w_{\max} , and controls $p_*(t)$ and $q_*(t)$ simultaneously take either values p_{\min} and q_{\max} , or values p_{\max} and q_{\min} . By the same arguments as in Case 2, we have to conclude that control $r_*(t)$ takes either value r_{\min} or value r_{\max} on the entire interval Δ_v or on its subinterval Δ'_v .

Differentiating the first equality of (4.9) on the interval Δ_v or on its subinterval Δ'_v , and then using the corresponding differential equations, we get equality

$$w_*(w_* - r_*)L_w(t) + q_*(p_* + q_* - r_*)L_q(t) = 0, \quad t \in \Delta_v (t \in \Delta'_v), \tag{4.22}$$

where $w_* \in \{w_{\min}; w_{\max}\}$, $(p_* + q_*) \in \{p_{\min} + q_{\max}; p_{\max} + q_{\min}\}$ and $r_* \in \{r_{\min}; r_{\max}\}$. The first equality in (4.9) and (4.22) form a linear homogeneous system with respect to unknown $L_w(t)$ and $L_q(t)$. By (4.21), the solution of this system must be non-trivial, and, hence, the determinant of this system must be equal to zero. That is, the equality

$$(w_* - r_*)(p_* + q_* - r_*)(w_* - q_*) = 0, \tag{4.23}$$

must hold. However, by the control restrictions (2.7)–(2.10) of Assumption 2, this equality is impossible. Therefore, this case is also impossible.

As all possible cases lead to a contradiction, we have to conclude that the hypothesis was incorrect and that the switching function $L_v(t)$ cannot be identically equal to zero on any subinterval of the interval $[0, T]$.

Now we have to consider whether switching function $L_w(t)$ can be equal to zero on a subinterval of the interval $[0, T]$. Let us assume that there is an interval $\Delta_w \subset [0, T]$ on which

$$L_w(t) = 0. \tag{4.24}$$

Then, almost everywhere on this interval

$$\dot{L}_w(t) = 0, \tag{4.25}$$

holds as well. Substituting (4.24) and (4.25) into the second equation of system (3.16) and using the constraints on control $v_*(t)$ from (2.5) and the properties of function $f(x, z)$ from Assumption 1, we have to conclude that the contradictory equality $L_v(t) = 0$ holds for all $t \in \Delta_w$. Hence, our assumption is incorrect and the switching function $L_w(t)$ cannot be identically equal to zero on any subinterval of the interval $[0, T]$.

Next, we consider a possibility that the switching function $L_q(t)$ is identically equal to zero on any subinterval of the interval $[0, T]$. Applying the same arguments as those for the function $L_w(t)$, but using the third equation of system (3.16) instead of the second equation, we come to the conclusion that the switching function $L_q(t)$ cannot be identically equal to zero on any subinterval as well.

Finally, we consider the switching function $L_r(t)$. Let us suppose the contrary and assume that there is the interval $\Delta_r \subset [0, T]$ on which

$$L_r(t) = 0. \tag{4.26}$$

Then, almost everywhere on this interval

$$\dot{L}_r(t) = 0, \tag{4.27}$$

holds. Again, we substitute (4.26) and (4.27) into the fourth equation of system (3.16). Using the constraints on control $p_*(t)$ from (2.5) we find the contradictory equality $L_q(t) = 0$ for $t \in \Delta_r$. Hence the above-made assumption was incorrect and the switching function $L_r(t)$ is not identically equal to zero on any subinterval of the interval $[0, T]$. This completes the proof. \square

Lemma 7 combined with equalities (3.8)–(3.13) and (3.15) immediately leads to the following Corollary.

Corollary 8. *The optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$ are the bang-bang functions taking values $\{u_{\min}; u_{\max}\}$, $\{v_{\min}; v_{\max}\}$, $\{w_{\min}; w_{\max}\}$, $\{p_{\min}; p_{\max}\}$, $\{q_{\min}; q_{\max}\}$ and $\{r_{\min}; r_{\max}\}$, respectively. Moreover, each of the pairs composed of controls $u_*(t)$ and $w_*(t)$, and $p_*(t)$ and $q_*(t)$ switches from the minimum values to the maximum values and vice versa at the same switching moments, which are specific for each of these pairs.*

In addition to this corollary, we note that the controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$ have no singular arcs (see [23, 24] for details).

5. Estimating the Number of Zeros of the Switching Functions

In previous section, we proved that the optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$ are the bang-bang functions with values $\{u_{\min}; u_{\max}\}$, $\{v_{\min}; v_{\max}\}$, $\{w_{\min}; w_{\max}\}$, $\{p_{\min}; p_{\max}\}$, $\{q_{\min}; q_{\max}\}$ and $\{r_{\min}; r_{\max}\}$, respectively. In this section, we proceed to estimating the number of switchings of these controls. Equalities (3.8)–(3.13) and (3.15) show that for this task it suffices to estimate the number of zeros of the switching functions $L_v(t)$, $L_w(t)$, $L_q(t)$ and $L_r(t)$ defined by system (3.16). From the analysis of the second and third equations of this system, we conclude that if we have an estimate for the number of zeros of function $L_v(t)$, then, applying the Generalized Rolle’s Theorem (see, e.g., [53]) to each of these equations, we can find estimates for the number of zeros of functions $L_w(t)$ and $L_q(t)$ as well. The use of this theorem with the fourth equation of system (3.16) gives us an estimate of the number of zeros of function $L_r(t)$. Thus, the analysis of switching function $L_v(t)$ is the basis for completing our task. The method that allows to estimate the number of zeros of the function $L_v(t)$ consists in the subsequent application of the above-mentioned Generalized Rolle’s Theorem to the equations of system (3.16).

Let us consider a possibility of application of the Generalized Rolle’s Theorem to the first equation of system (3.16). The direct application of the theorem to this equation is impossible, because, apart of the term containing the function $L_v(t)$, the equation has two terms independent of $L_v(t)$. Therefore, we introduce an auxiliary function

$$G_v(t) = (w_*(t) - r_*(t))L_w(t) + (p_*(t) + q_*(t) - r_*(t))L_q(t), \tag{5.1}$$

which, by Lemma 7 and relationships (2.7)–(2.10) of Assumption 2, is not equal to zero on any subinterval of $[0, T]$. Then, the first equation of system (3.16) can be rewritten as

$$\dot{L}_v(t) = (v_*(t)(f'_x(x_*(t), z_*(t)) + f'_z(x_*(t), z_*(t))) + r_*(t))L_v(t) + G_v(t). \tag{5.2}$$

In order to formulate the differential equation for function $G_v(t)$, we note that the switching functions $L_w(t)$ and $L_q(t)$ are absolutely continuous (as the solutions of system (3.16) and, therefore, are almost everywhere differentiable functions on the interval $[0, T]$ [63]. Theorem 3 of Chapter 4 in [60], gives us absolute continuity and hence, also almost everywhere differentiability of functions $w_*(t)L_w(t)$ and $(p_*(t) + q_*(t))L_q(t)$. Finally, arguments similar to these used by the authors in [34] allow to conclude that the functions $(w_*(t) - r_*(t))$ and $(p_*(t) + q_*(t) - r_*(t))$ are also differentiable almost everywhere on the interval $[0, T]$. Then, using the second and third equations of system (3.16), we formulate the required differential equation:

$$\begin{aligned} \dot{G}_v(t) = & v_*(t)((w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + (p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t)))L_v(t) \\ & + w_*(t)(w_*(t) - r_*(t))L_w(t) + q_*(t)(p_*(t) + q_*(t) - r_*(t))L_q(t). \end{aligned} \tag{5.3}$$

Let us introduce another auxiliary function

$$Q_v(t) = w_*(t)(w_*(t) - r_*(t))L_w(t) + q_*(t)(p_*(t) + q_*(t) - r_*(t))L_q(t). \tag{5.4}$$

Then, equation (5.3) can be rewritten as follows:

$$\begin{aligned} \dot{G}_v(t) = & v_*(t)((w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + (p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t)))L_v(t) + Q_v(t). \end{aligned} \tag{5.5}$$

By the above-presented arguments, function $Q_v(t)$ is also almost everywhere differentiable and satisfies

$$\begin{aligned} \dot{Q}_v(t) = & v_*(t)(w_*(t)(w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + q_*(t)(p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t)))L_v(t) \\ & + w_*^2(t)(w_*(t) - r_*(t))L_w(t) + q_*^2(t)(p_*(t) + q_*(t) - r_*(t))L_q(t). \end{aligned} \tag{5.6}$$

The last two terms of this equation can be expressed in terms of the auxiliary functions $G_v(t)$ and $Q_v(t)$. Then,

$$\begin{aligned} \dot{Q}_v(t) = & v_*(t)(w_*(t)(w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + q_*(t)(p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t)))L_v(t) \\ & - w_*(t)q_*(t)G_v(t) + (w_*(t) + q_*(t))Q_v(t). \end{aligned} \tag{5.7}$$

Let us denote

$$\begin{aligned} a_v(t) = & v_*(t)(f'_x(x_*(t), z_*(t)) + f'_z(x_*(t), z_*(t))) + r_*(t), \\ b_v(t) = & v_*(t)((w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + (p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t))), \\ c_v(t) = & v_*(t)(w_*(t)(w_*(t) - r_*(t))f'_x(x_*(t), z_*(t)) \\ & + q_*(t)(p_*(t) + q_*(t) - r_*(t))f'_z(x_*(t), z_*(t))), \\ d_v(t) = & w_*(t)q_*(t), e_v(t) = w_*(t) + q_*(t). \end{aligned} \tag{5.8}$$

Due to the constraints on the controls $v_*(t)$, $w_*(t)$, $q_*(t)$ and $r_*(t)$ in (2.5) and the properties of function $f(x, z)$ in Assumption 1, we conclude that functions $a_v(t)$, $d_v(t)$ and $e_v(t)$ are positive almost everywhere on the interval $[0, T]$. Moreover, functions $d_v(t)$ and $e_v(t)$ are piecewise-constant and functions $a_v(t)$, $b_v(t)$ and $c_v(t)$ are piecewise-continuous. Combining equations (5.2), (5.5) and (5.7), we write down the following system of differential equations for the switching function $L_v(t)$ and auxiliary functions $G_v(t)$ and $Q_v(t)$, which is more convenient for the subsequent analysis than system (3.16):

$$\begin{cases} \dot{L}_v(t) = a_v(t)L_v(t) + G_v(t), & t \in [0, T], \\ \dot{G}_v(t) = b_v(t)L_v(t) + Q_v(t), \\ \dot{Q}_v(t) = c_v(t)L_v(t) - d_v(t)G_v(t) + e_v(t)Q_v(t). \end{cases} \tag{5.9}$$

Let us rewrite system (5.9) in the following matrix form:

$$\begin{pmatrix} \dot{L}_v(t) \\ \dot{G}_v(t) \\ \dot{Q}_v(t) \end{pmatrix} = \begin{pmatrix} a_v(t) & 1 & 0 \\ b_v(t) & 0 & 1 \\ c_v(t) & -d_v(t) & e_v(t) \end{pmatrix} \begin{pmatrix} L_v(t) \\ G_v(t) \\ Q_v(t) \end{pmatrix}. \tag{5.10}$$

Now, using a special change of variables, we reduce the matrix of system (5.10) to the upper-triangular form. This

transformation is fairly standard, and is described in detail in Appendix A. Comparing systems (5.10) and (A.1) and applying arguments from Appendix A, we find the desired system in the form

$$\begin{cases} \dot{\tilde{L}}_v(t) = (a_v(t) - h_{21}(t))\tilde{L}_v(t) + \tilde{G}_v(t), \\ \dot{\tilde{G}}_v(t) = (h_{21}(t) - h_{32}(t))\tilde{G}_v(t) + \tilde{Q}_v(t), \\ \dot{\tilde{Q}}_v(t) = (h_{32}(t) + e_v(t))\tilde{Q}_v(t), \end{cases} \quad (5.11)$$

where $\tilde{L}_v(t) = L_v(t)$ is the switching function, and $\tilde{G}_v(t), \tilde{Q}_v(t)$ are the new auxiliary functions corresponding to $L_v(t)$. According to the arguments of Appendix A, the functions $h_{20}(t), h_{21}(t), h_{31}(t)$ and $h_{32}(t)$ perform such a transformation and, by (A.21) and (A.26), satisfy the following systems of differential equations:

$$\begin{cases} \dot{h}_{20}(t) = c_v(t) + (e_v(t) - a_v(t))h_{20}(t) + d_v(t)h_{21}(t) + h_{20}(t)h_{21}(t), \\ \dot{h}_{21}(t) = -b_v(t) - h_{20}(t) - a_v(t)h_{21}(t) + h_{21}^2(t), \\ h_{20}(0) = h_{20}^0 = h_{32}^0 h_{21}^0 - h_{31}^0, h_{21}(0) = h_{21}^0, \end{cases} \quad (5.12)$$

$$\begin{cases} \dot{h}_{31}(t) = -c_v(t) + (e_v(t) - a_v(t))h_{31}(t) - b_v(t)h_{32}(t) + h_{31}(t)h_{32}(t), \\ \dot{h}_{32}(t) = d_v(t) - h_{31}(t) + e_v(t)h_{32}(t) + h_{32}^2(t), \\ h_{31}(0) = h_{31}^0, h_{32}(0) = h_{32}^0. \end{cases} \quad (5.13)$$

Solutions $h_{20}(t), h_{21}(t)$ of system (5.12) and solutions $h_{31}(t), h_{32}(t)$ of system (5.13) are determined on the maximal intervals of existence of such solutions [60, 64]: specifically, we assume that functions $h_{20}(t)$ and $h_{21}(t)$ are determined on the interval $\Delta_2 = (t_0^2, t_1^2)$, and functions $h_{31}(t)$ and $h_{32}(t)$ are determined on the interval $\Delta_3 = (t_0^3, t_1^3)$. Without loss of generality, we suppose that $t_1^2 \leq T$ and $t_1^3 \leq T$.

As we noted in Appendix A, we need to find the restrictions on the functions $a_v(t), b_v(t), c_v(t), d_v(t), e_v(t)$ and initial conditions $h_{21}^0, h_{31}^0, h_{32}^0$, under which systems (5.12) and (5.13) have solutions defined on the entire interval $[0, T]$. To do this, using the fact that $d_v(t)$ and $e_v(t)$ are piecewise-constant functions, we transform the equations of systems (5.12) and (5.13) as follows:

$$\begin{cases} \frac{d}{dt}(h_{20}(t) + d_v(t)) = -(h_{20}(t) + d_v(t))(-h_{21}(t)) \\ \quad + (e_v(t) - a_v(t))(h_{20}(t) + d_v(t)) \\ \quad + (c_v(t) - (e_v(t) - a_v(t))d_v(t)), \\ \frac{d}{dt}(-h_{21}(t)) = -(-h_{21}(t))^2 + (h_{20}(t) + d_v(t)) \\ \quad - a_v(t)(-h_{21}(t)) + (b_v(t) - d_v(t)), \end{cases} \quad (5.14)$$

$$\begin{cases} \frac{d}{dt}(h_{31}(t) - k_v(t)) = -(h_{31}(t) - k_v(t))(-h_{32}(t) - 0.5e_v(t)) \\ \quad + (0.5e_v(t) - a_v(t))(h_{31}(t) - k_v(t)) \\ \quad + (b_v(t) - k_v(t))(-h_{32}(t) - 0.5e_v(t)) \\ \quad - (c_v(t) - 0.5(b_v(t) + k_v(t))e_v(t) + a_v(t)k_v(t)), \\ \frac{d}{dt}(-h_{32}(t) - 0.5e_v(t)) = -(-h_{32}(t) - 0.5e_v(t))^2 + (h_{31}(t) - k_v(t)). \end{cases} \quad (5.15)$$

Here $k_v(t) = d_v(t) - 0.25e_v^2(t)$ is a piecewise-constant function as well, and, as it is not difficult to see, the inequality $k_v(t) < 0$

holds for almost all $t \in [0, T]$. Performing in these systems the change of variables

$$\begin{aligned} g_{20}(t) &= h_{20}(t) + d_v(t), & g_{21}(t) &= -h_{21}(t), \\ g_{31}(t) &= h_{31}(t) - k_v(t), & g_{32}(t) &= -h_{32}(t) - 0.5e_v(t), \end{aligned} \quad (5.16)$$

and introducing values

$$g_{21}^0 = -h_{21}^0, g_{31}^0 = h_{31}^0 - k_v(+0), g_{32}^0 = -h_{32}^0 - 0.5e_v(+0), \quad (5.17)$$

where $e_v(+0) = \lim_{t \rightarrow +0} e_v(t)$ and $k_v(+0) = \lim_{t \rightarrow +0} k_v(t)$, we obtain the following systems of equations:

$$\begin{cases} \dot{g}_{20}(t) = \Phi_{20}(t, g_{20}(t), g_{21}(t)) \\ \quad = -g_{20}(t)g_{21}(t) + (e_v(t) - a_v(t))g_{20}(t) \\ \quad \quad + (c_v(t) - (e_v(t) - a_v(t))d_v(t)), \\ \dot{g}_{21}(t) = \Phi_{21}(t, g_{20}(t), g_{21}(t)) \\ \quad = -g_{21}^2(t) + g_{20}(t) - a_v(t)g_{21}(t) + (b_v(t) - d_v(t)), \\ g_{20}(0) = (g_{32}^0 + e_v(+0))g_{21}^0 - g_{31}^0, g_{21}(0) = g_{21}^0, \end{cases} \quad (5.18)$$

$$\begin{cases} \dot{g}_{31}(t) = \Phi_{31}(t, g_{31}(t), g_{32}(t)) \\ \quad = -g_{31}(t)g_{32}(t) + (0.5e_v(t) - a_v(t))g_{31}(t) \\ \quad \quad + (b_v(t) - k_v(t))g_{32}(t) - (c_v(t) - 0.5(b_v(t) \\ \quad \quad + k_v(t))e_v(t) + a_v(t)k_v(t)), \\ \dot{g}_{32}(t) = \Phi_{32}(t, g_{31}(t), g_{32}(t)) = -g_{32}^2(t) + g_{31}(t), \\ g_{31}(0) = g_{31}^0, g_{32}(0) = g_{32}^0. \end{cases} \quad (5.19)$$

Let us apply to systems (5.18) and (5.19) with nonnegative initial conditions

$$(g_{32}^0 + e_v(+0))g_{21}^0 - g_{31}^0 \geq 0, g_{21}^0 \geq 0, g_{31}^0 \geq 0, g_{32}^0 \geq 0, \quad (5.20)$$

the necessary and sufficient condition for solutions $g_{20}(t), g_{21}(t)$ and $g_{31}(t), g_{32}(t)$ to be nonnegative for all $t \in [0, t_1^2)$ and $t \in [0, t_1^3)$, respectively [65]. It consists in satisfying the so-called quasi-positivity condition

$$\begin{cases} \Phi_{20}(t, 0, g_{21}) \geq 0, & \text{for all } g_{21} \geq 0, \\ \Phi_{21}(t, g_{20}, 0) \geq 0, & \text{for all } g_{20} \geq 0, \end{cases} \quad (5.21)$$

for system (5.18), and

$$\begin{cases} \Phi_{31}(t, 0, g_{32}) \geq 0, & \text{for all } g_{32} \geq 0, \\ \Phi_{32}(t, g_{31}, 0) \geq 0, & \text{for all } g_{31} \geq 0, \end{cases} \quad (5.22)$$

for system (5.19).

As a result, we obtain the inequalities:

$$\begin{cases} b_v(t) \geq d_v(t), \\ a_v(t)d_v(t) + c_v(t) \geq d_v(t)e_v(t), \\ a_v(t)k_v(t) + c_v(t) \leq 0.5(b_v(t) + k_v(t))e_v(t), \end{cases} \quad (5.23)$$

which hold for almost all $t \in [0, T]$. These inequalities define the required restrictions on the functions $a_v(t), b_v(t), c_v(t), d_v(t)$ and $e_v(t)$. Moreover, as it is easy to see choosing appropriate values g_{21}^0, g_{31}^0 and g_{32}^0 , inequalities (5.20) are satisfied as well.

We consider the following assumption, the validity of which will be demonstrated in subsequent arguments.

Assumption 9. Let the system of inequalities

$$\begin{cases} v(t)((w(t) - r(t))f'_x(x(t), z(t)) + (p(t) + q(t) \\ - r(t))f'_z(x(t), z(t))) \geq w(t)q(t), \\ w(t)q(t)(v(t)(f'_x(x(t), z(t)) + f'_z(x(t), z(t))) + r(t)) \\ + v(t)(w(t)(w(t) - r(t))f'_x(x(t), z(t)) \\ + q(t)(p(t) + q(t) - r(t))f'_z(x(t), z(t))) \geq w(t)q(t)(w(t) + q(t)), \\ (w(t) - q(t))^2(v(t)(f'_x(x(t), z(t)) + f'_z(x(t), z(t))) \\ - 0.5((w(t) - r(t)) + (q(t) - r(t)))) \\ + 4v(t)(1 - 2w(t))(w(t) - r(t))f'_x(x(t), z(t)) + (1 - 2q(t)) \\ (p(t) + q(t) - r(t))f'_z(x(t), z(t)) \geq 0 \end{cases} \quad (5.24)$$

hold almost everywhere on the interval $[0, T]$ for arbitrary admissible controls $u(t), v(t), w(t), p(t), q(t)$ and $r(t)$ and the components $x(t), y(t)$ and $z(t)$ of the corresponding solution $(x(t), y(t), z(t))$ to system (2.11).

By definitions of functions $a_v(t), b_v(t), c_v(t), d_v(t), e_v(t)$ and $k_v(t)$, it is easy to see that under Assumption 9 the inequalities (5.23) hold. The verification of this assumption for the case of a bilinear function $f(x, z) = \kappa xz$, where κ is a positive constant, is discussed in Appendix B.

Finally, we have to verify the continuation of the nonnegative solutions $g_{20}(t), g_{21}(t), g_{31}(t)$ and $g_{32}(t)$ of systems (5.18) and (5.19) to the entire interval $[0, T]$ (see Theorem 1.6 in [65]). To do this, we define Lyapunov functions:

$$V_2(g_{20}, g_{21}) = 0.5(g_{20}^2 + g_{21}^2), \quad V_3(g_{31}, g_{32}) = 0.5(g_{31}^2 + g_{32}^2), \quad (5.25)$$

and estimate the derivatives dV_2/dt and dV_3/dt of these functions along the trajectories of the systems of differential equations (5.18) and (5.19), respectively. (The derivatives are the scalar products of the gradients of these functions and the right-hand sides of the corresponding systems.) The Lyapunov functions satisfy

$$\begin{aligned} \frac{dV_2}{dt}(t) &= -g_{21}(t)(g_{20}^2(t) + g_{21}^2(t)) + (e_v(t) - a_v(t))g_{20}^2(t) \\ &+ g_{20}(t)g_{21}(t) - a_v(t)g_{21}^2(t) + (c_v(t) - (e_v(t) - a_v(t))d_v(t))g_{20}(t) \\ &+ (b_v(t) - d_v(t))g_{21}(t) + (c_v(t) - (e_v(t) - a_v(t))d_v(t))g_{20}(t) \\ &+ (b_v(t) - d_v(t))g_{21}(t) \leq K_2V_2(g_{20}(t), g_{21}(t)). \end{aligned} \quad (5.26)$$

$$\begin{aligned} \frac{dV_3}{dt}(t) &= -g_{32}(t)(g_{31}^2(t) + g_{32}^2(t)) + (0.5e_v(t) - a_v(t))g_{31}^2(t) \\ &+ (b_v(t) - k_v(t) + 1)g_{31}(t)g_{32}(t) - (c_v(t) - 0.5(b_v(t) \\ &+ k_v(t))e_v(t) + a_v(t)k_v(t))g_{31}(t) \leq K_3V_3(g_{31}(t), g_{32}(t)). \end{aligned} \quad (5.27)$$

Here, K_2 and K_3 are constants defined as following:

$$K_2 = 2 \max_{t \in [0, T]} \{a_v(t) + e_v(t); 1; |b_v(t)| + d_v(t); |c_v(t)| + (a_v(t) + e_v(t))d_v(t)\}, \quad (5.28)$$

$$K_3 = 2 \max_{t \in [0, T]} \{a_v(t) + 0.5e_v(t); |b_v(t)| + |k_v(t)| + 1; |c_v(t)| + 0.5(|b_v(t)| + |k_v(t)|)e_v(t) + a_v(t)|k_v(t)|\}. \quad (5.29)$$

We have to note that these inequalities are obtained under constraints

$$g_{20}^2(t) + g_{21}^2(t) \geq 1, \quad g_{31}^2(t) + g_{32}^2(t) \geq 1, \quad (5.30)$$

which follow from the relationships

$$\lim_{t \rightarrow t_1^+ - 0} \|(g_{20}(t), g_{21}(t))^T\| = +\infty, \quad \lim_{t \rightarrow t_1^+ - 0} \|(g_{31}(t), g_{32}(t))^T\| = +\infty, \quad (5.31)$$

(see Lemma in Section 14, Chapter 4 in [66]). That is, the estimates

$$\frac{dV_2}{dt}(t) \leq K_2V_2(g_{20}(t), g_{21}(t)), \quad \frac{dV_3}{dt}(t) \leq K_3V_3(g_{31}(t), g_{32}(t)), \quad (5.32)$$

imply the required continuation of the nonnegative solutions $g_{20}(t)$ and $g_{21}(t)$ of system (5.18) and the nonnegative solutions $g_{31}(t)$ and $g_{32}(t)$ of system (5.19) to the entire interval $[0, T]$. Taking into consideration that functions $h_{20}(t), h_{21}(t), h_{31}(t)$ and $h_{32}(t)$ satisfy systems (5.12) and (5.13) and are also defined on the entire interval $[0, T]$, we conclude that system (5.11) is defined on this interval as well.

Using system (5.11), we can estimate the number of zeros of the switching function $L_v(t) = L_v(t)$. Taking into consideration that $L_v(T) = \beta - \alpha \neq 0$ (see (3.16)), we will show that this switching function has at most two distinct zeros on the interval $[0, T]$.

Let us suppose the contrary, that is that the function $L_v(t) = L_v(t)$ has at least three distinct zeroes on this interval. Then, by virtue of the Generalized Rolle's Theorem [53] applied to the first equation of system (5.11), the function $G_v(t)$ has at least two distinct zeroes on the interval $(0, T)$. Applying this theorem to the second equation of system (5.11), we have to conclude that the function $Q_v(t)$ has at least one zero on the interval $(0, T)$. However, this function satisfies the linear homogeneous equation, and, therefore, if $Q_v(t) = 0$ at a single point of interval $[0, T]$, then $Q_v(t) = 0$ on the entire interval $[0, T]$. Moreover, in this case the second equation of system (5.11) is also the linear homogeneous equation, and, hence, if function $G_v(t)$ has at least two zeros on interval $[0, T]$, then $G_v(t) = 0$ everywhere on the interval $[0, T]$ as well. Finally, applying the same line of arguments to the first equation of system (5.11), we have to conclude that in this case $L_v(t) = 0$ everywhere on the entire interval $[0, T]$. However, this contradicts to Lemma 7 and, hence, the assumption is incorrect. Thus, the required estimate for the number of zeros of the switching function $L_v(t) = \tilde{L}_v(t)$ is established.

Denote the zeros of function $L_v(t)$ by τ_1^* and τ_2^* ; $\tau_1^*, \tau_2^* \in [0, T]$. Depending on the sign of the difference $(\beta - \alpha)$, the function $L_v(t)$ is described by one of the following two relationships, namely:

(i) If $\beta - \alpha > 0$, then

$$L_v(t) = \begin{cases} > 0, & \text{if } 0 \leq t < \tau_1^*, \\ = 0, & \text{if } t = \tau_1^*, \\ < 0, & \text{if } \tau_1^* < t < \tau_2^*, \\ = 0, & \text{if } t = \tau_2^*, \\ > 0, & \text{if } \tau_2^* < t \leq T; \end{cases} \quad (5.33)$$

(ii) If $\beta - \alpha < 0$, then

$$L_v(t) = \begin{cases} < 0, & \text{if } 0 \leq t < \tau_1^*, \\ = 0, & \text{if } t = \tau_1^*, \\ > 0, & \text{if } \tau_1^* < t < \tau_2^*, \\ = 0, & \text{if } t = \tau_2^*, \\ < 0, & \text{if } \tau_2^* < t \leq T. \end{cases} \quad (5.34)$$

Please note that these formulas completely agree with the corresponding inequalities from (4.2) and (4.3) of Lemma 5.

Now we proceed to estimating the number of zeros of the switching function $L_w(t)$. We will show that, taking into consideration that $L_w(T) = 0$ (see (3.16)), function $L_w(t)$ has at most three distinct zeros on the interval $[0, T]$. To prove this proposition, we, again, suppose the opposite: that is, we suppose that function $L_w(t)$ has at least four distinct zeroes on interval $[0, T]$. Applying the Generalized Rolle's Theorem, we have to conclude that in this case function $L_v(t)$ has at least three distinct zeroes on the interval $(0, T)$, which contradicts to the above-found estimates for the number of zeros of switching function $L_v(t)$. Therefore, our assumption was incorrect, and the required estimate for the number of zeros of the switching function $L_w(t)$ is found.

Since $L_w(T) = 0$ (see (3.16)), one of these three zeros is the end point of the interval $[0, T]$. We denote the other two zeros by θ_1^* and θ_2^* ; $\theta_1^*, \theta_2^* \in (0, T)$. Then, taking into consideration the corresponding inequalities from Lemma 5, we conclude that, depending on the sign of $(\beta - \alpha)$, the switching function $L_w(t)$ is described by one of the following two relationships, namely:

(i) if $\beta - \alpha > 0$, then

$$L_w(t) = \begin{cases} < 0, & \text{if } 0 \leq t < \theta_1^*, \\ = 0, & \text{if } t = \theta_1^*, \\ > 0, & \text{if } \theta_1^* < t < \theta_2^*, \\ = 0, & \text{if } t = \theta_2^*, \\ < 0, & \text{if } \theta_2^* < t < T, \\ = 0, & \text{if } t = T; \end{cases} \quad (5.35)$$

(ii) If $\beta - \alpha < 0$, then

$$L_w(t) = \begin{cases} > 0, & \text{if } 0 \leq t < \theta_1^*, \\ = 0, & \text{if } t = \theta_1^*, \\ < 0, & \text{if } \theta_1^* < t < \theta_2^*, \\ = 0, & \text{if } t = \theta_2^*, \\ > 0, & \text{if } \theta_2^* < t < T, \\ = 0, & \text{if } t = T. \end{cases} \quad (5.36)$$

To estimate the number of zeros of the switching function $L_q(t)$, we take into consideration the boundary condition $L_q(T) = \beta > 0$ (see (3.16)). Applying the arguments that are analogous to those used to estimate the number of zeros of function $L_w(t)$, we can prove that the function $L_q(t)$ has at most three distinct zeros on the interval $[0, T]$. We omit the arguments, which are the same as the above-used. Taking into consideration the first inequality from (4.1) in Lemma 4, we have to conclude that the switching function $L_q(t)$ satisfies

$$L_q(t) = \begin{cases} < 0, & \text{if } 0 \leq t < \eta_1^*, \\ = 0, & \text{if } t = \eta_1^*, \\ > 0, & \text{if } \eta_1^* < t < \eta_2^*, \\ = 0, & \text{if } t = \eta_2^*, \\ < 0, & \text{if } \eta_2^* < t < \eta_3^*, \\ = 0, & \text{if } t = \eta_3^*, \\ > 0, & \text{if } \eta_3^* < t \leq T. \end{cases} \quad (5.37)$$

Here $\eta_1^*, \eta_2^*, \eta_3^* \in [0, T)$ are zeros of the switching function $L_q(t)$.

Please note that, by (3.15), relationships (5.35)–(5.37) also describe the behavior of the switching functions $L_u(t)$ and $L_p(t)$.

Finally, we estimate the number of zeros of switching function $L_r(t)$. Taking into consideration the boundary condition $L_r(T) = \alpha > 0$ (see (3.16)), we show that function $L_r(t)$ has at most four distinct zeros on interval $[0, T]$. Let us prove this by contradiction: assume that the function $L_r(t)$ has at least five distinct zeros on the interval $[0, T)$. Applying the Generalized Rolle's Theorem, we conclude that in this case the function $L_q(t)$ has at least four distinct zeros on the interval $(0, T)$. This conclusion contradicts to the above-established estimate, and, therefore, our assumption is incorrect.

Denoting the zeros of function $L_r(t)$ by $\chi_1^*, \chi_2^*, \chi_3^*$ and χ_4^* ; $\chi_1^*, \chi_2^*, \chi_3^*, \chi_4^* \in (0, T)$, and taking into consideration the second inequality from (4.1) in Lemma 4, we get the following relationship describing behavior of the switching function $L_r(t)$:

$$L_r(t) = \begin{cases} > 0, & \text{if } 0 \leq t < \chi_1^*, \\ = 0, & \text{if } t = \chi_1^*, \\ < 0, & \text{if } \chi_1^* < t < \chi_2^*, \\ = 0, & \text{if } t = \chi_2^*, \\ > 0, & \text{if } \chi_2^* < t < \chi_3^*, \\ = 0, & \text{if } t = \chi_3^*, \\ < 0, & \text{if } \chi_3^* < t < \chi_4^*, \\ = 0, & \text{if } t = \chi_4^*, \\ > 0, & \text{if } \chi_4^* < t \leq T. \end{cases} \quad (5.38)$$

6. Types of the Optimal Controls

The above-found qualitative profiles of the switching functions (5.33)–(5.38) combined with the formulas of the optimal controls (3.9), (3.10) and (3.12), (3.13) enable us to immediately find the possible types of the optimal controls $v_*(t)$, $w_*(t)$, $q_*(t)$ and $r_*(t)$. Specifically, depending on the sign of $(\beta - \alpha)$, the optimal controls $v_*(t)$ and $w_*(t)$ are of one of the following types:

(i) If $\beta - \alpha > 0$, then

$$v_*(t) = \begin{cases} v_{\max}, & \text{if } 0 \leq t \leq \tau_1^*, \\ v_{\min}, & \text{if } \tau_1^* < t \leq \tau_2^*, \\ v_{\max}, & \text{if } \tau_2^* < t \leq T, \end{cases} \quad (6.1)$$

$$w_*(t) = \begin{cases} w_{\min}, & \text{if } 0 \leq t \leq \theta_1^*, \\ w_{\max}, & \text{if } \theta_1^* < t \leq \theta_2^*, \\ w_{\min}, & \text{if } \theta_2^* < t \leq T. \end{cases}$$

(ii) If $\beta - \alpha < 0$, then

$$v_*(t) = \begin{cases} v_{\min}, & \text{if } 0 \leq t \leq \tau_1^*, \\ v_{\max}, & \text{if } \tau_1^* < t \leq \tau_2^*, \\ v_{\min}, & \text{if } \tau_2^* < t \leq T, \end{cases} \quad (6.2)$$

$$w_*(t) = \begin{cases} w_{\max}, & \text{if } 0 \leq t \leq \theta_1^*, \\ w_{\min}, & \text{if } \theta_1^* < t \leq \theta_2^*, \\ w_{\max}, & \text{if } \theta_2^* < t \leq T. \end{cases}$$

Here, $\tau_1^*, \tau_2^* \in [0, T]$ and $\theta_1^*, \theta_2^* \in [0, T]$ are the switchings of the optimal controls $v_*(t)$ and $w_*(t)$, respectively.

The optimal control $q_*(t)$ is of the following type:

$$q_*(t) = \begin{cases} q_{\min}, & \text{if } 0 \leq t \leq \eta_1^*, \\ q_{\max}, & \text{if } \eta_1^* < t \leq \eta_2^*, \\ q_{\min}, & \text{if } \eta_2^* < t \leq \eta_3^*, \\ q_{\max}, & \text{if } \eta_3^* < t \leq T, \end{cases} \quad (6.3)$$

where $\eta_1^*, \eta_2^*, \eta_3^* \in [0, T]$ are the moments of switching.

The optimal control $r_*(t)$ is described as following:

$$r_*(t) = \begin{cases} r_{\max}, & \text{if } 0 \leq t \leq \chi_1^*, \\ r_{\min}, & \text{if } \chi_1^* < t \leq \chi_2^*, \\ r_{\max}, & \text{if } \chi_2^* < t \leq \chi_3^*, \\ r_{\min}, & \text{if } \chi_3^* < t \leq \chi_4^*, \\ r_{\max}, & \text{if } \chi_4^* < t \leq T, \end{cases} \quad (6.4)$$

where $\chi_1^*, \chi_2^*, \chi_3^*, \chi_4^* \in [0, T]$ are the moments of switching.

Taking into consideration (3.15), we can immediately describe the type of controls $u_*(t)$ and $p_*(t)$ as well. Specifically, depending on the sign of the difference $(\beta - \alpha)$, the optimal control $u_*(t)$ belongs to one of the following two types:

(i) If $\beta - \alpha > 0$, then

$$u_*(t) = \begin{cases} u_{\max}, & \text{if } 0 \leq t \leq \theta_1^*, \\ u_{\min}, & \text{if } \theta_1^* < t \leq \theta_2^*, \\ u_{\max}, & \text{if } \theta_2^* < t \leq T. \end{cases} \quad (6.5)$$

(ii) If $\beta - \alpha < 0$, then

$$u_*(t) = \begin{cases} u_{\min}, & \text{if } 0 \leq t \leq \theta_1^*, \\ u_{\max}, & \text{if } \theta_1^* < t \leq \theta_2^*, \\ u_{\min}, & \text{if } \theta_2^* < t \leq T. \end{cases} \quad (6.6)$$

The optimal control $p_*(t)$ is of the following qualitative type:

$$p_*(t) = \begin{cases} p_{\max}, & \text{if } 0 \leq t \leq \eta_1^*, \\ p_{\min}, & \text{if } \eta_1^* < t \leq \eta_2^*, \\ p_{\max}, & \text{if } \eta_2^* < t \leq \eta_3^*, \\ p_{\min}, & \text{if } \eta_3^* < t \leq T. \end{cases} \quad (6.7)$$

7. Numerical Study

In order to illustrate our analytical results, we consider some numerical examples for problem (2.11), (2.15) using BOCOP-2.0.5 software package. BOCOP-2.0.5 is an optimal control interface, which is specifically designed for solving optimal control problems with general path and boundary constraints and with free or fixed final time and implemented in MATLAB [67]. Using discretization of time, problems under consideration are approximated by finite-dimensional optimization problems, which are thereby solved by the well-known open-source software package IPOPT designed for large-scale nonlinear optimization.

In our calculations, we use a time grid with 5000 nodes over time interval $[0, T]$ that is equivalent to the time step of $\Delta t = 2T \cdot 10^{-4}$. Since the problem is solved by a direct method using an iterative approach, at each step we impose the acceptable convergence tolerance $\epsilon_{rel} = 10^{-14}$. In the computations we use the sixth-order Lobatto III C discretization rule (see [67] for details). Parameters and initial conditions of system (2.11), the weights in (2.15), control constraints in (2.5) are as following:

$$\begin{aligned} \lambda &= 0.13 & \mu &= 0.013 & \sigma &= 0.4 & k &= 0.4 & \gamma &= 0.05 \\ x_0 &= 5.0 & y_0 &= 0.15 & z_0 &= 20.0 & \alpha &= 1.0 & \beta &= 0.01 \\ u_{\min} &= 0.10 & u_{\max} &= 0.13 & v_{\min} &= 0.8 & v_{\max} &= 1.0 & & \\ w_{\min} &= 0.013 & w_{\max} &= 0.018 & p_{\min} &= 0.1 & p_{\max} &= 0.4 & & \\ q_{\min} &= 0.05 & q_{\max} &= 0.08 & r_{\min} &= 0.4 & r_{\max} &= 0.6 & & \end{aligned} \quad (7.1)$$

These values are adopted from [68]. We note that $\alpha > \beta$.

Results of the numerical calculations for $T = 60.0$ are presented in Figures 1–4. Figures 1 and 2 correspond to the outcomes for the following bilinear incidence rate

$$f(x, z) = \kappa x z. \quad (7.2)$$

with $\kappa = 0.00216$. Figure 1 show the optimal solutions $x_*(t)$, $y_*(t)$ and $z_*(t)$, while Figure 2 display the corresponding optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$. $J_* = 0.270579$ is the minimum value of the functional $J(u, v, w, p, q, r)$ (see (2.15)). IPOPT used 40 iterations in order to solve the optimal control problem (2.11) and (2.15).

Figures 3 and 4 present results for a nonlinear incidence rate $f(x, z)$ of the Michaelis–Menten kinetics type with saturation in both variables:

$$f(x, z) = \frac{\kappa x}{1 + mx} \frac{z}{1 + lz}, \quad (7.3)$$

where $\kappa = 0.00216$, $m = 0.1$, $l = 0.001$. Figure 3 show the optimal solutions $x_*(t)$, $y_*(t)$ and $z_*(t)$, while Figure 4 display the corresponding optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$. $J_* = 0.250366$ is the minimum value of the functional $J(u, v, w, p, q, r)$ of (2.15). IPOPT used 45 iterations in order to solve the optimal control problem (2.11) and (2.15).

Results of the numerical calculations for $T = 120.0$ are presented in Figures 5–8. Figures 5 and 6 correspond to the outcomes for the bilinear incidence rate $f(x, z)$ of (7.2). Figure 5 shows the optimal solutions $x_*(t)$, $y_*(t)$ and $z_*(t)$, while Figure 5 displays the corresponding optimal controls $u_*(t)$, $v_*(t)$, $w_*(t)$, $p_*(t)$, $q_*(t)$ and $r_*(t)$. $J_* = 1.249370 \cdot 10^{-3}$ is the minimum value of the functional $J(u, v, w, p, q, r)$

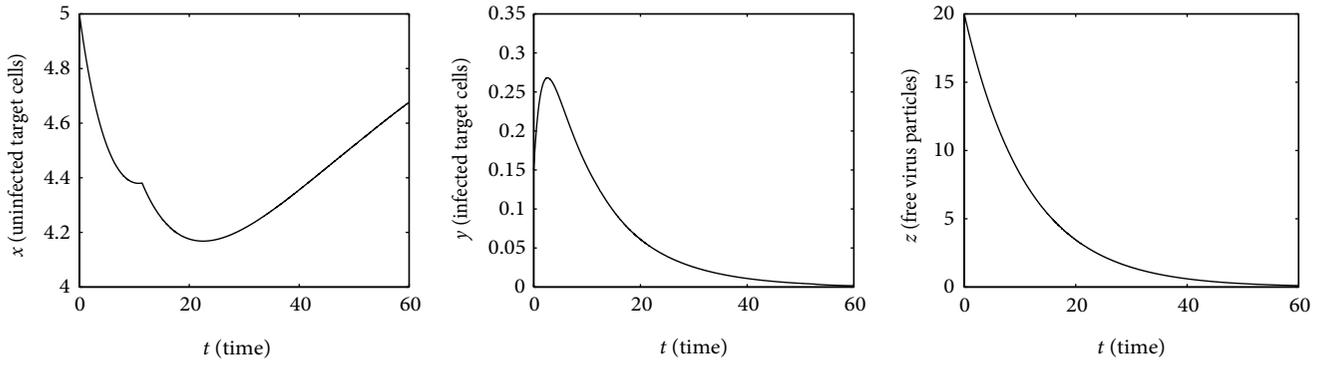


FIGURE 1: Optimal solutions for $T = 60.0$: $x_*(t)$, $y_*(t)$, $z_*(t)$.

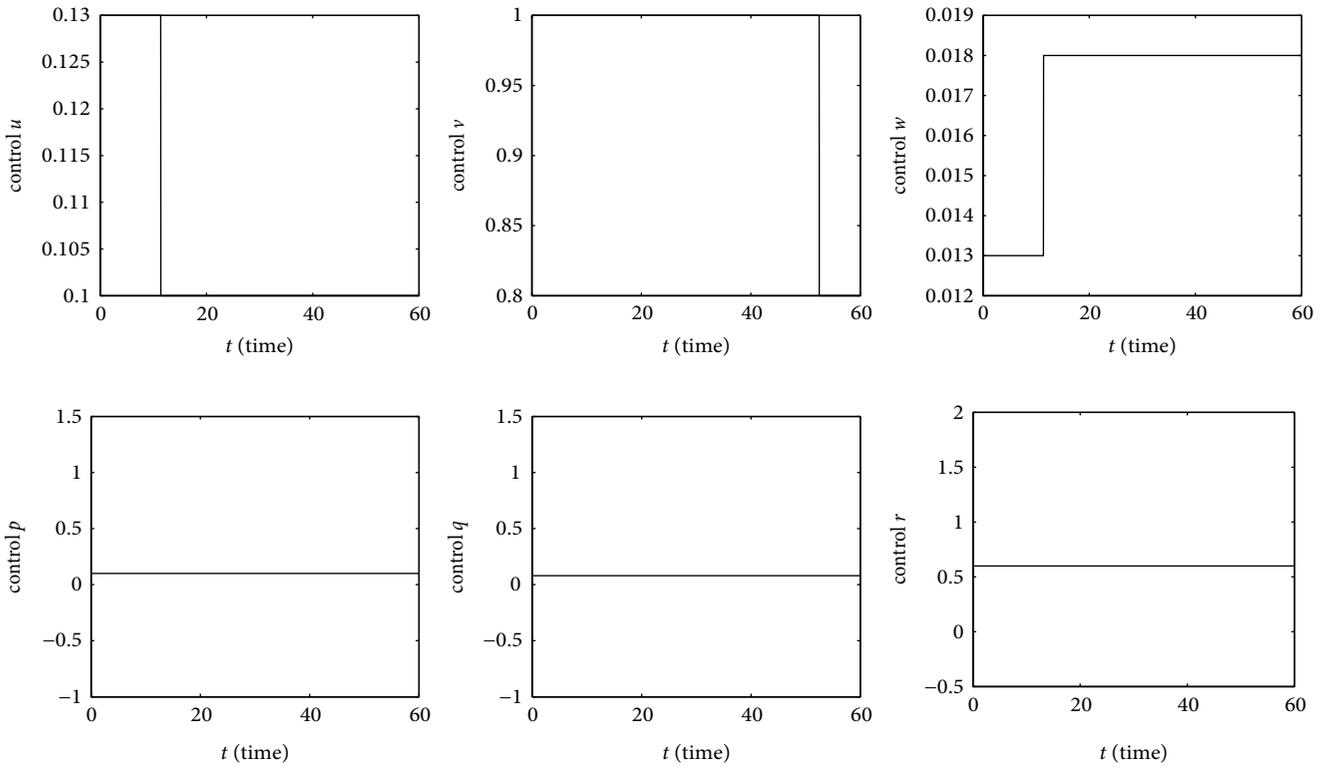


FIGURE 2: Optimal controls for $T = 60.0$: upper row: $u_*(t)$, $v_*(t)$, $w_*(t)$; lower row: $p_*(t)$, $q_*(t)$, $r_*(t)$.

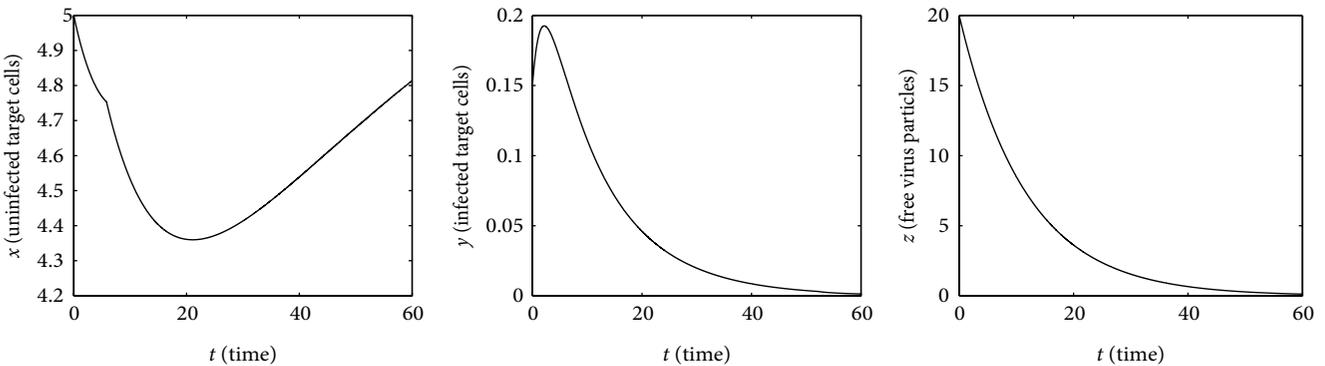


FIGURE 3: Optimal solutions for $T = 60.0$: $x_*(t)$, $y_*(t)$, $z_*(t)$.

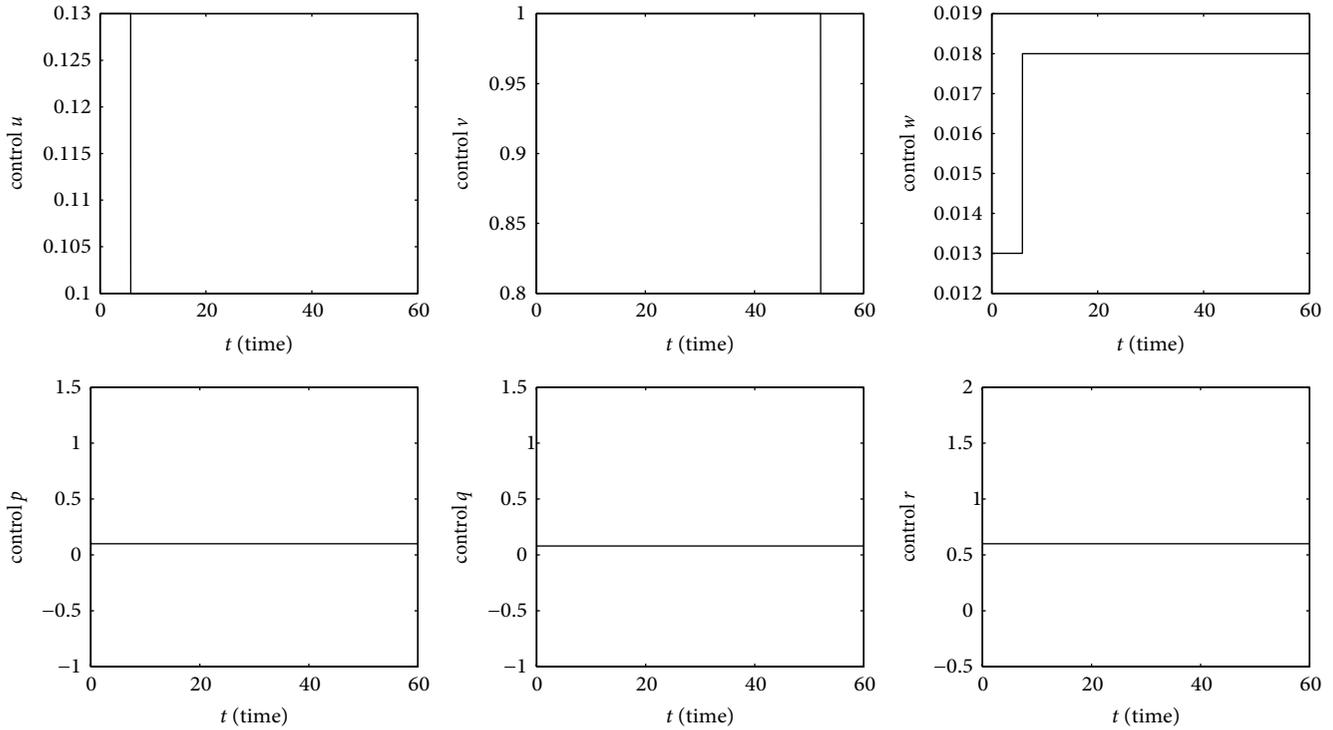


FIGURE 4: Optimal controls for $T = 60.0$: upper row: $u_*(t), v_*(t), w_*(t)$; lower row: $p_*(t), q_*(t), r_*(t)$.

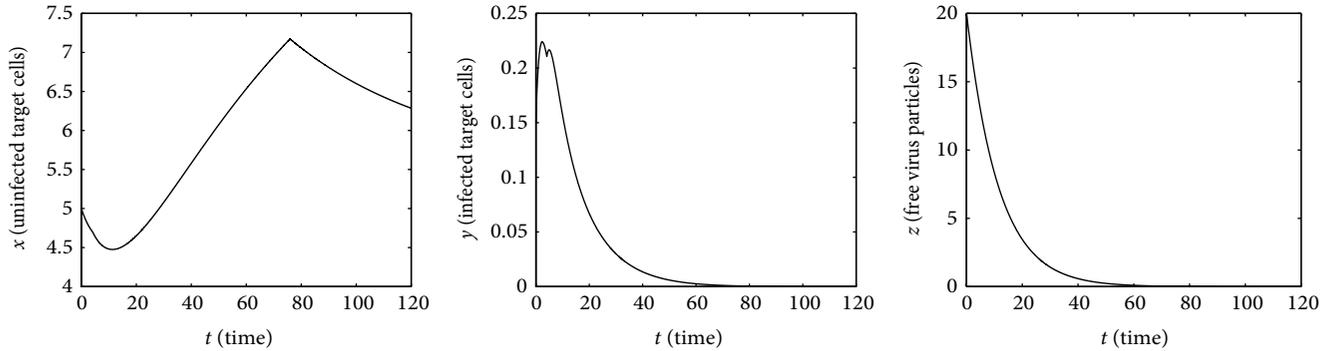


FIGURE 5: Optimal solutions for $T = 120.0$: $x_*(t), y_*(t), z_*(t)$.

(see (2.15)). IPOPT used 50 iterations in order to solve the optimal control problem (2.11), (2.15).

Figures 7 and 8 present results for the nonlinear incidence rate $f(x, z)$ of (7.3). Figure 7 show the optimal solutions $x_*(t), y_*(t)$ and $z_*(t)$, while Figure 8 display the corresponding optimal controls $u_*(t), v_*(t), w_*(t), p_*(t), q_*(t)$ and $r_*(t)$. $J_* = 1.425958 \cdot 10^{-3}$ is the minimum value of the functional $J(u, v, w, p, q, r)$ of (2.15). IPOPT used 55 iterations in order to solve the optimal control problem (2.11) and (2.15).

The numerical calculations lead to the following conclusions:

- (i) A change in the type of the incidence rate $f(x, z)$ does not lead to qualitative changes in the behavior of the optimal solutions $x_*(t), y_*(t)$ and $z_*(t)$.
- (ii) Taking into account the relationship (3.15) between the switching functions $L_u(t), L_w(t)$ and $L_p(t), L_q(t)$, it can be seen that the optimal controls $u_*(t), v_*(t)$ and

$w_*(t)$ are piecewise constant functions with switchings. At the same time, the optimal controls $p_*(t), q_*(t)$ and $r_*(t)$ are the constant functions over the whole interval $[0, T]$. Moreover, all optimal controls behave at the end of the time interval precisely in accordance with formulas (6.2)–(6.4), (6.6) and (6.7).

8. Conclusions

In this paper, we used an optimal control theory to model the highly active antiretroviral therapy for the purpose of finding both an optimal scheduler and an optimal drug combination to be used in HAART. To address this issue, we considered a compartmental model of within-host HIV dynamics based on the Nowak–May model. Compared with the classical Nowak–May model, our model assumes that the incidence rate is given by an unspecified nonlinear function $f(x, z)$ constrained by

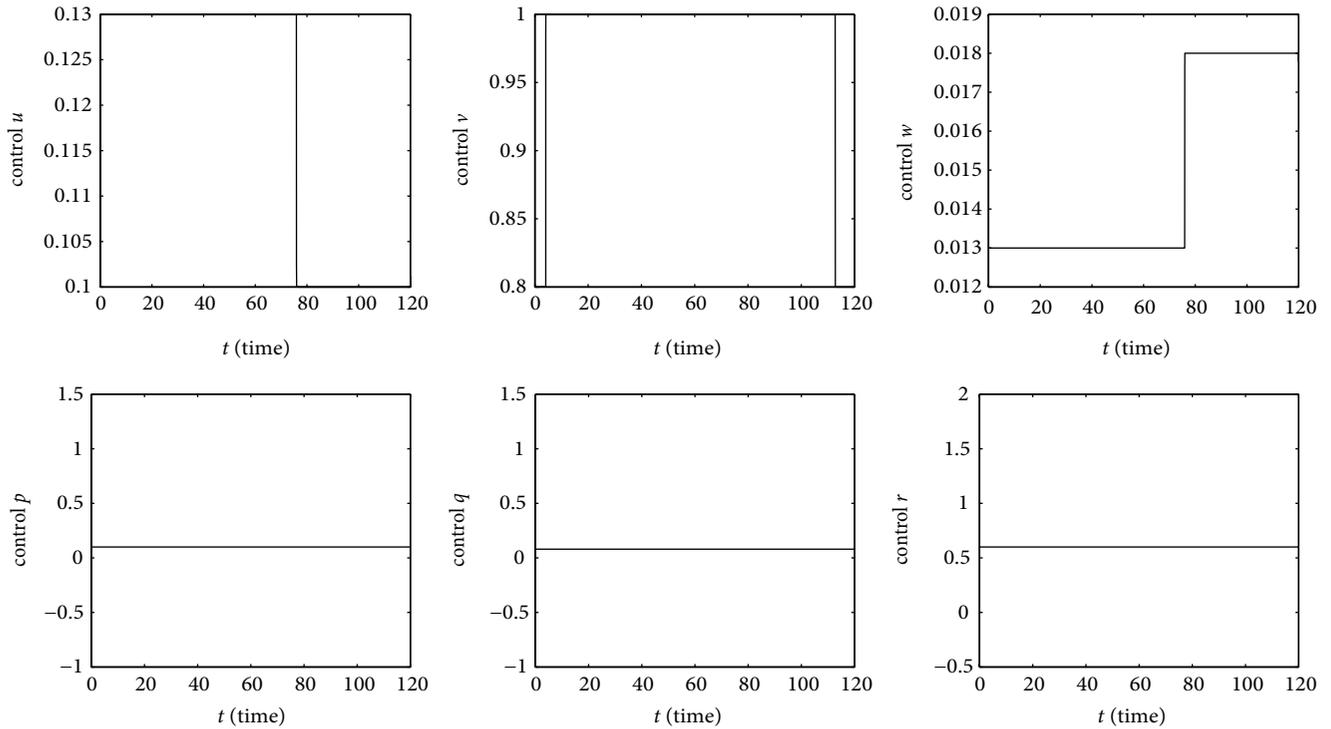


FIGURE 6: Optimal controls for $T = 120.0$: upper row: $u_*(t)$, $v_*(t)$, $w_*(t)$; lower row: $p_*(t)$, $q_*(t)$, $r_*(t)$.

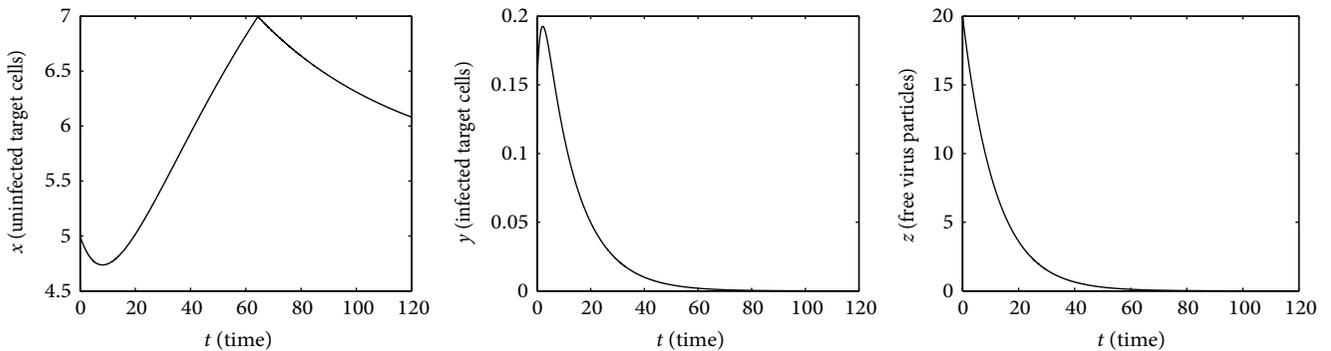


FIGURE 7: Optimal solutions for $T = 120.0$: $x_*(t)$, $y_*(t)$, $z_*(t)$.

a few biologically feasible conditions and takes account of the loss of free virus due to the target cell infections. The constraints imposed on incidence rate $f(x, z)$ summarized in Assumption 1 are biologically motivated and hold for any realistic incidence rate. These constraints are considerably less restrictive than the conditions for the uniqueness of a positive equilibrium state and the global asymptotic stability of the system [45–48]. Apart from biologically motivated Assumption 1 and for convenience of the analysis, we also introduce constraints in Assumption 9. Further, our numerical experiments show that constraints in Assumption 9 are sufficient but not necessary conditions.

To mirror HAART, we introduce six simultaneously acting bounded controls into this model. These six controls comprise all controls possible for this model and represent the actions that the existing anti-HIV drugs exhibit. The optimal control

problem of minimizing the viral load and the infected cell concentration at the end of a predetermined time interval is stated and solved with the objective functional independent on the controls.

Using Pontryagin Maximum Principle, we investigated optimal dynamic analytically. We proved that the optimal controls are bang-bang with no singular arcs and found all possible variants of the optimal controls. The realization of a specific variant depends on specific initial conditions $(x(0), y(0), z(0))$ and the maximal values of derivatives $f'_x(x, z)$ and $f'_z(x, z)$ reached in the feasible region of the model. When the maximal number of switchings of the optimal controls is known, the two-point boundary value problem for Pontryagin Maximum Principle can be immediately reduced to a considerably simpler problem of finite-dimensional optimization. For specific parameter values and initial

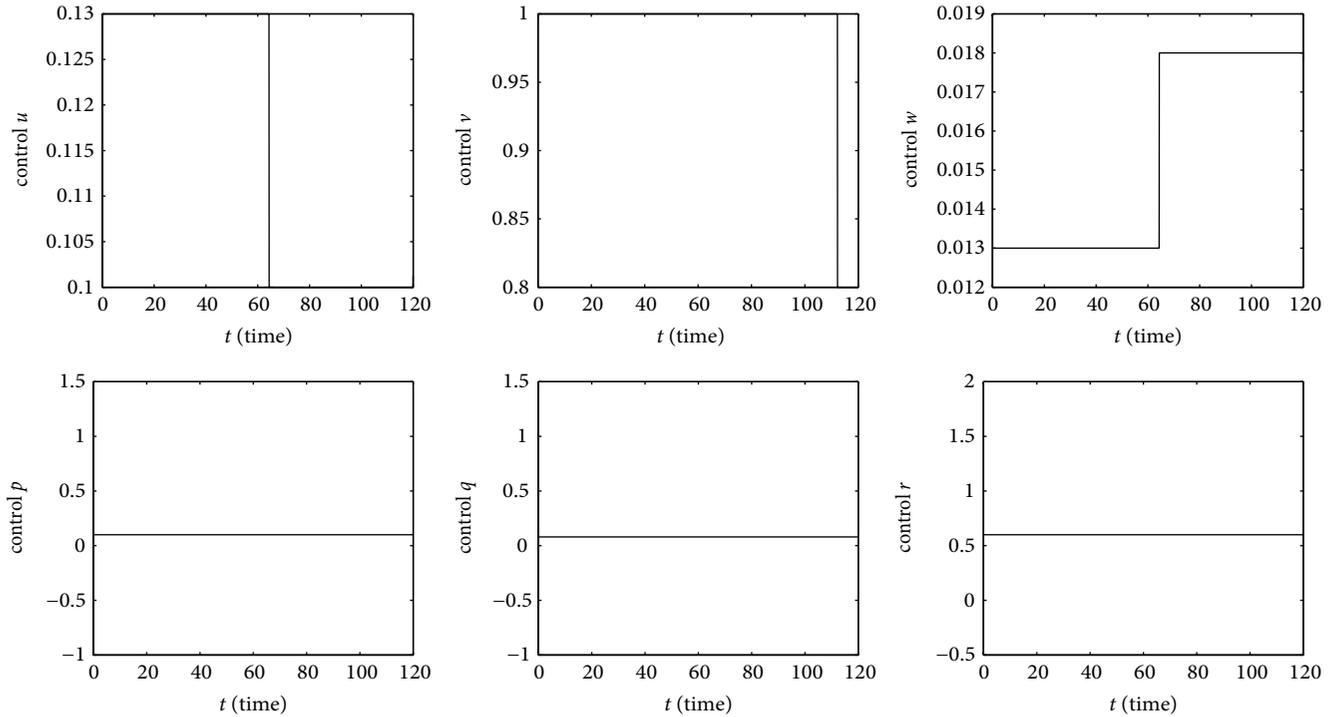


FIGURE 8: Optimal controls for $T = 120.0$: upper row: $u_*(t)$, $v_*(t)$, $w_*(t)$; lower row: $p_*(t)$, $q_*(t)$, $r_*(t)$.

conditions, the latter problem can be solved numerically. However, in this paper, we did not use this approach; we used, instead, an alternative numerical solver package BOCOP-2.0.5 to verify and illustrate our analytical findings.

It follows from our analysis that for the considered model with a nonlinear incidence rate satisfying few biologically feasible assumptions, the principal qualitative properties of the optimal controls, such as the maximum number of switchings of the bang-bang controls and the order of these switchings, do not depend on a particular form of the nonlinearity. Additionally, our analysis indicates that possible 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, however, 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 functionals which are independent of the controls and can be different for the explicitly depending on controls objective functionals.

While our numerical outcomes are presented for treatment duration of 60 days, we proved in this paper that our results do not depend on the length of the treatment and that the optimal schedule and drug combination would be the same for different time intervals.

Another practically relevant conclusion that follows from our analysis is that the six considered controls can be segregated into four groups such that the controls of each group should be considered as a single control and started, carried out and ended simultaneously.

- (i) The first group is composed of controls $\tilde{u}(t)$ (“immunization” of CD4+ T cells before they enter patient’s blood-stream) and $\tilde{w}(t)$ (“immunization” of the existing CD4+ T cells) and corresponds to action of the fusion inhibitors. (Drugs that “immunize” CD4+ T cells before they enter patient’s blood stream do not exist at the moment, but such drugs can appear in the nearest future).
- (ii) The second group combines controls $\tilde{p}(t)$ (inhibition of the virus production by the infected cells) and $\tilde{q}(t)$ (immobilization or removal of the free virus particles). The first of these controls correspond to actions of the protease and maturation inhibitors. Drugs represented by control $\tilde{q}(t)$ are not currently available, but they can appear in the nearest future. Action of such drugs is similar to that of antibodies.
- (iii) The third and the fourth groups contain a single control each; namely, $\tilde{r}(t)$ (killing or immobilizing of the infected cells, or removing them by other means) and $\tilde{v}(t)$ (inhibiting the infection rate). Control $\tilde{v}(t)$ corresponds to action of NRTI, NtRTI, NNRTI and the integrase inhibitors. Drugs represented by control $\tilde{r}(t)$ do not exist, but such drugs can be developed in the next few years.

It is remarkable that in the absence of control $\tilde{v}(t)$ the drugs of the first, second and third groups are completely independent and do not interfere. This implies that in the absence of control $\tilde{v}(t)$, the controls of other three groups can be considered separately (which simplifies the analysis), whereas in the presence of this control all four groups should be considered as a complex. What is more important for real-life

practice, this fact also means that a drug or drugs of the fourth control group should be included in HAART. For a continuing uninterrupted HAART, these results also indicate that life-long use of the same drug combination may be not optimal, and that patient-specific alterations of the combinations can be beneficial. However, in this case other considerations, and in particular a possibility of developing of drug-resistant strains, should be taken into consideration as well.

Finally, we note that the qualitative descriptions of the optimal controls (6.1)–(6.7) assuming that constrains (2.7)–(2.10) of Assumption 2 are held. However, while these constrains significantly simplify the analysis, they are not necessary: by arguments that were used in [29, 34], it is possible to establish the validity of relationships (6.1)–(6.7) for the cases where conditions (2.7)–(2.10) of Assumption 2 are violated.

Appendices

A. Reduction of the Matrix of the Linear System

On time interval $[0, T]$ we consider the following third-order system of linear nonautonomous homogeneous differential equations:

$$\dot{r}(t) = F(t)r(t), \tag{A.1}$$

where

$$F(t) = \begin{pmatrix} f_{11}(t) & f_{12}(t) & 0 \\ f_{21}(t) & f_{22}(t) & f_{23}(t) \\ f_{31}(t) & f_{32}(t) & f_{33}(t) \end{pmatrix}, \quad r(t) = \begin{pmatrix} r_1(t) \\ r_2(t) \\ r_3(t) \end{pmatrix}. \tag{A.2}$$

Here $f_{ij}(t)$ are given Lebesgue measurable bounded functions; $f_{12}(t)$ and $f_{23}(t)$ are sign-definite almost everywhere on the interval $[0, T]$ functions; $r_j(t)$, $j = 1, 2, 3$ are absolutely continuous solutions to system (A.1) satisfying the boundary conditions

$$r_1(T) = r_1^0, \quad r_2(T) = r_2^0, \quad r_3(T) = r_3^0, \tag{A.3}$$

where r_j^0 , $j = 1, 2, 3$ are constants. Usually in this problems $r_1(t)$ is the switching function and the object of investigation, and $r_2(t)$ and $r_3(t)$ are the corresponding auxiliary functions.

Our goal is to reduce the matrix $F(t)$ to the upper-triangular form:

$$F(t) = \begin{pmatrix} f_{11}(t) & f_{12}(t) & 0 \\ f_{21}(t) & f_{22}(t) & f_{23}(t) \\ f_{31}(t) & f_{32}(t) & f_{33}(t) \end{pmatrix} \rightarrow \begin{pmatrix} * & f_{12}(t) & 0 \\ 0 & * & f_{23}(t) \\ 0 & 0 & * \end{pmatrix} = D(t). \tag{A.4}$$

In our earlier papers [28, 34], we conducted such a transformation via a series of successive changes of variables. However, such a transformation can be achieved by a simultaneous nondegenerate transformation of a vector $r(t)$

$$\tilde{r}(t) = \begin{pmatrix} 1 & 0 & 0 \\ h_{21}(t) & 1 & 0 \\ h_{31}(t) & h_{32}(t) & 1 \end{pmatrix} \begin{pmatrix} r_1(t) \\ r_2(t) \\ r_3(t) \end{pmatrix}, \tag{A.5}$$

or, in a matrix form, $\tilde{r}(t) = H(t)r(t)$, where

$$H(t) = \begin{pmatrix} 1 & 0 & 0 \\ h_{21}(t) & 1 & 0 \\ h_{31}(t) & h_{32}(t) & 1 \end{pmatrix}, \quad \tilde{r}(t) = \begin{pmatrix} \tilde{r}_1(t) \\ \tilde{r}_2(t) \\ \tilde{r}_3(t) \end{pmatrix}. \tag{A.6}$$

Here, $h_{21}(t)$, $h_{31}(t)$ and $h_{32}(t)$ are functions should be chosen later. It is easy to see that such a transformation does not change function $r_1(t)$; that is, $\tilde{r}_1(t) = r_1(t)$ holds. The other two functions satisfy

$$\tilde{r}_2(t) = h_{21}(t)r_1(t) + r_2(t), \quad \tilde{r}_3(t) = h_{31}(t)r_1(t) + h_{32}(t)r_2(t) + r_3(t). \tag{A.7}$$

Moreover, the reverse transformation is also easily determined as

$$\begin{aligned} r_1(t) &= \tilde{r}_1(t), \quad r_2(t) = \tilde{r}_2(t) - h_{21}(t)\tilde{r}_1(t), \\ r_3(t) &= \tilde{r}_3(t) - h_{32}(t)\tilde{r}_2(t) - (h_{31}(t) - h_{32}(t)h_{21}(t))\tilde{r}_1(t). \end{aligned} \tag{A.8}$$

The corresponding inverse matrix in equality $r(t) = H^{-1}(t)\tilde{r}(t)$ is

$$H^{-1}(t) = \begin{pmatrix} 1 & 0 & 0 \\ -h_{21}(t) & 1 & 0 \\ -(h_{31}(t) - h_{32}(t)h_{21}(t)) & -h_{32}(t) & 1 \end{pmatrix}. \tag{A.9}$$

We apply transformation (A.5) to matrix $F(t)$. To do this, using variables $\tilde{r}_j(t)$, we rewrite system (A.1) as

$$\dot{\tilde{r}}(t) = (\tilde{H}(t) + H(t)F(t))H^{-1}(t)\tilde{r}(t). \tag{A.10}$$

Here we choose matrix $H(t)$, or that is the same, the functions $h_{21}(t)$, $h_{31}(t)$ and $h_{32}(t)$, to satisfy the equality

$$D(t) = (\tilde{H}(t) + H(t)F(t))H^{-1}(t). \tag{A.11}$$

We re-write equality (A.11) element-wise:

$$\begin{aligned} d_{11}(t) &= f_{11}(t) - f_{12}(t)h_{21}(t), \quad d_{12}(t) = f_{12}(t), \quad d_{13}(t) = 0, \\ d_{21}(t) &= (\dot{h}_{21}(t) + f_{11}(t)h_{21}(t) + f_{21}(t)) \\ &\quad - (f_{12}(t)h_{21}(t) + f_{22}(t))h_{21}(t) \\ &\quad - f_{23}(t)(h_{31}(t) - h_{32}(t)h_{21}(t)) = 0, \\ d_{22}(t) &= f_{12}(t)h_{21}(t) + f_{22}(t) - f_{23}(t)h_{32}(t), \quad d_{23}(t) = f_{23}(t), \end{aligned} \tag{A.12}$$

$$\begin{aligned} d_{31}(t) &= (\dot{h}_{31}(t) + f_{11}(t)h_{31}(t) + f_{21}(t)h_{32}(t) + f_{31}(t)) \\ &\quad - (\dot{h}_{32}(t) + f_{12}(t)h_{31}(t) + f_{22}(t)h_{32}(t) + f_{32}(t))h_{21}(t) \\ &\quad - (f_{23}(t)h_{32}(t) + f_{33}(t))(h_{31}(t) - h_{32}(t)h_{21}(t)) = 0, \end{aligned} \tag{A.13}$$

$$\begin{aligned} d_{32}(t) &= (\dot{h}_{32}(t) + f_{12}(t)h_{31}(t) + f_{22}(t)h_{32}(t) + f_{32}(t)) \\ &\quad - (f_{23}(t)h_{32}(t) + f_{33}(t))h_{32}(t) = 0, \\ d_{33}(t) &= f_{23}(t)h_{32}(t) + f_{33}(t). \end{aligned} \tag{A.14}$$

The matrix $D(t)$ that satisfies these equalities has the required upper-triangular form (A.2). On its main diagonal the matrix has values

$$\begin{aligned} d_{11}(t) &= f_{11}(t) - f_{12}(t)h_{21}(t), \\ d_{22}(t) &= f_{12}(t)h_{21}(t) + f_{22}(t) - f_{23}(t)h_{32}(t), \\ d_{33}(t) &= f_{23}(t)h_{32}(t) + f_{33}(t), \end{aligned} \tag{A.15}$$

and, due to equalities (A.12) and (A.14), functions $h_{21}(t)$ and $h_{32}(t)$ satisfy the following differential equations:

$$\begin{aligned} \dot{h}_{21}(t) &= -f_{21}(t) + (f_{22}(t) - f_{11}(t))h_{21}(t) \\ &\quad + f_{23}(t)h_{31}(t) + f_{12}(t)h_{21}^2(t) - f_{23}(t)h_{32}(t)h_{21}(t), \\ \dot{h}_{32}(t) &= -f_{32}(t) + (f_{33}(t) - f_{22}(t))h_{32}(t) \\ &\quad - f_{12}(t)h_{31}(t) + f_{23}(t)h_{32}^2(t). \end{aligned} \tag{A.16}$$

To find a similar equation for function $h_{31}(t)$, we substitute (A.14) into equality (A.13). This yields the equation

$$\dot{h}_{31}(t) = -f_{31}(t) + (f_{33}(t) - f_{11}(t))h_{31}(t) - f_{21}(t)h_{32}(t) + f_{23}(t)h_{32}(t)h_{31}(t). \tag{A.17}$$

Combining equations (A.16) and (A.17), we finally get a system of nonautonomous nonhomogeneous quadratic differential equations for functions $h_{21}(t)$, $h_{31}(t)$ and $h_{32}(t)$:

$$\begin{cases} \dot{h}_{21}(t) = -f_{21}(t) + (f_{22}(t) - f_{11}(t))h_{21}(t) + f_{23}(t)h_{31}(t) \\ \quad + f_{12}(t)h_{21}^2(t) - f_{23}(t)h_{32}(t)h_{21}(t), \\ \dot{h}_{31}(t) = -f_{31}(t) + (f_{33}(t) - f_{11}(t))h_{31}(t) \\ \quad - f_{21}(t)h_{32}(t) + f_{23}(t)h_{32}(t)h_{31}(t), \\ \dot{h}_{32}(t) = -f_{32}(t) + (f_{33}(t) - f_{22}(t))h_{32}(t) \\ \quad - f_{12}(t)h_{31}(t) + f_{23}(t)h_{32}^2(t). \end{cases} \tag{A.18}$$

The transformed system (A.10) in this case has the following form:

$$\begin{cases} \dot{\tilde{r}}_1(t) = (f_{11}(t) - f_{12}(t)h_{21}(t))\tilde{r}_1(t) + f_{12}(t)\tilde{r}_2(t), \\ \dot{\tilde{r}}_2(t) = (f_{12}(t)h_{21}(t) + f_{22}(t) - f_{23}(t)h_{32}(t))\tilde{r}_2(t) + f_{23}(t)\tilde{r}_3(t), \\ \dot{\tilde{r}}_3(t) = (f_{23}(t)h_{32}(t) + f_{33}(t))\tilde{r}_3(t). \end{cases} \tag{A.19}$$

Initial conditions

$$h_{21}(0) = h_{21}^0, h_{31}(0) = h_{31}^0, h_{32}(0) = h_{32}^0, \tag{A.20}$$

where h_{21}^0, h_{31}^0 and h_{32}^0 are constants, can be added to system (A.18).

As a result, we have the Cauchy problem, and, as a consequence of the Existence and Uniqueness of Solution Theorem [64], solutions $h_{21}(t), h_{31}(t), h_{32}(t)$ to this problem can be defined only locally, in a neighborhood of the value $t = 0$. However, for this problem it is important to have solutions of the Cauchy problem (A.18), (A.20) defined on the entire interval $[0, T]$ (see [28, 34]). That is we have to find the restrictions on functions $f_{ij}(t)$ such that under these restrictions solutions of problem (A.18) and (A.20) were defined on the entire interval $[0, T]$. Then system (A.19) would be also defined on the entire interval and, hence, it would be possible to use the Generalized Rolle's Theorem [53] to estimate the number of zeros of switching function $r_1(t) = \tilde{r}_1(t)$.

The analysis of the 3-dimensional quadratic system (A.18) can be significantly simplified by splitting this system into two simpler 2-dimensional quadratic subsystems. Specifically, it is easy to see that equations for $h_{31}(t)$ and $h_{32}(t)$ contain no variable $h_{21}(t)$ and, hence, can be considered separately. These two equations form the first of these quadratic subsystems:

$$\begin{cases} \dot{h}_{31}(t) = -f_{31}(t) + (f_{33}(t) - f_{11}(t))h_{31}(t) - f_{21}(t)h_{32}(t) \\ \quad + f_{23}(t)h_{32}(t)h_{31}(t), \\ \dot{h}_{32}(t) = -f_{32}(t) + (f_{33}(t) - f_{22}(t))h_{32}(t) - f_{12}(t)h_{31}(t) \\ \quad + f_{23}(t)h_{32}^2(t), \\ h_{31}(0) = h_{31}^0, h_{32}(0) = h_{32}^0. \end{cases} \tag{A.21}$$

To formulate the second 2-dimensional subsystem, we introduce variable

$$h_{20}(t) = h_{32}(t)h_{21}(t) - h_{31}(t). \tag{A.22}$$

Substituting (A.22) into the first equation of system (A.18), we get equation

$$\dot{h}_{21}(t) = -f_{21}(t) + (f_{22}(t) - f_{11}(t))h_{21}(t) - f_{23}(t)h_{20}(t) + f_{12}(t)h_{21}^2(t). \tag{A.23}$$

Besides, function $h_{20}(t)$ satisfies the following differential equation:

$$\dot{h}_{20}(t) = f_{31}(t) + (f_{33}(t) - f_{11}(t))h_{20}(t) - f_{32}(t)h_{21}(t) + f_{12}(t)h_{21}(t)h_{20}(t). \tag{A.24}$$

Initial condition $h_{20}^0 = h_{20}(0)$ for variable $h_{20}(t)$ immediately follows from (A.22):

$$h_{20}^0 = h_{32}^0 h_{21}^0 - h_{31}^0. \tag{A.25}$$

Combining (A.23)–(A.25) and adding the initial condition for function $h_{21}(t)$ from (A.20), we get the second 2-dimensional quadratic subsystem

$$\begin{cases} \dot{h}_{20}(t) = f_{31}(t) + (f_{33}(t) - f_{11}(t))h_{20}(t) - f_{32}(t)h_{21}(t) \\ \quad + f_{12}(t)h_{21}(t)h_{20}(t), \\ \dot{h}_{21}(t) = -f_{21}(t) + (f_{22}(t) - f_{11}(t))h_{21}(t) - f_{23}(t)h_{20}(t) \\ \quad + f_{12}(t)h_{21}^2(t), \\ h_{20}(0) = h_{20}^0 = h_{32}^0 h_{21}^0 - h_{31}^0, h_{21}(0) = h_{21}^0. \end{cases} \tag{A.26}$$

We would like to emphasize that the 3-dimensional quadratic system (A.18) is equivalent to two 2-dimensional quadratic subsystems (A.21) and (A.26). As analysis of 2-dimensional systems is usually simpler, further we will deal with the later. In particular, we have to analyze these quadratic subsystems to find the restrictions on functions $f_{ij}(t)$ which guarantee that there exist solutions of subsystems (A.21) and (A.26) defined on the entire interval $[0, T]$.

B. Verifying Assumption 9

To verify that Assumption 9 holds for the optimal controls that we found in this paper and the corresponding optimal solutions, we consider system (2.11) for the case of bilinear incidence rate $f(x, z) = \kappa xz$, where κ is a positive constant:

$$\begin{cases} \dot{x}(t) = u(t) - \kappa v(t)x(t)z(t) - w(t)x(t), \quad t \in [0, T], \\ \dot{y}(t) = \kappa v(t)x(t)z(t) - r(t)y(t), \\ \dot{z}(t) = p(t)y(t) - \kappa v(t)x(t)z(t) - q(t)z(t), \\ x(0) = x_0, \quad y(0) = y_0, \quad z(0) = z_0; \quad x_0, y_0, z_0 > 0. \end{cases} \tag{B.1}$$

For this case Assumption 9 is

$$\left\{ \begin{aligned} &\kappa v(t)((w(t) - r(t))(z(t) + (p(t) + q(t) - r(t))x(t))) \geq w(t)q(t), \\ &w(t)q(t)(\kappa v(t)(z(t) + x(t)) + r(t)) \\ &\quad + \kappa v(t)(w(t)(w(t) - r(t))z(t) + q(t)(p(t) + q(t) - r(t))x(t)) \\ &\quad \geq w(t)q(t)(w(t) + q(t)), \\ &(w(t) - q(t))^2(\kappa v(t)(z(t) + x(t)) \\ &\quad - 0.5((w(t) - r(t)) + (q(t) - r(t)))) \\ &\quad + 4\kappa v(t)((1 - 2w(t))(w(t) - r(t))z(t) \\ &\quad + (1 - 2q(t))(p(t) + q(t) - r(t))x(t)) \geq 0. \end{aligned} \right. \quad (B.2)$$

Our goal is to find the restrictions on the parameters and the initial conditions of system (2.11), as well as on values u_{\min} , u_{\max} , v_{\min} , v_{\max} , w_{\min} , w_{\max} , p_{\min} , p_{\max} , q_{\min} , q_{\max} , r_{\min} and r_{\max} defined in (2.5), which ensure that inequalities (B.2) are held. Taking into consideration that all the controls are bang-bang and taking either the minimal, or the maximal possible value, we immediately see that the first two inequalities in (B.2) hold if the smallest values of their left-hand sides are greater than or equal to the largest values of the right-hand sides. That is, we have the inequalities:

$$\begin{aligned} &\kappa v_{\min}((w_{\min} - r_{\max})z_{\min} + (p_{\min} + q_{\min} - r_{\max})x_{\min}) \geq w_{\max}q_{\max}, \\ &w_{\min}q_{\min}(\kappa v_{\min}(z_{\min} + x_{\min}) + r_{\min}) + \kappa v_{\min}(w_{\min}(w_{\min} - r_{\max})z_{\min} \\ &\quad + q_{\min}(p_{\min} + q_{\min} - r_{\max})x_{\min}) \geq w_{\max}q_{\max}(w_{\max} + q_{\max}). \end{aligned} \quad (B.3)$$

Here x_{\min} and z_{\min} are the positive lower bounds for solutions $x(t)$ and $z(t)$, respectively. In inequalities (2.12) of Lemma 3 for the general incidence rate $f(x, z)$, these values are zero. However, for the specific function $f(x, z) = \kappa xz$ the lower bounds x_{\min} and z_{\min} can be refined. Specifically, integrating the first equation in (B.1), we get the inequality

$$x(t) \geq x_{\min}, \quad t \in [0, T], \quad (B.4)$$

where

$$x_{\min} = x_0 e^{-(\kappa v_{\max} z_{\max} + w_{\max})T} + \frac{u_{\min}}{\kappa v_{\max} z_{\max} + w_{\max}} (1 - e^{-(\kappa v_{\max} z_{\max} + w_{\max})T}). \quad (B.5)$$

To find the lower bound z_{\min} , we, firstly, have to find the positive lower bound y_{\min} for $y(t)$ by integrating the second equation of system (B.1). This yields

$$y(t) \geq y_{\min}, \quad t \in [0, T], \quad (B.6)$$

where $y_{\min} = y_0 e^{-r_{\max}T}$. Then, using this inequality, we integrate the third equation of system (B.1) and get the inequality

$$z(t) \geq z_{\min}, \quad t \in [0, T], \quad (B.7)$$

where

$$z_{\min} = z_0 e^{-(\kappa v_{\max} x_{\max} + q_{\max})T} + \frac{p_{\min} y_{\min}}{\kappa v_{\max} x_{\max} + q_{\max}} (1 - e^{-(\kappa v_{\max} x_{\max} + q_{\max})T}). \quad (B.8)$$

We would like to note that the values x_{\max} and z_{\max} in inequalities (B.4) and (B.7) can be also found by the integration of the first and third equations of system (B.1) and a subsequent evaluation of the results. These values are found in (2.12), Lemma 3:

$$\begin{aligned} x_{\max} &= \min \left\{ \frac{x_0 + \lambda T; (x_0 + y_0) + \lambda}{\min\{\mu; \sigma\}} \right\}, \\ z_{\max} &= \min \left\{ \frac{z_0 + \kappa y_{\max} T; (y_0 + z_0) + \kappa y_{\max}}{\min\{\sigma; \gamma\}} \right\}. \end{aligned} \quad (B.9)$$

The value y_{\max} is also defined in (2.12).

The third inequality in (B.2) is satisfied if the smallest possible values of its left-hand side is greater than or equal to zero. Hence, we have the following inequalities:

$$\begin{aligned} &\kappa v_{\min}(z_{\min} + x_{\min}) - 0.5((w_{\max} - r_{\min}) + (q_{\max} - r_{\min})) \geq 0, \\ &(1 - 2w_{\max})(w_{\min} - r_{\max})z_{\min} \\ &\quad + (1 - 2q_{\max})(p_{\min} + q_{\min} - r_{\max})x_{\min} \geq 0. \end{aligned} \quad (B.10)$$

Combining inequalities (B.3) and (B.10), we obtain the following system:

$$\left\{ \begin{aligned} &\kappa v_{\min}((w_{\min} - r_{\max})z_{\min} + (p_{\min} + q_{\min} - r_{\max})x_{\min}) \geq w_{\max}q_{\max}, \\ &w_{\min}q_{\min}(\kappa v_{\min}(z_{\min} + x_{\min}) + r_{\min}) \\ &\quad + \kappa v_{\min}(w_{\min}(w_{\min} - r_{\max})z_{\min} + q_{\min}(p_{\min} + q_{\min} - r_{\max})x_{\min}) \\ &\quad \geq w_{\max}q_{\max}(w_{\max} + q_{\max}), \\ &\kappa v_{\min}(z_{\min} + x_{\min}) - 0.5((w_{\max} - r_{\min}) + (q_{\max} - r_{\min})) \geq 0, \\ &(1 - 2w_{\max})(w_{\min} - r_{\max})z_{\min} \\ &\quad + (1 - 2q_{\max})(p_{\min} + q_{\min} - r_{\max})x_{\min} \geq 0. \end{aligned} \right. \quad (B.11)$$

If these inequalities hold, then inequalities (B.2) are satisfied as well.

We verified the validity of inequalities (B.11) by direct calculations for the following parameters and initial conditions of system (2.11) and specific values u_{\min} , v_{\min} , v_{\max} , w_{\min} , w_{\max} , p_{\min} , q_{\min} , q_{\max} , r_{\min} and r_{\max} from (2.5):

$$\begin{aligned} \lambda &= 0.2720 & \kappa &= 1.2500 & k &= 1.0050 \\ \mu &= 0.0050 & \sigma &= 0.0024 & \gamma &= 0.0048 \\ x_0 &= 100.0 & y_0 &= 80.0 & z_0 &= 50 & T &= 50.0 \\ u_{\min} &= 0.2717 & v_{\min} &= 0.9990 & v_{\max} &= 1.0000 \\ w_{\min} &= 0.0050 & w_{\max} &= 0.0060 & p_{\min} &= 1.0039 \\ q_{\min} &= 0.0048 & q_{\max} &= 0.0058 & r_{\min} &= 0.0024 & r_{\max} &= 0.0026. \end{aligned} \quad (B.12)$$

Other specific values of the parameters and initial conditions of system (2.11) can be found in [69, 70].

Data Availability

The software used to support the findings of this study is free open-access software that can be freely downloaded from <https://www.bocop.org/>. The values of parameters and initial conditions used in computations to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

E. Khailov is funded by the Russian Foundation for Basic Research according to the research project 18-51-45003 IND_a.

References

- [1] “United Nations Programme on HIV/AIDS,” www.unaids.org.
- [2] World Health Organization, “HIV/AIDS Data,” <https://www.who.int/hiv/data/en/>.
- [3] A. Maxmen, “Generic HIV drugs will widen US treatment net,” *Nature*, vol. 488, p. 267, 2012.
- [4] C. Sonenklar, *Treatment for HIV and AIDS*, Twenty-First Century Books, Minneapolis, 2010.
- [5] “U.S. Department of Health and Human Services,” <https://aidsinfo.nih.gov/guidelines>.
- [6] Panel on Antiretroviral Guidelines for Adults and Adolescents, “Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents, US Department of Health and Human Services, 2010,” <https://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf>.
- [7] S. Butler, D. Kirschner, and S. Lenhart, “Optimal control of the chemotherapy affecting the infectivity of HIV,” *Advances in Mathematical Population Dynamics – Molecules, Cells and Man*, vol. 6, O. Arino, D. Axelrod, and M. Kimmel, Eds., pp. 557–569, World Scientific, Singapore, 1997.
- [8] D. Kirschner, S. Lenhart, and S. Serbin, “Optimal control of the chemotherapy of HIV,” *Journal of Mathematical Biology*, vol. 35, no. 7, pp. 775–792, 1997.
- [9] B. M. Adams, H. T. Banks, M. Davidian et al., “HIV dynamics: modeling, data analysis, and optimal treatment protocols,” *Journal of Computational and Applied Mathematics*, vol. 184, no. 1, pp. 10–49, 2005.
- [10] B. M. Adams, H. T. Banks, H. D. Kwon, and H. T. Tran, “Dynamic multidrug therapies for HIV: optimal and STI control approaches,” *Mathematical Biosciences and Engineering*, vol. 1, no. 2, pp. 223–241, 2004.
- [11] S. Ahmed, A. Alim, and S. Rahman, “A controlled treatment strategy applied to HIV immunology model,” *Numerical Algebra Control and Optimization*, vol. 8, no. 2, pp. 309–324, 2018.
- [12] S. Anita, V. Arnautu, and V. Capasso, *Introduction to Optimal Control Problems in Life Sciences and Economics*, p. 241, Birkhäuser, USA, 2011.
- [13] R. V. Culshaw, S. Ruan, and R. J. Spiteri, “Optimal HIV treatment by maximising immune response,” *Journal of Mathematical Biology*, vol. 48, no. 5, pp. 545–562, 2004.
- [14] K. R. Fister, S. Lenhart, and J. S. McNally, “Optimizing chemotherapy in an HIV model,” *Electronic Journal of Differential Equations*, vol. 1998, no. 32, pp. 1–12, 1998.
- [15] H. R. Joshi, “Optimal control of an HIV immunology model,” *Optimal Control Applications & Methods*, vol. 23, pp. 199–213, 2002.
- [16] J. J. Kutch and P. Gurfil, “Optimal control of HIV infection with a continuously-mutating viral population,” *Proceedings of American Control Conference, IEEE*, Anchorage, Alaska, pp. 4033–4038, 2002.
- [17] S. Lenhart and J. T. Workman, *Optimal Control Applied to Biological Models*, CRC Press Taylor & Francis Group, London, 2007.
- [18] E. Shudo and Y. Iwasa, “Dynamic optimization of host defence, immune memory, and post-infection pathogen levels in mammals,” *Journal of Theoretical Biology*, vol. 228, pp. 17–29, 2004.
- [19] R. F. Stengel, R. Ghigliazza, N. Rulkarni, and O. Laplace, “Optimal control of innate immune response,” *Optimal Control Applications & Methods*, vol. 23, pp. 91–104, 2002.
- [20] V. V. Velichenko and D. A. Pritykin, “Control of the medical treatment of AIDS,” *Automation and Remote Control*, vol. 67, no. 3, pp. 493–511, 2006.
- [21] D. Wodarz and M. A. Nowak, “Mathematical models of HIV pathogenesis and treatment,” *BioEssays*, vol. 24, no. 12, pp. 1178–1187, 2002.
- [22] H. G. Zadeh, H. C. Nejad, M. M. Abadi, and H. M. Sani, “A new fast optimal control for HIV-infection dynamics based on AVK method and fuzzy estimator,” *American Journal of Scientific Research*, vol. 32, pp. 11–16, 2011.
- [23] 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.
- [24] H. Schättler and U. Ledzewicz, *Geometric Optimal Control: Theory, Methods and Examples*, Springer, New York-Heidelberg-Dordrecht-London, 2012.
- [25] E. V. Grigorieva, N. V. Bondarenko, E. N. Khailov, and A. Korobeinikov, “Three-dimensional nonlinear control model of wastewater biotreatment,” *Neural, Parallel, and Scientific Computations*, vol. 20, pp. 23–36, 2012, <https://www.dynamicpublishers.com/Neural/npscontent.htm>.
- [26] E. V. Grigorieva, N. V. Bondarenko, E. N. Khailov, and A. Korobeinikov, “Finite-dimensional methods for optimal control of autothermal thermophilic aerobic digestion,” *Industrial Waste, InTech*, CroatiaK. -Y. Show and X. Guo, Ed., pp. 91–120, 2012.
- [27] 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*, vol. 20, no. 2, pp. 103–118, 2013. ISSN:14092433.
- [28] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “Parametrization of the attainable set for a nonlinear control model of a biochemical process,” *Mathematical Biosciences and Engineering*, vol. 10, no. 4, pp. 1067–1094, 2013.
- [29] E. V. Grigorieva and E. N. Khailov, “Optimal preventive strategies for SEIR type model of the 2014 Ebola epidemics dynamics of continuous discrete and impulsive,” *Systems Series B*, vol. 24, pp. 155–182, 2017.
- [30] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “Optimal control for a SIR infectious disease model,” *Journal of Coupled Systems and Multiscale Dynamics*, vol. 1, no. 3, pp. 324–331, 2013.
- [31] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “An optimal control problem in HIV treatment,” *Discrete and Continuous Dynamical Systems*, pp. 311–322, 2013.
- [32] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “Optimal control for an epidemic in a population of varying size,” *Discrete and Continuous Dynamical Systems*, pp. 549–561, 2015.
- [33] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “Optimal control for a SIR epidemic model with nonlinear incidence rate,” *Mathematical Modelling of Natural Phenomena*, vol. 11, no. 4, pp. 90–105, 2016.
- [34] E. V. Grigorieva, E. N. Khailov, and A. Korobeinikov, “Optimal control for an SEIR epidemic model with nonlinear incidence rate,” *Studies in Applied Mathematics*, vol. 141, no. 3, pp. 353–398, 2018.

- [35] E. V. Grigorieva, N. V. Bondarenko, E. N. Khailov, and A. Korobeinikov, "Modeling and optimal control for antiretroviral therapy," *Journal of Biological Systems*, vol. 22, no. 2, pp. 199–217, 2014.
- [36] D. Wodarz, J. Christensen, and A. Thomsen, "The importance of lytic and nonlytic immune responses in viral infections," *Trends in Immunology*, vol. 23, no. 4, pp. 194–210, 2002.
- [37] E. F. Arruda, C. M. Dias, C. V. de Magalhães, D. H. Pastore, R. C. A. Thomé, and H. M. Yang, "An optimal control approach to HIV immunology," *Applied Mathematics*, vol. 6, no. 6, pp. 1115–1130, 2015.
- [38] M. Roshanfekr, M. H. Farahi, and R. Rahbarian, "A different approach of optimal control on an HIV immunology model," *Ain Shams Engineering Journal*, vol. 5, no. 1, pp. 213–219, 2014.
- [39] Y. Zhou, Y. Liang, and J. Wu, "An optimal strategy for HIV multitherapy," *Journal of Computational and Applied Mathematics*, vol. 263, pp. 326–337, 2014.
- [40] Y. Zhou, K. Yang, K. Zhou, and C. Wang, "Optimal treatment strategies for HIV with antibody response," *Journal of Applied Mathematics*, vol. 2014, pp. 1–13, 2014.
- [41] G. Akudibillah, A. Pandey, and J. Medlock, "Optimal control for HIV treatment," *Mathematical Biosciences and Engineering*, vol. 16, no. 1, pp. 373–396, 2018.
- [42] K. O. Okosun, O. D. Makinde, and I. Takaidza, "Impact of optimal control on the treatment of HIV/AIDS and screening of unaware infectives," *Applied Mathematical Modelling*, vol. 37, pp. 3802–3820, 2013.
- [43] A. Kolmogorov, "On the Volterra theory of the struggle for existence," *The Italian Institute of Actuaries*, vol. 7, pp. 74–80, 1936.
- [44] A. Korobeinikov, and P. K. Maini, "A Lyapunov function and global properties for SIR and SEIR epidemiological models with nonlinear incidence," *Mathematical Biosciences and Engineering*, vol. 1, no. 1, pp. 57–60, 2004.
- [45] A. Korobeinikov and P. K. Maini, "Nonlinear incidence and stability of infectious disease models," *Mathematical Medicine and Biology*, vol. 22, no. 2, pp. 113–128, 2005.
- [46] A. Korobeinikov, "Global properties of infectious disease models with nonlinear incidence," *Bulletin of Mathematical Biology*, vol. 69, no. 6, pp. 1871–1886, 2007.
- [47] A. Korobeinikov, "Global asymptotic properties of virus dynamics models with dose dependent parasite reproduction and virulence, and nonlinear incidence rate," *Mathematical Medicine and Biology*, vol. 26, no. 3, pp. 225–239, 2009.
- [48] A. Korobeinikov, "Stability of ecosystem: global properties of a general prey-predator model," *Mathematical Medicine and Biology*, vol. 26, no. 4, pp. 309–321, 2009.
- [49] H. Behncke, "Optimal control of deterministic epidemics," *Optimal Control Applications and Methods*, vol. 21, no. 6, pp. 269–285, 2000.
- [50] S. Bonhoeffer, R. M. May, G. M. Shaw, and M. A. Novak, "Virus dynamics and drug therapy," *Proceedings of the National Academy of Sciences USA*, vol. 94, no. 13, pp. 6971–6976, 1997.
- [51] M. A. Novak and R. M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University Press, NY, 2000.
- [52] H. L. Smith and P. De Leenheer, "Virus dynamics: a global analysis," *SIAM Journal of Applied Mathematics*, vol. 63, pp. 1313–1327, 2003.
- [53] 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*, vol. 30, no. 5, pp. 1087–1091, 1992.
- [54] M. A. Novak and R. M. May, "Mathematical biology of HIV infection: antigenic variation and diversity threshold," *Mathematical Biosciences*, vol. 106, pp. 1–2, 1991.
- [55] A. Korobeinikov, "Global properties of basic virus dynamical models," *Bulletin of Mathematical Biology*, vol. 66, no. 4, pp. 879–883, 2004.
- [56] A. Korobeinikov, "Lyapunov functions and global properties for SEIR and SEIS epidemic models," *Mathematical Medicine and Biology*, vol. 21, no. 2, pp. 75–83, 2004.
- [57] H. Gaff and E. Schaefer, "Optimal control applied to vaccination and treatment strategies for various epidemiological models," *Mathematical Biosciences and Engineering*, vol. 6, no. 3, pp. 469–492, 2009.
- [58] U. Ledzewicz and H. Schättler, "On Optimal singular controls for a general SIR-model with vaccination and treatment," *Discrete and Continuous Dynamical Systems*, pp. 981–990, 2011.
- [59] M. S. Nikol'skii, "On an optimal control problem associated with the general "predator-prey" model of A.D. Bazykin," *Differential Equations*, vol. 41, pp. 1660–1667, 2011.
- [60] E. B. Lee and L. Marcus, *Foundations of Optimal Control Theory*, John Wiley & Sons, NY, 1967.
- [61] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *Mathematical Theory of Optimal Processes*, John Wiley & Sons, NY, 1962.
- [62] V. A. Ilyin and E. G. Poznyak, *Foundations of Mathematical Analysis Part 1*, Mir Moscow, Russia, 1982.
- [63] A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis*, Dover Publications NY, 1999.
- [64] P. Hartman, *Ordinary Differential Equations*, John Wiley & Sons, NY, 1964.
- [65] M. A. Krasnosel'skii, *The Operator of Translation along the Trajectories of Differential Equations*, AMS, Providence, RI, 1968.
- [66] B. P. Demidovich, *Lectures on Stability Theory*, Nauka, Moscow, Russia, 1967.
- [67] F. Bonnans, P. Martinon, D. Giorgi et al., "BOCOP 2.0.5 – User Guide 8 February 2017," <https://bocop.org>.
- [68] G. Huang, Y. Takeuchi, and A. Korobeinikov, "HIV evolution and progression of the infection to AIDS," *Journal of Theoretical Biology*, vol. 307, pp. 149–159, 2012.
- [69] N. Dalal, D. Greenhalgh, and X. Mao, "Mathematical modelling of internal HIV dynamics," *Discrete and Continuous Dynamical Systems – Series B*, vol. 12, no. 2, pp. 305–321, 2009.
- [70] H. C. Tuckwell and F. Y. M. Wan, "Nature of equilibria and effects of drug treatments in some simple viral population dynamical models," *IMA Journal of Mathematics Applied in Medicine and Biology*, vol. 17, pp. 311–327, 2000.