A SIR epidemic model structured by immunological variables

OSCAR ANGULO
Departamento de Matematica Aplicada, Universidad de Valladolid,
Pso. Belén, 15, 47011 Valladolid, Spain
oscar@mat.uva.es

FABIO MILNER*
School of Mathematical and Statistical Sciences, Arizona State University,
P.O. Box 871804 Tempe, AZ 85287, United States
milner@asu.edu

LAURENTIU SEGA
Department of Mathematics, Georgia Regents University,
1120 15th Street, Augusta, GA 30912, United States
lsega@gru.edu

Received (Day Month Year)
Accepted (Day Month Year)

Standard mathematical models for analyzing the spread of a disease are usually either epidemiological or immunological. The former are mostly ODE-based models that use classes like susceptibles, recovered, infectives, latently infected, and others to describe the evolution of an epidemic in a population. Some of them also use structure variables, such as size or age. The latter describe the evolution of the immune system/pathogen in the infected host—evolution that usually results in death, recovery or chronic infection. There is valuable insight to be gained from combining these two types of models, as that may lead to a better understanding of the severity of an epidemic. In this article we propose a new type of model that combines the two by using variables of immunological nature as structure variables for epidemiological models. We prove the well-posedness of the proposed model under some restrictions and conclude with a look at a practical application of the model.

Keywords: epidemic model; immunological model; PDE.

1. Introduction

1.1. Motivation

Why an immuno-epidemiological model? In almost all epidemiological models, the death rates and the recovery rates are either constant or depending on chronological age or age of infection. We do know, however, that what ultimately decides whether

*Corresponding author
an individual recovers from an infection (with or without immunity) or dies because of it is how the immune system of that individual protects her/him from the disease. With HIV, for example, the depletion of CD4+ T cells, not the age, is the best indicator of an individual’s state of health and prognosis. So a model that takes into account the immune status of all individuals in a population seems reasonable and desirable.

A similar approach was used in Ref. 1. The authors presented a framework for modeling host-parasite coevolution using nested modeling, and illustrated it by analysing a simple host-parasite system. Refs. 2 and 3 considered models similar to the one presented in this paper, although their approach was geared more towards understanding the evolution of virulence and natural selection. Kostova presented a model of viral spread in a population that embeds an immune response model in an epidemic network model.

We will present in the next two sub-sections some background on epidemiological models and immunological models, respectively, including a specific model of the immune reaction of the human body against HIV infection. We shall later use that immune reaction model in Section 4 to provide a real-life example of the application of our general immuno-epidemiological model presented in Section 2, where variables of immunological nature are used as structure variables for epidemiological models. In Section 3, we will describe an immunologically structured SIR-type model that describes the evolution of an HIV epidemic and prove its well-posedness. In Section 4 we describe a numerical method we use to approximate the solution of a realization of our model corresponding to HIV in San Francisco. These simulations are interesting because they provide some insight into the evolution of the immune systems of HIV-positive individuals for the population, without really knowing the state of any specific individual. Finally, in Section 5 we summarize our findings and describe future directions of this research.

1.2. Epidemiological models

Most epidemiological models are compartmental models as the population is divided into epidemiologically relevant compartments such as susceptible, infected, recovered, latently infected, exposed, quarantined, etc. Assumptions about the rate of transfer from one compartment to another are made, and the number of compartments used usually depends on the characteristics of the specific disease studied.

Some types of epidemiological models that are frequently encountered include SIR, SIS, SEIR, SEIS, SIQR or SEIQR. SIR (susceptible - infected - recovered) models are used to analyze diseases that impart immunity against reinfection (such as influenza, measles, rubella). SIS (susceptible - infected - susceptible) models are used for diseases that confer no immunity against reinfection (such as tuberculosis or gonorrhea). SEIR and SEIS models take into account the existence of an "exposed" period between becoming infected and being infective. SIQR models consider a class called "quarantined", and SEIQR extend those models by consid-
ering a class called "exposed" (asymptomatic individuals who have been in contact with the disease agent). Most of the models described above are formulated as systems of differential equations. Other models, more complicated in nature, may also include partial differential equations or integral equations. These in general are structured models where one or more rates of transfer from one compartment to another depend on size, chronological age or age of infection.\textsuperscript{12–14

1.3. Immunological models

Immunological, or in-host models, that are concerned with describing the reaction of immune system to infection and the ways the host defends itself (or may also be helped via treatment) against viral or bacterial infections or parasites. They were first used to study viral diseases,\textsuperscript{15} and later for helminthic infections\textsuperscript{16,17} or malaria,\textsuperscript{18} and other diseases as well. A significant number of papers are dedicated to the study of HIV/AIDS.\textsuperscript{19,20}

We present now a modification of the Perelson and Nelson model,\textsuperscript{19} with the addition of the $T^*$ term in the first equation. The variables from this model will be used later in this paper as structure variables for our immuno-epidemiological model. We will also establish here some basic properties of the solution of this model of immune response that we will need later in the proof of the main theorem of our article.

Let $T$, $T^*$ and $V$ be, respectively, the densities of CD4+ T cells, infected CD4+ T cells and virus in the plasma of an individual. Then, their interaction can be modeled by the following system:

\begin{align}
\frac{dT}{dt} &= s + pT \left( 1 - \frac{T + T^*}{T_{\text{max}}} \right) - d_T T - kVT, \quad (1.1) \\
\frac{dT^*}{dt} &= kVT - \delta T^*, \quad (1.2) \\
\frac{dV}{dt} &= b\delta T^* - cV. \quad (1.3)
\end{align}

Here $s$ is the rate at which the CD4+ T cells are created from sources within the body (naive stem cell differentiation); the term $pT(1 - (T + T^*)/T_{\text{max}})$ gives the rate at which new CD4+ T cells appear from existing T-cells by cell division ($p$ is the maximum proliferation rate and $T_{\text{max}}$ is the T cell population density at which the proliferation shuts off); $d_T$ is the natural T cell death rate; $k$ is the per virion infection rate, $\delta > d_T$ is the death rate of infected T cells, $b$ is the number of virions an infected T cell produces during its lifetime (referred to as burst rate), and $c$ is the viral clearance rate. In Ref. 20, it is mentioned that the condition $d_T T_{\text{max}} > s$ should be imposed (this is equivalent to saying that if the population ever reaches $T_{\text{max}}$ then it should decrease). In the same paper, a list of realistic values for the parameters $p$, $d_T$, $s$, $T_{\text{max}}$, $k$, $\delta$, $b$ and $c$ is given.
We will next establish some basic properties of $T, T^*$ and $V$, namely positivity and boundedness.

**Proposition 1.1** Consider the system (1.1)-(1.3), together with the initial conditions: $T(0) = T_0, T^*(0) = T_0^*, V(0) = V_0$. If $T_0, T_0^*$ and $V_0$ are positive, then $T, T^*$ and $V$ are positive and bounded for all $t \geq 0$.

## 2. A general model

We first present the general form of our model resembling so-called size-structured models, and then present a model that is specific to HIV. We will allow any number of structure variables that will be represented by a vector, and assume the total population is undergoing logistic growth through linearly increasing mortality but no reduction in fertility.

Consider a vector $\mathbf{Y}$ that has as components the immunological (or other type) variables that will be used as structure variables for the epidemiological model, and consider the infected individuals in the population as being structured by these variables:

$$i = i(t, \mathbf{Y})$$

will represent the density of individuals who, at time $t$, are at state $\mathbf{Y}$. The total number of infected individuals is then given by:

$$I(t) = \int i(t, \mathbf{Y}) d\mathbf{Y},$$

where the integral is computed over the whole domain of possible values for the structure variables.

The new immunologically structured SIR model is now given by:

$$\frac{\partial i}{\partial t} = -\nabla_{\mathbf{Y}}(i) \cdot \frac{d\mathbf{Y}}{dt} - \gamma(\mathbf{Y})i - \mu(\mathbf{Y})i \tag{2.1}$$

$$\frac{dN}{dt} = \beta N - (\bar{\mu} + aN)N \tag{2.2}$$

$$\frac{dS}{dt} = \beta N - (\bar{\mu} + aN)S - \rho \frac{SI}{N} \tag{2.3}$$

$$N = S + I + R \tag{2.4}$$

Here, $S$, $I$ and $R$ represent, respectively, the numbers of susceptible, infective, and recovered individuals, $N$ is the total population, $\gamma$ is the recovery rate, $\mu$ is the natural mortality rate, $\beta$ is the birth rate, $a$ is a (very small) coefficient representing the per capita increase in mortality over baseline that gives the logistic behavior, and $\rho$ is the per capita infection rate.

Of course, the PDE (2.1) has to be complemented by appropriate boundary conditions, and this is not a trivial issue. In particular, for the case motivated in the previous section, $\mathbf{Y} = (V, T, T^*)$, it is obvious that individuals who have not
been infected should have initial values $V_0 = T_0^* = 0$, but such conditions on the structure variables at any time force $V(t) = T^*(t) = 0$ for all future times $t$ and, therefore, new infections may not be reflected as Dirichlet boundary conditions on $V$. Thus we are led to the following boundary conditions for $i$:

$$i(0, T, T^*) = 0, \quad T \geq 0, \quad T^* > 0, \quad (2.5)$$

$$i(V, 0, T^*) = 0, \quad V, T^* \geq 0, \quad (2.6)$$

$$i((V, T, 0) = 0, \quad V > 0, \quad T \geq 0, \quad (2.7)$$

$$\frac{\partial i}{\partial V}(0, T, 0) = \frac{1}{T_{max}} \rho \frac{SI}{N}, \quad T > 0. \quad (2.8)$$

The first three of these conditions reflect, respectively, the fact that no individual can have no virus if (s)he has infected CD4+ T cells, that nobody may have no healthy T cells, and that nobody can have no infected T cells in the presence of virus. The last condition gives the flow of newly infected individuals—who prior to infection had $V = T^* = 0$—into the correct domain for $Y = (V, T, T^*)$, where $T$ can have any allowable value (i.e. whatever it was prior to infection), and $V = T^* = 0$ with $V$ increasing at the rate that accounts for all new infections from (2.3) when integrated in $T$.

The analysis of the general model (2.1)-(2.8) is very complex and remains an open problem. We shall consider a double simplification of this problem by assuming there is a single structure variable, and make an assumption to allow the use of Dirichlet boundary condition. This will actually make the analysis very similar to that of classical size-structured models.

### 3. A simplified model

We will assume now that the infected class is structured by a single immunological variable, namely $V$, the viral load. In other words, we now have

$$i = i(t, V),$$

and the PDE (2.1) now has the simpler form

$$\frac{\partial i}{\partial t} + \frac{\partial i}{\partial V} \frac{dV}{dt} = -\gamma(V)i - \mu(V)i. \quad (3.1)$$

We impose now the boundary condition

$$i(t, \epsilon) = \rho \frac{S(t)I(t)}{N(t)}, \quad (3.2)$$

and the initial condition

$$i(0, V) = \phi(V). \quad (3.3)$$

Why do we use $\epsilon$ in the boundary condition? This provides a way to circumvent the fact that, if no virus is initially present, the infection will not "take off". So,
we will assume that every individual in the population has some minimum, albeit extremely small, positive density of virus, $\epsilon$.

We now impose also the compatibility condition

$$\frac{S(0)I(0)}{N(0)} = \phi(\epsilon),$$

so that the solution of (3.1)-(3.3) will be continuous.

A standard way to solve a partial differential equation of the type (3.1) is by using the method of characteristics. The method consists of reducing the partial differential equation to a family of ordinary differential equations that can be integrated along the characteristics.

Let us start by noting that in this case the characteristic curves in the $tV$-plane starting at the points on the boundary of the domain

$$\Omega = \{(t,V) \mid t \geq 0, V \geq \epsilon\}$$

satisfy the equation

$$\frac{dV}{dt} = b\delta T^* - cV,$$

with initial condition $V(t_0) = V_0$, where $(t_0, V_0) \in \Omega$.

Let us separate the boundary of $\Omega$ as a disjoint union of a "$t$-boundary" and a "$V$-boundary":

$$\Gamma = \partial \Omega = \Gamma_1 \cup \Gamma_2,$$

where

$$\Gamma_1 = \{V = \epsilon, t \geq 0\}, \quad \Gamma_2 = \{V > \epsilon, t = 0\}.$$

Let us represent the solution of the initial value problem

$$\frac{dV}{dt} = b\delta T^* - cV, \quad V(t_0) = V_0$$

using the solution operator $G$ as follows:

$$V = G(t; V_0, t_0).$$

Assuming that $V'(0) > 0$, or, in other words, that $b\delta T^* > c\epsilon$ (i.e. that $V$ starts off increasing, which is a realistic biological assumption), we then have that there exists $\eta > 0$ such that $G(t; V_0, t_0)$ is an increasing function of $t$ on the interval $[0, \eta]$. That means $G$ has an inverse, which we will denote by $H$:

$$H(V; V_0, t_0) = G^{-1}(G(t; V_0, t_0)) = G^{-1}(V).$$

Hence, for every characteristic curve starting at $(t_0, V_0) \in \Gamma$, we have two equivalent representations:

$$C(t_0, V_0) = \{V = G(t; V_0, t_0), t \in [0, \eta]\} = \{t = H(V; V_0, t_0), V \in [\epsilon, \bar{V}]\},$$
for $\epsilon < \bar{V} = G(\eta; V_0, t_0) \leq V_{\text{max}}$. The characteristic $C(0, \epsilon)$ is important in that it separates the set of characteristics originating on $\Gamma_1$ from those originating on $\Gamma_2$.

Let us denote by $\mathcal{D}$ the rectangle in which the solution of the PDE (3.1) will be found,

$$\mathcal{D} = \{(t, V); t \in [0, \eta], V \in [\epsilon, \bar{V}]\}.$$

**Proposition 3.1** To each pair $(t, V) \in \mathcal{D}$, there corresponds a unique pair $(t_0, V_0) \in \Gamma$ such that if $V \leq G(t; \epsilon, 0)$, then $(t_0, V_0) \in \Gamma_1$ and if $V > G(t; \epsilon, 0)$, then $(t_0, V_0) \in \Gamma_2$.

**Proposition 3.2** Using the notation $\alpha = \mu + \gamma$, the solution of the partial differential equation (3.1) with the boundary conditions (3.2)-(3.3) is given (implicitly) by the formula:

$$i(t, V) = \begin{cases} \rho S(t_0)I(t_0) e^{-\int_{t_0}^{t} \alpha(G(\sigma, \epsilon, t_0)) d\sigma}, & \epsilon \leq V \leq G(t; \epsilon, 0), \\ \phi(G(0; V, t)) e^{-\int_{t_0}^{t} \alpha(G(\sigma; V, t)) d\sigma}, & V \geq G(t; \epsilon, 0). \end{cases}$$

Looking carefully at Proposition 3.2, we see that the solution of our PDE is given in an implicit form, since $i(t, V)$ appears both on the left hand side and on the right hand side of the defining relation (implicitly in $I(t_0)$). To prove the (local) existence and uniqueness of the solution we will use the contraction mapping theorem, which states that if $(X, d)$ is a non-empty complete metric space and $\Phi : X \to X$ is a contraction, then $\Phi$ has a unique fixed point.

We let $X$ be the space of continuous functions from $\mathcal{D}$ to $\mathcal{D}$,

$$X = C(\mathcal{D}),$$

and note that $X$ is a non-empty complete metric space. Indeed, the set $[0, \eta] \times [\epsilon, \bar{V}]$ is compact, and the space of continuous functions on a compact set is a Banach space with the $sup$ norm

$$\|f\|_{\infty} = sup_{x \in X} |f(x)|$$

We now define an operator $\Phi : X \to X$ via the equation $\Phi(i) = \bar{i}$, where $\bar{i}$ is the right hand side of the formula in Proposition (3.6) with the input $i$. Let us first note that $\Phi$ is well defined since the compatibility condition (3.4) ensures the continuity of the expression on the right hand side across the characteristic $G(t; \epsilon, 0)$. Then, observe that the local existence and uniqueness of solutions for our PDE is equivalent to the following:

**Theorem 3.1** For $\rho$ and $\eta$ sufficiently small $\Phi$ is a contraction; i.e. there exists a nonnegative constant $q < 1$ such that

$$\|\Phi(i_1) - \Phi(i_2)\|_{\infty} \leq q \|i_1 - i_2\|_{\infty} \text{ for every } i_1, i_2 \in X.$$
4. Numerical results and simulations

Models such as the one presented in this article cannot be solved analytically without some strong restrictive assumptions. Therefore, the use of numerical methods is the most suitable mathematical tool for studying them and it is often the only one available. For the purpose of presenting a scenario as close to real-life as possible, we perform our simulations based on the full model (1.2)-(2.8). However, the 4-dimensional domain of the structure variables $T, V, T^*, t$, creates a serious computer memory problem to run our numerical method. Therefore, we do not take into account the variable $T$ in the simulations by using the reduction suggested in Ref. 19. More specifically, since the $T$ cell levels generally change very slowly within any one individual, we hold $T = T_0$ constant though not necessarily satisfying the parameter constraint $c = b k T_0$ resulting from $V$ and $T^*$ being at equilibrium. We then use this reduced model with $T = T_0 =$ taking the values 100, 400, and 1000:

\[
\frac{dT^*}{dt} = k T_0 V - \delta T^*, \quad \frac{dV}{dt} = b \delta T^* - c V, \\
\frac{dN}{dt} = \beta N - (\bar{\mu} + a N) N, \quad \frac{dS}{dt} = \beta N - (\bar{\mu} + a N) S - \rho \frac{SI}{N}, \quad N = S + I + R, \\
\frac{\partial i}{\partial t} = - (b \delta T^* - c V) \frac{\partial i}{\partial V} - (k T_0 V - \delta T^*) \frac{\partial i}{\partial T^*} - \gamma V, T^*) i - \mu V, T^*) i,
\]

with boundary conditions

\[
\frac{\partial i}{\partial V}(0, T^*) = \frac{1}{T_{max}} \frac{SI}{N}, \quad i(V, 0) = 0.
\]

During the last decades different numerical algorithms have been proposed for the approximation of solutions to size-structured models; we refer to Ref. 21 and the references therein for a review and a description of several numerical methods applied to this kind of model. Also, numerical methods have been successfully applied to structured models to replicate available field and/or laboratory data, for a variety of different systems.\textsuperscript{22–28}

The algorithm we present and use here is based on a similar formula to the one provided by Proposition 3.3, in discretized form. Thus, we establish the domain of our simulations as the Cartesian product of the biologically feasible values of $(V, T^*)$, $[0, V_{max}] \times [0, T_{max}^*]$ and we fix the maximum time of integration as $t_{max}$. The numerical method is based on a nonuniform bidimensional spatial grid and on integration along the characteristic curves; it also uses a composite quadrature rule to approximate the nonlocal term just as in Ref. 29.

4.1. Parameters

For the parameters in the immunological model, we used the following estimates found in Ref. 20:

\[
s = \text{rate of supply of CD4+ T cells from precursors} = 10 \text{ day}^{-1} \text{mm}^{-3} \\
p = \text{rate of growth for the CD4+ cell population} = 0.03 \text{ day}^{-1}
\]
$T_{\text{max}}$ = maximum CD4+ cell population level = 1500 mm$^{-3}$

$d_T$ = the death rate of uninfected (healthy) T cells = 0.02 day$^{-1}$

$\delta$ = the death rate of infected T cells = 0.24 day$^{-1}$

$c$ = the death rate of free virus = 2.4 day$^{-1}$

$k$ = the rate at which CD4+ cells become infected by free virus = $2.4 \times 10^{-5}$ mm$^3$ day$^{-1}$

While we find reliable estimates in the literature for most of the parameters in the model, it seems that the burst size $b$—the number of virions produced by an actively infected T cell during its lifetime—varies greatly.\textsuperscript{30} suggested that $b$ is between 50 and 1000; Ref. 31 estimated that $b > 300$ from data on the minimum concentration of soluble CD4+ needed to block HIV infectivity in an in vitro assay; others found that values of $b$ well above 1000 are possible.\textsuperscript{32}

The per capita death rate per day, $\mu$, was calculated from the life expectancy at birth in US being 77.9 years, using the relation $1 - 1/77.9 = (1 - \mu)^{365}$ that gives $\mu = 0.000035396843$. Similarly, the birth rate per day was calculated to be $\beta = 0.00003863122$. To estimate $\rho$, we used data obtained from the San Francisco Department of Public Health, HIV/AIDS Epidemiology Annual Report, HIV Epidemiology Section, 2008 and from the US Census website, reported below in Table 1. Taking into account the number of new cases for every year, a good estimate of $\rho$ seems to be $0.000095 \pm 0.000025$ per day (different years will give different results, as the number of new cases seems to change significantly from one year to another).

Table 1. Number of cases newly diagnosed with HIV, 2004-2008, San Francisco

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of new cases diagnosed</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>639</td>
</tr>
<tr>
<td>2005</td>
<td>518</td>
</tr>
<tr>
<td>2006</td>
<td>520</td>
</tr>
<tr>
<td>2007</td>
<td>518</td>
</tr>
<tr>
<td>2008</td>
<td>434</td>
</tr>
</tbody>
</table>

The initial values for $T^{*}_0$ and $V_0$ were taken from uniform distributions on [0, 80] and [0, 5000] respectively. Assuming for the USA a carrying capacity $K = 1,000,000,000$, we estimated $a$ using the formula $a = (\beta - \mu)/K = 3.234377 \times 10^{-15}$.

4.2. Numerical method

We introduce the notation $W$, $J^0_V$, $J^0_T$, for the number of steps in time and the initial number of subintervals for the state variables. Then, we define the time and spatial discretization parameters as $h_t = t_{\text{max}}/W$, $h_V = V_{\text{max}}/J^0_V$, $h_T^* = T_{\text{max}}^*/J^0_T$. The discrete time levels are $t^n = n h_t$, $0 \leq n \leq W$, and the initial grid nodes are
Also, we denote with \( N \) the numerical approximations at the time level \( t \), \( S \leq J^0 \), \( 0 \leq j_V \leq J^0 \), \( 0 \leq j_T \leq J^0 \). We assume that the approximations to the theoretical solution in those nodes are known, \( I^{n}_{j_V, j_T} \), \( 0 \leq j_V \leq J^0 \), \( 0 \leq j_T \leq J^0 \). These are approximations of the initial state density of the population with respect to the structure variables \( V \) and \( T^* \).

Let us now denote by \( I^{n}_{j_V, j_T} \), our numerical approximation at time level \( t^n \) of the analytical solution of the problem on the \( n \)-time grid, namely \( i((V, T^*)^{n}_{j_V, j_T}, t^n) \), where \( (V, T^*)^{n}_{j_V, j_T}, 0 \leq j_V \leq J^0 \), \( 0 \leq j_T \leq J^0 \), represents the grid for the structure variables at that time level \( t^n \) with \( (V, T^*)^{n}_{0, j_T} = (0, j_T \cdot h_T) \) and \( (V, T^*)^{n}_{j_V, 0} = (j_V \cdot h_V, 0) \). We shall see that the number of nodes is different at each time step. Also, we denote with \( N^n \), \( S^n \) and \( R^n \) the corresponding approximations to \( N(t^n) \), \( S(t^n) \) and \( R(t^n) \), \( 0 \leq n \leq W \).

We now describe the general step of our numerical method. Assume that we know the numerical approximations at the time level, \( t^n \), and we compute the approximations at the level \( t^{n+1} \), \( 0 \leq n \leq W - 1 \) as follows. First, we define the grid by means of a semi-implicit Euler method

\[
V^{n+1}_{j_V, j_T} = \frac{V^n_{j_V, j_T} + h_t b \delta T^{n, n}_{j_V, j_T}}{1 + c h_t}, \tag{4.5}
\]
\[
T^{n+1, n}_{j_V, j_T} = \frac{T^n_{j_V, j_T} + h_t k V^n_{j_V, j_T} T}{1 + h_t \delta}, \tag{4.6}
\]

\( 0 \leq j_V \leq J^0 \), \( 0 \leq j_T \leq J^0 \).

Next we calculate the corresponding approximation to the theoretical solution by means of the rectangle quadrature rule

\[
I^{n+1}_{j_V, j_T} = I^n_{j_V, j_T} \exp \left( -h_t \left[ \gamma (V^n_{j_V, j_T}, T^{n, n}_{j_V, j_T}) + \mu (V^n_{j_V, j_T}, T^{n, n}_{j_V, j_T}) \right] \right),
\]

\( 0 \leq j_V \leq J^0 \), \( 0 \leq j_T \leq J^0 \).

We compute next approximations to the populations \( N \) and \( S \) by means of the explicit Euler method,

\[
N^{n+1} = N^n + h_t (\beta N^n - (\bar{\mu} + a N^n) N^n), \tag{4.8}
\]
\[
S^{n+1} = S^n + h_t \left( \beta N^n - (\bar{\mu} + a N^n) S^n - \rho S^n \frac{I^n}{N^n} \right). \tag{4.9}
\]

We define \( I^{n+1} = \mathcal{Q}^{n+1}((V, T^*), I) \) as an approximation to the total number of infected, where the composite quadrature rule is based on the rectangle quadrature in which we do not use the first node in order to make our method explicit.

Finally, we discretize the boundary condition

\[
I^{n+1}_{j_V, 0} = 0, \quad 0 \leq j_V \leq J_V^{n+1}, \tag{4.10}
\]
\[
I^{n+1}_{0, j_T} = h_V \rho \frac{S^{n+1} I^{n+1}}{N^{n+1}}, \quad 1 \leq j_T \leq J_T^{n+1}. \tag{4.11}
\]

Note that at the new time level we have a new node that flux from zero, so we define \( J_V^{n+1} = J_V^n + 1 \) and \( J_T^{n+1} = J_T^n + 1 \). We stress the fact that all the steps in our algorithm are fully explicit.
4.3. Simulations

We performed several simulations with our numerical method in order to show the behavior of the model in different situations. As explained in the previous subsection, in order to avoid RAM problems we use a bidimensional grid for \( V \) and \( T^* \), and we select several values of \( T \).

We used a computational domain given by \((V, T^*) \in [0, 5000] \times [0, 80]\). The initial number of susceptibles was chosen as \( S^0 = 775046 \), the initial distribution of infected \( i(V, T^*; 0) \) as uniform with \( \tilde{I}^0 = 0.05 \), \( S^0 = 38752 \), and the total population as \( N^0 = S^0 + \tilde{I}^0 = 813798 \).

We computed at time \( t_{\text{max}} = 100 \) the approximations to the distribution of infected using an initial grid with \( J^0_{V} = 40000 \) and \( J^0_{T^*} = 640 \), and using different values of \( T \) and \( b \).

![Fig. 1. T = 100, b = 50. Left plot total population and susceptibles. Right plot total population of infected](image)

The left plot of Figure 1 shows the evolution of the total population and of the population of susceptibles, and in the right plot the evolution of infected population. We observe that the infected population tends to disappear. This is due to the domain getting smaller and the population density concentrating in a very small range along time, as we see in figure 2.

This computational (and immunological) domain increases with the value of \( T \), as we see in figure 3 and also with the value of \( h_i \) as we see in figure 4. Note that \( T = 100 \) corresponds to individuals with AIDS, \( T = 400 \) to infected immunocompromised, and \( T = 1000 \) to healthy.
In the last plots, figure 5, we show how infected individuals decrease slowly $T = 1000$ and $b = 100$, or how it increases quickly, $T = 700$, $b = 150$.

5. Conclusions

We have presented a new framework for modeling the evolution of a fairly general epidemic where infected individuals are structured by relevant immunological variables, the model consisting of coupled ordinary and partial differential equations. The model fits the general class of SIR epidemic models, though we do not discuss in this paper the meaning or specific modeling of the class of removed or recovered individuals, $R$.

The analysis of the resulting system proved to be unmanageable, and even local existence of solutions remains an open problem for the full model. We were able to prove well-posedness in the simpler case of a single structure variable reducing...
We then described in detail the particular case of three structure variables that model a viral infection, namely densities of CD4+ T cells, infected CD4+ T cells, and virions. In this case the boundary conditions that correspond to new infections cannot be of homogeneous Dirichlet type because uninfected individuals with no viral load at the time of infection would then evolve virus-free. We then described for this case the appropriate boundary conditions as homogeneous Dirichlet on the three boundary planes and on two of their lines of intersection, but non-homogeneous Neumann corresponding to increasing viral load at a rate proportional to new infections along the line determined by no viral load and no infected T cells.

We finally described in detail an explicit algorithm for the numerical approximation of solutions to the full model, and used it to obtain simulations of several situations that correspond to healthy individuals, HIV-positive, and AIDS, with parameters and population sizes drawn from the San Francisco epidemic. We showed how the biologically relevant domain for the variables corresponding to measures of severity of infection changes according to the T cell density. The results are interesting because they provide an insight into the immunological status of infected individuals in the population, without knowledge of the status of any specific individual. This can be useful, for example, for public health planning.

The full model we proposed is very general and can applied to many different epidemics. However, it provides difficult challenges both analytically and computationally due to the convergence of characteristics to equilibria of the dynamical system that governs the evolution of the structure variables. Local existence and uniqueness, as well as conservation of positivity of solutions are open problems, as is
Fig. 5. Evolution of infected individuals. Right plot: $T = 1000$ and $b = 100$. Left plot: $T = 700$ $b = 150$ individuals.

global existence even for the simplified problem we described and analyzed. These are problems the authors will address in their continuing research.

Acknowledgements

Óscar Angulo was supported in part by by Ministerio de Ciencia e Innovación (Spain), project MTM2011-25238.

References

Appendix A. Proofs

Proof. (of Proposition 1.1) Let us note that:

\[ dT \left| \begin{array}{c}
T=0 \\
T^*=0
\end{array} \right. = s > 0, \quad dT^* \left| \begin{array}{c}
T=0 \\
T^*=0
\end{array} \right. = kVT \geq 0, \quad dV \left| \begin{array}{c}
V=0 \\
T^*=0
\end{array} \right. = b\delta T^* \geq 0 \]

So if they ever reach 0, \( T, T^* \) and \( V \) will start growing. That takes care of the positivity part. To show they are bounded, let us first introduce the following notation,

\[ T_{\text{total}} = T + T^*. \]

By summing (1.1) and (1.2) side-by-side, we see that

\[ \frac{dT_{\text{total}}}{dt} = s + pT(1 - \frac{T_{\text{total}}}{T_{\text{max}}}) - dT - \delta T^*. \]

Since \( dT < \delta \) (the death rate of infected \( T \) cells is greater than the death rate of the healthy ones) we see that

\[ \frac{dT_{\text{total}}}{dt} \leq s + pT(1 - \frac{T_{\text{total}}}{T_{\text{max}}}) - dT_{\text{total}}. \]

Hence,

\[ \frac{dT_{\text{total}}}{dt} \left| \begin{array}{c}
T_{\text{total}} = T_{\text{max}} \\
T_{\text{total}} = T_{\text{max}}
\end{array} \right. = s - dT_{\text{max}} \leq 0, \]
according to the assumption that, if the total population of T cells ever reaches $T_{\text{max}}$, it will decrease. It follows that $T_{\text{total}} \leq T_{\text{max}}$, and the positivity of $T$ and $T^*$ ensures that $T \leq T_{\text{max}}$ and $T^* \leq T_{\text{max}}$. We note here that the assumption $dT/dt > s$ is purely biological, and that the result still holds in its absence.

To prove that $V$ is bounded, let us note that $V$ is bounded above by an obvious supersolution, $\tilde{V}$, solution of the ordinary differential equation:

$$\frac{d\tilde{V}}{dt} = b\delta T_{\text{max}} - c\tilde{V}, \quad \tilde{V}(0) = V_0,$$

which can be easily solved as

$$\tilde{V}(t) = (V_0 - \frac{b\delta T_{\text{max}}}{c})e^{-ct} + \frac{b\delta T_{\text{max}}}{c}.$$

That shows $\tilde{V}$ is bounded above, and concludes our argument. We will denote the least upper bound for $V$ as $V_{\text{max}}$.

**Proof. (of Proposition 3.1)**

It is obvious that if such a pair exists, then it is unique. Now we have to look at two separate cases:

1) Suppose $(t, V) \in D$ and $V \leq G(t; \epsilon, 0)$. Define $V_0 = \epsilon$ and $t_0 = H(\epsilon; V, t)$. Note that in this case the equation (3.5) is satisfied, and $(t_0, V_0) \in \Gamma_1$.

2) Suppose $(t, V) \in D$ and $V > G(t; \epsilon, 0)$. Define $t_0 = 0$ and $V_0 = G(0; V, t)$. Then (3.5) is again satisfied, and $(t_0, V_0) \in \Gamma_2$.

**Proof. (of Proposition 3.2)** The characteristic equations of (3.1) are:

$$\frac{dt}{ds} = 1, \quad t(0) = t_0, \quad (A.1)$$

$$\frac{dV}{ds} = b\delta T^*(t) - cV(t), \quad V(0) = V_0, \quad (A.2)$$

$$\frac{di}{ds} = -\alpha(V)i, \quad i(0) = i_0. \quad (A.3)$$

The solution of the initial value problem (A.1) is given by $t = s + t_0$ while that of (A.2) is given by $V = G(t; V_0, t_0)$, since we are assuming now that $T^*(t)$ is a known function. Finally, we find $i$ by integrating the separable equation (A.3):

$$\frac{di}{ds} + \alpha(G(t; V_0, t_0))i = \frac{di}{ds} + \alpha(G(s + t_0; V_0, t_0))i = 0, \quad (A.4)$$

to obtain

$$i(s) = i_0e^{-\int_{t_0}^{t} \alpha(G(\sigma + t_0; V_0, t_0))d\sigma} = i_0e^{-\int_{t_0}^{t} \alpha(G(\sigma; V_0, t_0))d\sigma}. \quad (A.5)$$

Now we have to consider two cases: when characteristics originate on the line $V = \epsilon$ ($\Gamma_1$) and when characteristics originate on the line $t = 0$ ($\Gamma_2$).
we have
\[ i(t,V) = \frac{\rho S(t_0) I(t_0)}{N(t_0)} e^{-\int_{t_0}^{t} \alpha(G(\sigma,x,t_0)) d\sigma} \tag{A.6} \]
where \( t_0 = H(\epsilon;V,t) \).

**Case 2:** The characteristics originate on \( \Gamma_2 \). The appropriate boundary condition to use now is \( i(0,V) = \phi(V) \). When \( s = 0 \), we have \( V = V_0 \) and \( i = \phi(V_0) \). According to the proof of Proposition 3.2, \( t_0 = 0 \) and \( V_0 = G(0;V,t) \). We then have
\[ G(\sigma;V_0,0) = G(\sigma;G(0;V,t),0) = G(\sigma;V,t) \]
which concludes our proof. \( \square \)

**Proof. (of Theorem 3.1)**

In order to evaluate \( \| \Phi(i_1) - \Phi(i_2) \|_{\infty} \) we have to look at two separate cases.

**Case 1:** \( (t,V) \in [0,\eta] \times [0,\bar{V}] \) with \( V \leq G(t;\epsilon,0) \). For such values of \( t \) and \( V \), we have
\[ \tilde{i}(t,V) = \frac{\rho S(t_0) I(t_0)}{N(t_0)} e^{-\int_{t_0}^{t} \alpha(G(\sigma,x,t_0)) d\sigma} \]
Note that \( N \) does not depend on \( I \), but \( S \) actually does. We then have
\[
|\Phi(i_1)(t,V) - \Phi(i_2)(t,V)| = \left| \frac{\rho S_1(t_0) I_1(t_0)}{N(t_0)} e^{-\int_{t_0}^{t} \alpha(G(\sigma,x,t_0)) d\sigma} \right| - \left| \frac{\rho S_2(t_0) I_2(t_0)}{N(t_0)} e^{-\int_{t_0}^{t} \alpha(G(\sigma,x,t_0)) d\sigma} \right|
\]
\[
= \frac{\rho}{N(t_0)} \left| S_1(t_0) I_1(t_0) - S_2(t_0) I_2(t_0) \right|
\]
if we assume that \( N(t) \) is always greater than or equal to 1. It remains now to evaluate the term \( |S_1(t_0) I_1(t_0) - S_2(t_0) I_2(t_0)| \). We know that
\[ I_1(t) = \int_{\epsilon}^{V_{\max}} i_1(t,V) dV, \quad I_2(t) = \int_{\epsilon}^{V_{\max}} i_2(t,V) dV, \]
\[
\frac{dN}{dt} = \beta N - (\mu + aN) N,
\]
\[
\frac{dS}{dt} = \beta N - (\mu + aN) S - \frac{\rho S(t) I(t)}{N(t)}.
\]
Thus, for every $t$ we have
\[
|S_1(t)I_1(t) - S_2(t)I_2(t)| \leq |S_1(t)I_1(t) - S_1(t)I_2(t)| + |S_1(t)I_2(t) - S_2(t)I_2(t)|
\]
\[
\leq |S_1(t)||I_1(t) - I_2(t)| + |I_2(t)||S_1(t) - S_2(t)|,
\]
and, writing $I_1$ and $I_2$ explicitly in terms of $i_1$ and $i_2$ we obtain
\[
|I_1(t) - I_2(t)| = \left| \int_{\epsilon}^{V_{\max}} i_1(t, V) dV - \int_{\epsilon}^{V_{\max}} i_2(t, V) dV \right|
\]
\[
= \int_{\epsilon}^{V_{\max}} |i_1(t, V) - i_2(t, V)| dV
\]
\[
\leq \int_{\epsilon}^{V_{\max}} |i_1(t, V)| - |i_2(t, V)| dV
\]
\[
\leq \int_{\epsilon}^{V_{\max}} \|i_1 - i_2\|_{\infty} dV
\]
\[
\leq (V_{\max} - \epsilon)\|i_1 - i_2\|_{\infty}.
\]

Also, using the fact that $S$ is bounded, say by $S_{\max}$, we have
\[
|S_1(t)I_1(t) - S_1(t)I_2(t)| \leq S_{\max}(V_{\max} - \epsilon)\|i_1 - i_2\|_{\infty} \tag{A.10}
\]
Now we need to evaluate the remaining term in (A.8), $|S_1(t) - S_2(t)|$. We solve (2.3) for $S$ in terms of $I$ and $N$ as
\[
S(t) = S_0 e^{-\int_0^t f(s)ds} + e^{\int_0^t f(s)ds} \int_0^t \beta N(\tau) e^{\int_0^\tau f(s)ds} d\tau, \tag{A.11}
\]
where $f(t) = \mu + aN(t) + \frac{\rho(t)}{N(\tau)}$. We now have
\[
|S_1(t) - S_2(t)| = \left| S_0 e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} + e^{\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau \right|
\]
\[
- S_0 e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau
\]
\[
- e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau
\]
\[
\leq \left| S_0 e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} - S_0 e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \right|
\]
\[
+ \left| e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau \right|
\]
\[
- e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau
\]
\[
+ \left| e^{-\int_0^t (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho(s)}{N(\tau)})ds} d\tau \right|. \tag{A.12}
\]
We now introduce the following notation:

\[
M = \left| S_0 e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} - S_0 e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \right|
\]

and

\[
P = \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

Now, for every \( t \), \( |S_1(t) - S_2(t)| \leq M + P \) and we need to bound \( M \) and \( P \).

\[
M = \left| S_0 e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} - S_0 e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \right| \leq S_0 \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} - e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \right| \tag{A.14}
\]

Now, by the Mean Value Theorem, there exists a number \( r < 0 \), between \( -\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \) and \( -\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \), such that

\[
e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} - e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} = e^r \left[ -\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds + \int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \right],
\]

and using this relation in (A.14) we see that

\[
M \leq S_0 e^r \left| \int_0^t \frac{r}{N(s)} (I_1(s) - I_2(s)) \, ds \right| \leq S_{\text{max}} r e^r (V_{\text{max}} - \epsilon) \|i_1 - i_2\|_{\infty}.
\]

Now we consider \( P \). We have

\[
P = \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

\[
\leq \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

\[
+ \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

\[
+ \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

\[
+ \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]

\[
+ \left| e^{-\int_0^t (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds} \int_0^t \beta N(\tau) e\int_0^\tau (\hat{\mu} + aN(s) + \frac{\rho I_1(s)}{N(s)}) \, ds \, d\tau \right|
\]
Since
\[
\int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} d\tau \leq \eta \beta N_{\text{max}} e^{\eta(\mu + aN_{\text{max}} + \rho I_{\text{max}})},
\]
where \( N_{\text{max}} = \sup_{0 \leq t \leq T} \{ N(t) \} \), it follows that the first term can be bounded as follows:

\[
|(1)| \leq \eta \beta N_{\text{max}} e^{\eta(\mu + aN_{\text{max}} + \rho I_{\text{max}})} \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} d\tau \leq \eta \beta N_{\text{max}} e^{\eta(\mu + aN_{\text{max}} + \rho I_{\text{max}})} \eta \rho (V_{\text{max}} - \epsilon) ||i_1 - i_2||_{\infty} = \eta^2 \beta N_{\text{max}} (V_{\text{max}} - \epsilon) e^{\eta(\mu + aN_{\text{max}} + \rho I_{\text{max}})} ||i_1 - i_2||_{\infty}.
\]

Concerning the second term, we have

\[
|(2)| \leq \left| \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} d\tau - \int_0^t \beta N(\tau) e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} d\tau \right| \\
\leq \int_0^t \beta N(\tau) \left| e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} d\tau - e^{\int_0^\tau (\mu + aN(s) + \frac{\rho I(s)}{N(s)}) ds} \right| d\tau \\
\leq \int_0^t \beta N(\tau) e^{\mu + aN_{\text{max}}} \rho \eta (V_{\text{max}} - \epsilon) ||i_1 - i_2||_{\infty} d\tau \\
\leq \beta N_{\text{max}} e^{\mu + aN_{\text{max}}} \rho \eta^2 (V_{\text{max}} - \epsilon) ||i_1 - i_2||_{\infty}.
\]

Using these relations we now obtain

\[
||\Phi(i_1)(t, V) - \Phi(i_2)(t, V)|| \leq q ||i_1 - i_2||,
\]

where \( q = \rho S_{\text{max}} (V_{\text{max}} - \epsilon) + \rho^2 \eta I_{\text{max}} S_{\text{max}} (V_{\text{max}} - \epsilon) + \rho^2 \eta^2 \beta N_{\text{max}} (V_{\text{max}} - \epsilon) e^{\eta(\mu + aN_{\text{max}} + \rho I_{\text{max}})} + \rho^2 \eta^2 \beta N_{\text{max}} e^{\mu + aN_{\text{max}} (V_{\text{max}} - \epsilon)} \)

We note that, for \( \rho \) and \( \eta \) sufficiently small, we have \( q < 1 \).

**Case 2:** \((t, V) \in [0, \eta] \times [0, V] \) with \( V > G(t; \epsilon, 0) \). For such values of \( t \) and \( V \), we have:

\[
\tilde{i}(t, V) = \phi(G(0; V, t)) e^{-\int_0^t \alpha(G(s; V, t)) ds}
\]

and thus

\[
||\Phi(i_1)(t, V) - \Phi(i_2)(t, V)||_{\infty} = 0.
\]

We have established that \( \Phi \) is a contraction, and hence it admits a unique fixed point, which is precisely the solution of our partial differential equation. \( \square \)