SIR Models

Overview and Application

Danilo Menegatti and Christian D'Amico

La Sapienza, University of Rome
Control Engineering
Optimal Control

Contents

1 Abstract

This project, starting from the definition of epidemiology, gives an insight about compartmental models, focusing on the SIR one presented in [33]. Control and parameter identification problems are formulated. In particular, real-life problems concerning Ebola are addressed.

2 Epidemiology

Epidemiology is a branch of medicine which deals with the incidence, distribution, and possible control of diseases and other factors relating to health. Literally, it means "the study of what is upon the people", since it is derived from Greek **epi**, meaning "upon, among", **demos**, meaning "people, district" and **logos**, meaning "study, word, discourse". As a result, its etymology suggests that it applies only to human populations. It was first used by the Spanish physician Villalba in *Epidemiologìa Españ ola* to describe the stupy of epidemics[1] in 1802. Nowadays, the term is widely used for the description and causation of disease in general and even non-disease like blood pressure and obesity.

The goal of an epidemiological investigation is the identification of **casual** relationship between *exposures* and *outcomes*. The former are defined as the contact between an *agent* - the cause of the disease, a microorganism, chemical substance or a form of radiation for example - and a target, a person or, in general, a group of people referred to as **target population**. The latter are defined as a broad term representing a defined disease, state of health, or health-related event - disease or death for example.

When it comes to the design of an epidemiological study, one may choose between an *observational* or *experimental* one. In the former, one observes nature following its course - exposure and disease study of each study participant. One may think about John Snow's studies of cholera in London. In the latter, one controls all of the factors entering a certain case study. One may think about a clinical trial of a new vaccine. Moreover, observational studies can be generally categorized as *descriptive* and *analytical*. The former seek answers to question like "Who, What, Where and When?" about the event. The latter tries to answer the question "What is happening in a specific time period?" in order to find casual associations between exposures and outcomes.

Experimental studies distinguish in *randomized controlled trials*, *field trials* and *community trials*. In order to fix ideas, one may think about the test of a new medicine, tests on those at high risk of contracting a disease and research on social originating diseases respectively. It's important to point out that a study refers to a *study sample* because if the target consists in a large number of people, then it is impossible to study each one of them, that's why a sample from the target population is taken.

[1] One refers to a disease that is "visited upon" a population.

3 Models

The need of understanding how an health-related event works and of predicting its outcomes in the future, led to the use of mathematical models. A **mathematical model** is an abstract model that uses mathematical language to describe the behaviour of a system. More specifically, Eykhoff defined it as "a representation of the essential aspects of an existing system - or a system to be constructed - which presents knowledge of that system in usable form".

With respect to epidemiology, it can be used to study the mechanisms by which a disease spread, to predict its future course and to evaluate control strategies. The first scientist who tried to quantify in a systematic way the causes of death was John Graunt in his book *Natural and Political Observations made upon the Bills of Mortality*, in 1662. The first mathematical modelling of a disease' spread was performed by Daniel Bernoulli in 1766. A traditional model of infectous disease causation is the *Epidemiological Triad*. It consists in an external **agent**, a **host** and an **enviroment**. The host is the organism which carries the disease since it's hosptial or attractive to the agent; it can get sick or not. The enviroment corresponds to outside factors which affect the epidemiological outbreak. One may think about temperature, the quality of drinking water and so on. It is even possible to include a **vector** which transmits infection by conveying the pathogen[2] from one host to another one without causing disease itself. The agent infects the host and transmission occurs when the agent leaves its host or reservoir through a **port of exit**, is conveyed by a **mode of transmission** to enter through an appropriate **portal of entry** to infect a **susceptible** host.

Transmission may be **direct** or **indirect**. The former corresponds to direct contact host-to-host while the latter to the transfer of the agent from a reservoir to a susceptible host by suspended air particles, inanimate objects - vehicles - or animate intermediaries - vectors -. In order to fix ideas, let's consider a Smoking-Related-Disease where the *agent* is a carcinogen[3] in the smoke of a cigarette, the **host** is a person who smokes cigarettes or inhale second-hand smoke and the *enviroment* is susceptibility to marketing efforts, smoking frequency, the high addiction of tobacco products and so on. One recalls that a susceptible host is a member of a population who is at risk of becoming infected by a disease. Talking about individuals let us suddenly and explicitely think about the population they are members of; in this perspective, it's resonable to think that susceptible individuals eventually may get sick or not, leading to a division of population in classes.

This is the basic idea under **compartimental models**, mathematical models investigated through a set of ordinary differential euqations, where the population is divided into compartment, with the assumption that each individual in it has the same characteristics. Being introduced in the early 1920, one of the first models to successfully predict the behaviour of outbreakes very similar to the one observed in many epidemics was the *Kermack-McKendrick* one. In particular, it considered a fixed population with only three compartements: **susceptible**, **infected** and **recovered**, denoted rispectively with S(t), I(t) and R(t). S(t) represents the number of individuals, members of the population, that are

[2] A bacterium, virus, or other microorganism that can cause disease.

[3] Any substance, radionuclide, or radiation that promotes carcinogenesis, the formation of cancer.

not yet infected with the disease at time t and are at risk of being infected. I(t) representes the number of individuals, members of the population, that are infected with the disease at time t and are capable of spreading it to the susceptible ones. R(t) represents the number of individuals, members of the population, that have been infected with the disease but recovered from it, either due to immunization[4] or due to death. One assumes that those in this cathegory are not able to be infected again or to transmit the infection to others.

One notices that they are function of time, which means that the number of susceptible, infected and recovered individuals may vary over time. This is reasonable if one refers to real-life experience like "influenza". It is a respiratory illness caused by a virus depending on the which it may last from three to seven days in most people; after that they are completely recovered. Moreover, one points out that each disease has its specificity. It means that, for example, it has its own *latent period* - period between exposure and infection during which there are no symptoms or sign of infection in the host -, its own *basic reproduction number* - number of people that the disease can spread to, on average, by one infected individual, its own *average duration* in addition to its specific symptoms and so on.

One should be able to add these features, also by means of parameters, in order to make the model consistent with the real-life experience. However, due to the complexity of the phenomenon - one may think about the specificity and subjectivity of a disease as long as the fact that many factors contribute to it - *simplification hypothesis* are made in order to simplify the problem as much as possible. One means that the problem "should be made simple as possible, but not simpler"[5]. This means that the model would be able to exactly follow the disease but can be able to approximate it. In compact form, this concept can be expressed as the following: "All models are wrong, but same are useful"[6].

Before adressing a particular SIR model, one recalls that there are a large number of modifications of it. One may talk about *SIS* models where there is no immunity, *SIRS*, where immunity lasts only for a short period of time, *SEIS* and *SEIR* where there is a latent period of the disease where the individual is not infectious and *MSIR* where infants can be born with immunity.

3.1 Building of the SIR model

Now, one considers the building of a SIR model with time-varying population which accounts for deaths from a disease. Suppose to have a population of size N at a given instant of time t and a disease so that the former can be divided into susceptible, infected and recovered. Suppose that the number of members of the population may vary from time to time since there are newborns and deaths, with the latter either natural or due to the disease. Suppose that, at a given instant of time, a certain number of individuals get sick and change their "status" from susceptible to infected. Suppose that, at a given instant of time, a certain number of infected recovers without the need of any cure while a certain number of recovered become susceptible again.

[4]Process whereby a person is made immune or resistant to an infectious disease, typically by the administration of a vaccine.

[5]Adressed to Einstein, it compactly articulates the principle of Occam's razor.

[6]George E. P. Box, one the greatest statistical minds of the 20th century.

By means of the previous considerations, the basic assumptions upon which it is built are the following:

- The population considered has constant size N for a fixed instant of time and is sufficiently large so that the sizes of each class can be considered as continuous variables.
- The entire population of new born is affected by the disease.
- Once recovered, an individual becomes susceptible again.
- Age, sex, social status do not affect the probability of being infected.
- There is no inherited immunity.
- The member of the population mix homogeneously - have the same interactions with one another to the same degree.

It follows that constant parameters are defined; γ represents the birth rate per unit of time, ν the regular death rate in the population and it's equally applied to all three compartments, μ the deaths due to the disease, α rate at which infectious individuals overcome the disease naturally without treatment, β models the probability that a contact between an individual from susceptible class with another one is strong enough for a transmission of the disease to occur, ρ rate at which recovered individuals become susceptible.

As a result, the dynamics are described by the following equations:

$$\begin{cases} \dot{S} = \gamma N - \nu S - \beta \frac{IS}{N} + \rho R, & S(0) = S_0 \\ \dot{I} = \beta \frac{IS}{N} - (\nu + \mu) I - \alpha I, & I(0) = I_0 \\ \dot{R} = -\nu R - \rho R + \alpha I, & R(0) = R_0 \end{cases} \tag{1}$$

With the addition of the dynamics of the population:

$$\dot{N} = (\gamma - \nu) N - \mu I, N(0) = N_0 \tag{2}$$

The fact that the population varies in time matters in the transition of susceptible to infected individuals.

Supposing to model this transfer in a probabilistic way, then the flow from S to I is given by $\beta \frac{I(t)}{N(t)} S(t)$. As a result, for an individual from the susceptible class in contact with another individual, the probability that this individual is infectus is given by $\frac{I(t)}{N(t)}$ while β has the previously defined meaning.

Before moving to the mathematical analysis, one wants to further investigate the notion of time. Even if one is dealing with continous variables, in real-life time is not continuous but discrete. It means that one has to interpret it as days or weeks. This is reasonable since the number of individuals of a population cannot be monitored every second since it would be useless because it's very unlikely that there is a considerable increment/decrement of population within the given time interval.

4 Mathematical analysis

One recalls that since the population varies in time, this system has no equilibrium point. However, in order to further investigare the model's mathematical properties, one considers the associated prevalence model with no time-varying population. One recalls that prevalence looks at existing cases and it's defined as:

$$\frac{\textbf{number of cases}}{\textbf{population}} \times 100 \tag{3}$$

Defining the proportions of individuals in the respective compartments,

$$x = \frac{S}{N}, y = \frac{I}{N}, z = \frac{R}{N}$$

with $x, y, z \geq 0$ since one is talnking about individuals.

Moreover, since $S + I + R = N$, it follows that, by means of substitutions, $x + y + z = 1$ and that assigning two of the three variables means that the third one is set.

As a result, the system is redundant - one may consider only two of the three equations.

$$\begin{cases} \dot{x} = (\mu - \beta)xy - \gamma x + \rho z + \gamma \\ \dot{y} = y(\beta x + \mu y - (\gamma + \mu + \alpha)) \\ \dot{z} = \mu yz + \alpha y - (\gamma + \rho)z \end{cases} \tag{4}$$

with $\alpha, \beta, \gamma, \mu, \rho \geq 0$.

By means of the previous considerations on x, y, z, it follows that the domain for the given dynamical system is:

$$\Sigma = \{(x, y, z) : x \geq 0, y \geq 0, z \geq 0, x + y + z = 1\} \tag{5}$$

In particular it is a positive invariant set since, considering an initial point belonging to it - $(x_0, y_0, z_0) \in \Sigma$, since $x + y + z = 1$ and $x \geq 0, y \geq 0, z \geq 0$, it is constrained to be in Σ $\forall t$.

One would like to investigate what happens when no disease is present. In other words, one is interested in the **disease-free equilibrium** (**DFE**) which is defined as the point at which no disease is present in the population. It can be found setting $y = 0$ and analyzing the corresponding system:

$$\begin{cases} \dot{x} = -\gamma x + \rho z + \gamma \\ \dot{z} = -(\gamma + \rho)z \end{cases} \tag{6}$$

It follows that $(x_0, y_0, z_0) = (1, 0, 0)$ is the DFE one is looking for. In order to analyze its stability, one computes the Jacobian matrix associated to the system and evaluates the latter at that point.

$$J = \begin{bmatrix} -\gamma & \mu - \beta & \rho \\ 0 & \beta - (\gamma + \mu + \alpha) & 0 \\ 0 & \alpha & -(\gamma + \rho) \end{bmatrix} \tag{7}$$

It follows that its eigenvalues are $\lambda_1 = -\gamma$, $\lambda_2 = -(\gamma + \rho)$, $\lambda_3 = \beta - (\gamma + \mu + \alpha)$. Recalling that these parameters are non-negative, DFE is locally asymptotically stable if $\beta < (\gamma + \mu + \alpha)$. That is because as soon as one has one infectious individual, one goes out of that point. In particular, if one has a disease, it's feasible to think that there are infectious people. Moreover, it's possible to introduce the *basic reproduction number* R_0 which plays a key role in analyzing the spread of the disease. It is defined as the average number of secondary infections caused by a single infectious individual during their entire infectious lifetime. More easily, it is the number of people that the disease can be spread to (on average) by one infected individual. It was originally developed for demographics (1886), studied for malaria (1911,1927) and then used or infectious disease (1975).

$$R_0 = \frac{\beta}{\gamma + \mu + \alpha} \tag{8}$$

Considering $\dot{I} = \beta \frac{IS}{N} - (\nu + \mu)I - \alpha I$, calling $\beta = F$ and $(\nu + \mu + \alpha) = V$, it is defined as $R_0 = FV^{-1} = \frac{\beta}{\gamma+\mu+\alpha}$. It gives a threshold criterion, because depending on its value, one gets information about the spread of the disease.

- If $R_0 < 1$, the disease will die out since each individual produces, on average, less than one new infected individual.
- If $R_0 > 1$, the disease will take hold since each individual produces more than one new infected individual.
- If $R_0 = 1$, the disease will become endemic - it will move thorughout the population but not increase or decrease since each person will infect exactly on other person.

Even though this threshold is useful, R_0 is rarely measured in the field since it is an individual parameter. **If** $R_0 < 1$, the DFE equilibrium solution is globally asymptotically stable since Σ lies in its region of attraction. It follows from the *Poincaré Bendixson* theorem: "*Given* Ω *a subset of* R^2*, supposing that exists* $K \subset \Omega$*, compact such that*

$$\Phi^t(K) \subset K \tag{9}$$

Let $x \in K$ *such that* $L_w(x)$ *doesn't contain critical points, then it is a periodic orbit of the system*". That is because, by Σ's invariance any solution to the dynamics starting in Σ exists for all times $t \geq 0$ and remains in Σ. Moreover, by means of its compactness, its corresponding *w-limit* is nonempty.

Since the DFE is the only equilibrium solution in Σ, it is the w-limit set. It follows that all trajectories starting in Σ, converge to the given point.

One recalls that $\Phi^t(K)$ is the flux associated to K. One recalls that K is said compact since for every open cover of Ω, the exists a finite subcover of Ω, where for cover one refers to a collection of sets whose union contains Ω as a subset. One recalls that critical points are related with the notion of derivative of a function or a mapping, denoting when it is not defined[7]. One recalls that the *w-limit* $L_w(x)$, suppsoing that the flux exists $\forall t > 0$ is the subset of R^N defined as

[7]One is referring to the concept, rather than considering particular cases.

$$L_w(x) = \{y \in R^N : \exists t_k, t_k \to +\infty, lim\Phi^{t_k}(x) = y\} \tag{10}$$

If $R_0 > 1$, there exists a unique endemic equilibrium point (x_*, y_*, z_*) which is locally asymptotically stable and its region of attraction $A(E)$ is given by

$$A(E) = \{(x_0, y_0, z_0) \in \Sigma : y_0 > 0\} \tag{11}$$

In this case the DFE is an unstable equilibrium point, in particular it is an *hyperbolic* one since the Jacobian evaluated at that point has no eigenvalues with zero real part. Moreover, there are several properties holding around its neighborhood, like the existance of a stable manifold and an unstable one. One recalls that a *manifold* is a space where any neighborhood of a point is diffeomorphic - "look like" - to a neighborhood of R^n, supposing to be in that space. It follows that the *hyperplane* - subspace whose dimension is one less than the ambient space - $\{y = 0\}$ is a stable manifold.

The dynamics for $y = 0$ is given by

$$\begin{cases} \dot{x} = -\gamma x + \rho z + \gamma \\ \dot{z} = -(\gamma + \rho)z \end{cases} \tag{12}$$

so that $z(t) \to 0$ as $t \to \infty$ and consequently $x(t) \to 1$ as $t \to \infty$.

As a result, all trajectories in the plane $\{y = 0\}$ converge to $(1, 0, 0)$. However, there exists another equilibrium point in Σ which can be computed considering $y > 0$. Considering $\dot{y} = 0$ and "dividing" for y, it follows that

$$\beta x + \mu y - (\gamma + \mu + \alpha) = 0 \tag{13}$$

$$\beta x + \mu y = (\gamma + \mu + \alpha) \tag{14}$$

Since $x + y + z = 1$, one explicits $x = 1 - y - z$

$$\beta(1 - y - z) + \mu y = (\gamma + \mu + \alpha) \tag{15}$$

$$\beta - \beta y - \beta z + \mu y = (\gamma + \mu + \alpha) \tag{16}$$

Expliciting z

$$z = \frac{\beta - \beta y + \mu y - \gamma - \mu - \alpha}{\beta} \tag{17}$$

$$z = 1 - \frac{\beta - \mu}{\beta} - \frac{\gamma + \mu + \alpha}{\beta} \tag{18}$$

By means of R_0

$$z = 1 - \frac{1}{R_0} + \frac{\mu - \beta}{\beta} y \tag{19}$$

Substituting into $\dot{z} = 0$

$$\alpha y + (\mu y - (\gamma + \rho))(1 - \frac{1}{R_0} + \frac{\mu - \beta}{\beta} y) = 0 \tag{20}$$

This expression can be re-written as $Q(y) = 0$, with

$$Q(y) = a_0 + a_1 y + a_2 y^2 \tag{21}$$

with

$$\begin{cases} a_0 = (\gamma + \rho)(\frac{1}{R_0} - 1) \\ a_1 = \alpha + \mu(1 + \frac{1}{R_0}) - \frac{(\gamma+\rho)(\mu-\beta)}{\beta} \\ a_2 = \frac{\mu-\beta}{\beta} y \end{cases} \tag{22}$$

In order to find the value of y, since $Q(y)$ is a quadratic function, by means of the *fundamental theorem of algebra*[8], it admits two roots y_1 and y_2.

Considering $\tilde{y_1} = \frac{\gamma+\rho}{\mu}$ and $y_2 = (1 - \frac{1}{R_0})\frac{\beta}{\beta-\mu}$ - they correspond to getting $Q(y) = \alpha y$ - one gets

$$Q(\tilde{y_1}) = \alpha \frac{\gamma + \rho}{\mu} = \alpha \tilde{y_1} \tag{23}$$

$$Q(\tilde{y_2}) = \alpha \frac{\beta}{\beta - \mu}(1 - \frac{1}{R_0}) = \alpha \tilde{y_2} \tag{24}$$

It follows that, if $\tilde{y_1} > 0$, then $Q(\tilde{y_1}) > 0$ and if $\tilde{y_2}$, then $Q(\tilde{y_2}) > 0$.

As a result, $Q(y)$ may be represented as a parabola with concavity towards the bottom. Recalling that y_1 and y_2 are $Q(y)$'s roots, and that since $R_0 > 1$, $0 < \tilde{y_2} < 1$; moreover, since $\tilde{y_1} > 0$, $0 < y_1 < 1$, it is possible to write that:

$$y_1 < min\{\tilde{y_1}, \tilde{y_2}\} < max\{\tilde{y_1}, \tilde{y_2}\} < y_2 \tag{25}$$

Since $x + y + z = 1$, $0 \leq y < 1$, one chooses y_1.

In order to explicit x, one considers

$$\beta x = \gamma + \mu(1 - y) - \alpha \tag{26}$$

$$x = \frac{\gamma + \mu(1 - y) + \alpha}{\beta} \tag{27}$$

In particular, considering y_1

$$x_1 = \frac{\gamma + \mu(1 - y) + \alpha}{\beta} \tag{28}$$

In order to explicit z, one considers the combination of its expression with the $Q(y) = 0$ one

[8] *Any non-constant single-variable polynomial with complex coefficients has at least one complex root.*

$$\alpha y + (\mu y - (\gamma + \rho))z = 0 \tag{29}$$

$$z = \frac{\alpha y}{\gamma + \rho - \mu y} \tag{30}$$

In particular, considering y_1

$$z_1 = \frac{\alpha y_1}{\gamma + \rho - \mu y_1} \tag{31}$$

$z_1 > 0$ since $y_1 < \tilde{y_1}$. As a result, there exists and it is unique an equilibrium point $(x_1, y_1, z_1) = (x_*, y_*, z_*)$ which belongs to the interior of Σ.

By means of the previous considerations, for y_2 one would consider an equilibrium point lying outside Σ. Applying the same reasoning as before - Poincaré-Bendixson theorem - one shows that it is locally asymptotically stable. The same results can also be shown using *Lyapunov stability theorem* with respect to

$$V(y, z) = y - y_* lny + \frac{1}{2}\frac{\beta}{\mu z + \alpha}(z - z*)^2 \tag{32}$$

and applying LaSalle theorem in order to prove local asymptotic stability.

5 Control problem

Once one modelled the spread of a disease through a model - SIR, in particular -, it is even possible to understand how to counteract its outbreak. The basic idea under counteraction is **vaccination** and **treatment**. Even though in both cases drugs are injected, the former is a form of prevention while the latter cures disease. One recalls that prevention is a prodecure through which individuals, particularly those with risk factors for a disease, are treated in order to prevent a disease from occurring. Treatment normally begins either before signs and symptoms of the disease occur, or shortly thereafter. When it comes to their implementation in the model, one starts with assumptions.

- Only the susceptible individuals of the population undergoes vaccination.
- Individuals that have been indentified with the disease ma undergo treatment.

One denotes by u the vaccination rate - $0 \leq u \leq u_{MAX}$ and by v the rate at which infectious individuals are treated - $0 \leq v \leq v_{MAX}$. Moreover, one introduces two parameters, κ and η, representing the effectiveness of vaccination and treatment respectively.

$$\begin{cases} \dot{S} = \gamma N - \nu S - \beta\frac{IS}{N} + \rho R - \kappa S u, & S(0) = S_0 \\ \dot{I} = \beta\frac{IS}{N} - (\nu + \mu)I - \alpha I - \eta I v, & I(0) = I_0 \\ \dot{R} = -\nu R - \rho R + \alpha I + \eta I v, & R(0) = R_0 \end{cases} \tag{33}$$

The goal is, given initial population sizes of all three classes, S_0, I_0, R_0, to find the best strategy in terms of the combined efforts of vaccination and treatment that minimizes the

number of infectious persons while at the same time also taking into account the cost of vaccination and treatment. Since such a goal has to be expressed mathematically, assumptions are made. In particular, one assumes that the cost of intervention modalities is proportional to the vaccination and treatment rates and therefore model the objective as an affine functional in the controls.

Before considering the *optimal control problem formulation*, one writes the system in a different way. Denoting the state of the system by x, $x = (S, I, R)^T$, the system can be written as a multi-input control-affine system of the vector form

$$\dot{x} = f(x) + g_1(x)u + g_2(x)v \tag{34}$$

with drift vector field f given by

$$f(x) = \begin{pmatrix} \gamma N - \nu S - \beta \frac{IS}{N} + \rho R \\ \beta \frac{IS}{N} - (\nu + \mu + \alpha) I \\ \alpha I - (\nu R + \rho) R \end{pmatrix} \tag{35}$$

and control vector fields g_1 and g_2 given by

$$g_1(x) = \begin{pmatrix} -\kappa S \\ 0 \\ \kappa S \end{pmatrix} \tag{36}$$

$$g_2(x) = \begin{pmatrix} 0 \\ -\eta I \\ \eta I \end{pmatrix} \tag{37}$$

It follows that, *given the nonlinear system* $\dot{x} = f(x) + g_1(x)u + g_2(x)v$ *affine with respect to inputs, with fixed initial state* $x_0 = (S_0, I_0, R_0)^T$, *initial time* $t = 0$ *and final time* T, *with* $u \in [0; u_{MAX}]$ *and* $v \in [0; v_{MAX}]$, *minimize*

$$J = J(u, v) = \int_0^T (aI(t) + bu(t) + cv(t))dt \tag{38}$$

Find u_*, v_*, x_* *which satisfy the dynamic, initial constraints and minimize the objective.*

One recalls that $aI(t)$ represents an average of the number of infectious individuals over the full interval and a, b, c are weights of choice. As a result, the objective is a weighted average of the number of infectious persons and cost of vaccination and treatment. In order to solve this problem, it is convenient to define the Hamiltonian function $H = H(x, u, v, \lambda_0, \lambda)$ as

$$H = \lambda_0 L(x) + \lambda^T (f(x) + g_1(x)u + g_2(x)v) \tag{39}$$

where λ_0 is a scalar, λ is a vectorial function of three components and $L(x)$ is the lagrangian function in the objective.

Let (x^*, u^*, v^*) be an admissible solution, necessary and sufficient conditions to be a local minimum is that there exists a constant $\lambda_0^* \geq 0$ and $\lambda^* : [0, T] \to (R^3)^*$ not simultanously zero such that

- $\dot{\lambda^*}(T) = -\frac{\partial H}{\partial x}|^{*^T}$
- $\lambda^*(T) = 0$
- $H(x^*(t), w, \lambda_0^*, \lambda^*(t)) \geq H(x^*(t), u^*(t), \lambda_0^*, \lambda^*(t)), \forall w \in U$
- $H|^* = k, \forall t \in [0, T], k \in R$

One notices that since λ_0 and $\lambda(t)$ do not vanish simultaneoudly and since $\lambda(T) = 0$, without loss of generality one chooses $\lambda_0 = 1$.

As a result, all extremals - solutions of the Pontryagin maximum principle - are *normal*.

Since u and v are restricted to being between an upper and lower bound, one expects the optimal controller to be a **bang-bang solution**. However, one notices that the Hamiltonian is linear with respect to u and v, so the minimization splits into two separate one-dimensional problems. Denoting with $H(u)$ and $H(v)$ the Hamiltonian referred to only u and v respectively, in order to minimize the former, for example, one needs to make u as big or as small possible, depending on the sign of the switching function $\Phi_1(t)$ which corresponds to "quantity multiplying u". In particular, if $\Phi_1(t) = 0$ is not sustained over an interval of time, then the control is *bang-bang*

$$u^*(t) = \begin{cases} u_{MAX}, & \Phi_1(t) < 0 \\ 0, & \Phi_1(t) > 0 \end{cases} \tag{40}$$

If $\Phi_1(t) = 0$ over an interval of time, the value of u^* is *singular*; its value belongs in the interior of the control set and is determined by other information.

Before considering them, one recalls that the same reasoning is applied with respect to v - its switching function is $\Phi_2(t)$ - and that the times when the optimal controller swtiches from one boundary value to the other or to a singular control are called **switch times**. The minimum condition is equivalent to minimizing the switching function $\Phi_1(t)$ over the interval $[0; u_{MAX}]$. Since it vanishes on an open interval I, then also all its derivatives must vanish. It follows that one may repeatedly differentiate the switching function - $\frac{\partial H}{\partial u}$ - with respect to time until the control appears again; when it happens one solves for the control. In other words, during the interval when the switching function remains at zero, the control is determined by the requirement that the singularity condition continues to hold. The resulting solution is a *singular arc*, a portion of the optimal trajectory in which the Hamiltonian is not an explicit function of the control inputs, requiring higher-order necessary condition to be applied in the precess of the solution.

From a mathetematical point of view, one finds out that

$$\frac{d^k}{dt^k}(\Phi_1(t)) = 0, k = 0, .., r-1 \tag{41}$$

For some r, the input appears and one defines the appropriate control laws. In particular,

$$\Phi_1(t) = b + \langle \lambda(t), g_1(x_*(t)) \rangle \tag{42}$$

$$\dot{\Phi}_1(t) = \langle \lambda(t), [f, g_1]x(t) \rangle \tag{43}$$

$$\ddot{\Phi}_1(t) = -\langle \nabla L(x(t)) \cdot [f, g_1](x(t)) \rangle + \langle \lambda(t), [f + g_1 u + g_2 v, [f, g_1]](x(t)) \rangle \tag{44}$$

Moreover, a necessary condition for optimality of the singular control is the *Legendre-Clebsh condition*

$$\langle \lambda(t), [g_1[f, g_1]] x(t) \rangle \leq 0 \tag{45}$$

The singular control is said to be of *order 1* on Is if everywhere on the interval this quantity does not vanish. It follows that, if the singular control is order 1, then

$$\langle \lambda(t), [g_1[f, g_1]] x(t) \rangle < 0 \tag{46}$$

which is the *strengthened Legendre-Clebsh condition* which implies some *local* optimality properties of the singular control. In particular, in dimension 3 singular controls are locally optimal.

The last condition is equivalent to the *Kelley*'s one

$$(-1)^k \frac{\partial}{\partial u} [(\frac{d}{dt})^{2k} \Phi_1(t)] \geq 0, k = 0, 1, .. \tag{47}$$

The Legendre-Clebsh condition can be used in order to check wheter u and v are singular or not. The basic idea is to suppose that they are singular over an open interval Is and to verify that necessary condition for optimality. If it is violated, then the controller is not singular, if not, it is singular. As a result, u is singular and v not.

5.1 Numerical solution

The above mentioned problem, is quite cumbersome to solve analytically. Particularly, in the non bang-bang case, when u is singular, the value of u^* is of difficult computation. Thus, for ease of use, and didactic purposes, the present comparison will solve the optimal control problem using numerical methods.

Many commercial solvers are available that can solve a wide variety of similar optimization problems. Some of these, even require minimal mathematical knowledge, and can take as input problem statements that are quite close to natural mathematical language. However, the best ones are commercially sold programs. Moreover, many of the lesser ones are difficult to use and sometimes old, no longer mantained and with difficult to read documentations. In view of all this, in this work, custom MATLAB code, using the forwards-backwards method has been developed in order to solve the optimization problem.

5.2 Forwards-Backwards method

A simple idea to solve a generic nonlinear minimization problem, is start with a random guess, then move the guess towards the negative derivative of the function. This is what many of the simplest iterative learning algorithms work. A similar idea might be employed to solve the problem at hand too:

- Start with a random input

- Solve for the system's evolution in the given time frame
- Adjust the input in order to step towards the negative derivative of the Hamiltonian with respect to said input

This procedure, simple as it is, hides a major difficulty. In order to update the control, the derivative of the Hamiltonian with respect to the input has to be known. But, this derivative depends on the covector of parameters λ in the adjoint equations. Thus, knowledge of costates at any time instant is required. However, equations that model the behavior of costates have two problems

- They may only be applied to the optimal trajectory. That is to say, it is possible to compute the evolution of lambdas only if the input is the optimal one.
- Their known boundary condition is only available at the end of the simulating time T

Note how, the first limitation alone would not be a nuisance. If the boundary condition was provided at the start of the simulation, it would be trivial to solve the problem. It would simply be possible to compute the adjoint quantities at each time stem, then use them to update the input. Thus, by repeating this procedure at each time step, a solution would be found.
The same idea can be used for an actually implementable algorithm. This works as follows:

1. Guess a generic input vector
2. Provide initial conditions
3. Compute the evolution of the system from the initial conditions at time $t = t_0$
4. Compute the evolution of the adjoint equations from the condition at time $t = t_f = T$
5. Compute the derivative of the Hamiltonian with respect to the input at each time step
6. Update the input in order to minimize the Hamiltonian. That is to say, update the input in the direction that lowers the Hamiltonian.
7. Go back to step 2

The trick is designing an algorithm that will eventually converge to the optimal solution. Even if the adjoint solutions are wrong during the first iterations, they will be close to the truth around the initial time instant of the control interval. Which will make the control close to optimal at initial time. And this renders the adjoints closer to the actual value. Which makes the control even better. And repeating this loop for a reasonably long amount of time makes the algorithm converge to an actually optimal input with a real evolution of the adjoint equations.
This simplifies the symbolic computation part of the problem significantly. Only an expression of the system, the adjoint equations and of the derivative of the Hamiltonian are needed.
In the case when more than one input is available. The same algorithm can be duplicated to make each of the inputs minimize the respective derivative of the Hamiltonian. Thus, this

generalization, only requires the computation of 1 more derivative.
An interesting showing of this idea's effectiveness is computation time. When the input is far from the optimal one, iterations take longer and are widely different from one another. But, as the convergence gets closer, cycles get more and more similar and require less computational effort.

6 Estimation of a SIR model

For these models to be used to effectively control an epidemic, they must be tailored to the problem at hand. Pathogens have a variety of different ways of spreading and attacking populations. Therefore, parameters inside the SIR model presented in the previous sections have to be estimated. It is important to note that, even when knowledge of physiology of a certain pathogen is available in clinical experience, it is still very hard to model its behavior in an epidemic. In fact, viruses that have been known for decades, can mutate, change their effects and cause devastating epidemics. An example of this is the Ebola epidemic that spread in South-West Africa in 2014. This will be taken as an example in the control problems discussed later. Moreover, even for a well known virus for which we have previous epidemiological data, a new epidemic will always have new dynamics. In fact, different regions have populations of different kinds. Important factors in the spread of epidemics are:

- Genetic characteristics of the population.
- Population density
- Hygienic conditions and availability of healthcare.

Thus, the problem of estimating a well-suited model to a pathogen is of great importance. A good estimate has to be given while the epidemic is in action in order to better design control policy.
The problem of estimating a dynamical model can be split in two subsequent parts; first, a correct structure of the model has to chosen, then, the parameters in it have to be estimated.

6.1 Model choice

An important decision in system identification, is the choice of model parametrization. Based on the a priori information that is available, it's possible to differentiate between:

- White-box models, where all necessary information is available
- Black-box models of which there is no a priori information available
- Most real systems are grey-box models where limited information is known in advance

This choice, when it comes to epidemiological cases, is significantly simplified by the availability of many kinds of model structures that have been proven to be effective in predicting future evolution. Thus, most problems will be solved by a "white-box" approach.

The problem of choosing the right model among the many possible ones is however quite a cumbersome one. The basic structure is suggested by the characteristics of the disease. Even if the disease at hand clearly points to a specific structure of the model, many varieties of it are possible. For instance, the model varying SIR model presented in the previous sections, is not really common to use. In fact, despite being able to model in a more precise way by including population dynamics, it proves complicated to handle. In particular, the estimation process becomes increasingly complex, as the few parameters of the basic SIR explode in the eight described above. Moreover, the entire population dynamics is hard to describe in a real environment. The population refers to the sum of susceptible, infected and recovered people. In the real world, having even a basic estimate of susceptible population is not trivial. Using the entire population of a country is obviously overestimating. On the contrary, using the population of an area is misleading: the epidemic might spread outside of it with time. In other cases, only a small section of the population, geographically isolated may be really in danger. Thus, as will be shown during the simulations, the choice of a more complex model carries some costs. In many practical cases, a less comprehensive, but easier-to-handle model is a better choice.

6.2 Mathematical formulation of the problem

Once the model structure has been chosen, parameters have to be estimated. Assuming that all relevant variables are Gaussian random variables, then the optimal solution is given by the Ordinary Least Squares method. For this to be applied, the model has to be rewritten and parametrized in terms of some parameters in a structure like:

$$\boldsymbol{Y}_{\pi}(\boldsymbol{d}, \boldsymbol{u})\boldsymbol{\pi} = \boldsymbol{d} \tag{48}$$

where

- $\boldsymbol{d}(t)$ is a data vector containing

$$\boldsymbol{d}(t) = \begin{bmatrix} \dot{\boldsymbol{S}}(t) \\ \dot{\boldsymbol{I}}(t) \\ \dot{\boldsymbol{R}}(t) \end{bmatrix} \tag{49}$$

- $\boldsymbol{\pi}$ is a vector containing the parameters
- $\boldsymbol{u}$ is a vector containing the input vectors
- $\boldsymbol{Y}_{\pi}$ is the regressor matrix.

For the SIR model described in [33], one possible choiche of parameters vector is

$$\boldsymbol{\pi} = \begin{bmatrix} \gamma & \nu & \beta & \rho & \kappa & \eta & \alpha & \mu \end{bmatrix}^T \tag{50}$$

Note that many other choices were possible by regrouping parameters. For instance, in the Infected dynamics, the term $(\mu + \nu + \alpha)$ can be grouped in a single parameter.

With the above-mentioned parameters vector, the regressor is defined as:

$$\boldsymbol{Y}_{\pi} = \begin{bmatrix} S+R+I & -S & -\frac{SI}{S+I+R} & R & -Su & 0 & 0 & 0 \\ 0 & -I & \frac{SI}{S+I+R} & 0 & 0 & -Iv & -I & I \\ 0 & -R & 0 & -R & Su & Iv & I & 0 \end{bmatrix} \tag{51}$$

The data vector contains the derivatives of the state variables. These are not immediately available, they must be estimated numerically. In practice, this may be done by

$$\boldsymbol{d}_i(t) = \begin{bmatrix} \frac{S_{i+1}-S_{i-1}}{2dT} & \frac{I_{i+1}-I_{i-1}}{2dT} & \frac{R_{i+1}-R_{i-1}}{2dT} \end{bmatrix}^T \tag{52}$$

By stacking all the data at the available times, it is possible to build a tall regressor and data vectors with $3N$ rows (where N is the number of data points). Once the complete matrix has been built, the optimal estimation of the parameters vector is given by

$$\hat{\boldsymbol{\pi}} = \boldsymbol{Y}_{\pi}^{\#}\boldsymbol{d} \tag{53}$$

where $\boldsymbol{A}^{\#}$ denotes the pseudoinverse of matrix A as defined by:

$$\boldsymbol{A}^{\#} := (\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}^T \tag{54}$$

For this method to be useful, the regressor has to carry as much information as possible. An useful measure of how much "independent" information the rows of the regressor are able to carry, is the conditioning number. This, defined as

$$cond(A) = ||A|| \cdot ||A||^{\#} \tag{55}$$

In practice, this is usually computed using the 2-norm for matrices. Thus, for the identification problem to be correctly solved, the conditioning number of the matrix should be finite, not getting too high.

6.3 Problems in Identification processes

Despite the analytical knowledge of an optimal solution that is easy to compute, and the availability of good and tested model structures, the identification process is still quite difficult in practice. As already mentioned above, the choice of the model can be cumbersome, and some models may be unusable in practice because of the difficulty in estimating their parameters.

Real worlds situations carry more difficulties. For instance, the SIR model with varying populations, has 3 different variables. Which means, that to estimate them, 3 independent measured variables would be needed. But, as stated above, susceptible individuals are practically impossible to measure. Knowledge of the total population is also impossible to determine, being this linked to the disease-ridden area, and not a total population of any easily distinguishable region. Moreover, available data is not really easy to process. In fact, often the number of deceased individuals is provided, as they are, sadly, easy to measure. But estimating the number of infected is not trivial. In addition to this, from the number of deceased, it should be possible to determine the recovered, but this process is far from trivial.

This may be done by adding a delay: after some time, infected become either deceased or recovered. But, this neglects the fact that deseases have sometimes widely varying duration of infection. Thus, an individual may recover after weeks, while another can die withing days.
In the case of disease with quick duration of infection, a possible solution is to simply neglect that duration and estimate the recovered to be equal to the infected minus the dead at time t. For instance, the Ebola epidemic in West Africa in 2014 had medium disease length of eight days. Therefore, over the span of the 2 years, it is not too unreasonable to neglect it.
Another obvious additional problem, a quite common one in identification theory, is that data is noisy. In particular, during an epidemic, in a confused environment and with a shortage of resources, it is impossible for medics to accurately guess the amount of infected. Thus, some data processing is usually required before the estimation process takes place. Most of the time, this can be simply achieved by using a low-pass filter that removes most of the high-frequency noise.
Additionally, data is no only noisy, but also depending on a wide variety of macroscopic (and low frequency) external factors. For instance, it's, again, impossible to correctly gauge the control effort applied in any reasonable way. Thus, models have to be estimated assuming no control effort, or by guessing one via heuristics.

7 Simulations

In this section, simulations of various scenarios of epidemics will be presented. All the optimal control problems will be solved by the forward-backwards method, all of the estimation problems will be solved by the Ordinary Least Squares method.
The first few presented simulations will use the SIR model presented in the previous sections. Then, that model will be applied to real world data. However, the population varying SIR will provide disappointing results. Thus, a simplified version of the model will be constructed by using additional assumptions. This will be used to study, in an optimal control framework, the Ebola epidemic of 2014 in Guinea.

7.1 Code structure

On the basis of the two main models used, the code has been organized into two folders: one for the complete theoretical model, the other for the simplified one. The first contains:

- A "control.m" file where the forwards backwards method is implemented to control the epidemic from a given initial state.
- an "identification.m" file where an attempt to identify the parameters in the system from fictional data was made.
- a "symbolic_math.m" where the symbolic mathematical expressions used in the control have been generated

The second folder contains equivalent "control_simplified.m", "symbolic_math_simplified.m" as well as:

- A "data_visualization.m" script that extracts the data of the real worlds epidemic and displays it in a plot
- An "identification_simplified.m" script that uses imported data to identify the simplified model.
- An "ebola_data.xlsx" excel sheet with the epidemiological data. This script has been modified to have an additional "dates" section in order to extract more easily the time stamp of the data. In all simulations, the dates will be presented as "number of days since the start of the epidemic".
- An "import_epidemics_data.m" function that extracts the data from the excel file and reorders it in more easily manageable vectors.

7.2 Population varying SIR model

Before addressing the control problem, it is essential to discuss the validity of the model in providing reasonable results. Model parameters have been set to the values in the table. These values were taken from the simulation described in [33]. It is important to note how these are specifically chosen for a starting population of 1000 susceptible, 10 infected and no recovered. Because of numerical errors, parameters have to be tuned to the population in the study. Moreover, looking at α, we may notice it to be extremely small, much more than any reasonable "natural healing rate" for any disease. This has been chosen to be so small in order to stress the importance and effect of the control in the simulations. Thus, a really bleak scenario will present itself in control-less simulations.

PARAMETERS	**VALUE**
γ	0.00683
ν	0.00288
β	0.2426
μ	0.005
α	0.00002
ρ	0.007
κ	0.3
η	0.1

7.2.1 Free evolution

As stated above, the free evolution with these parameters makes the disease infect a large portion of the population. A natural recovery factor of 0.00002 makes it so, for a population of 1000 infected, a single infected is naturally healed every 50 days, which is completely unrealistic. Thus, during the whole simulation time, I remains really low (around 1 individual). It can be noted that, while the total population grows a little at first, it remains quite stable in the end.
In the last section of the simulation, the variation in infected and susceptible gets smaller and smaller. This is because, the dynamics of these two variables slows down significantly. In fact, three quantities become quite close (Figure 3)

- Newborns added to the population each day via γN
- The rate of new infected each day via $\beta \frac{SI}{N}$

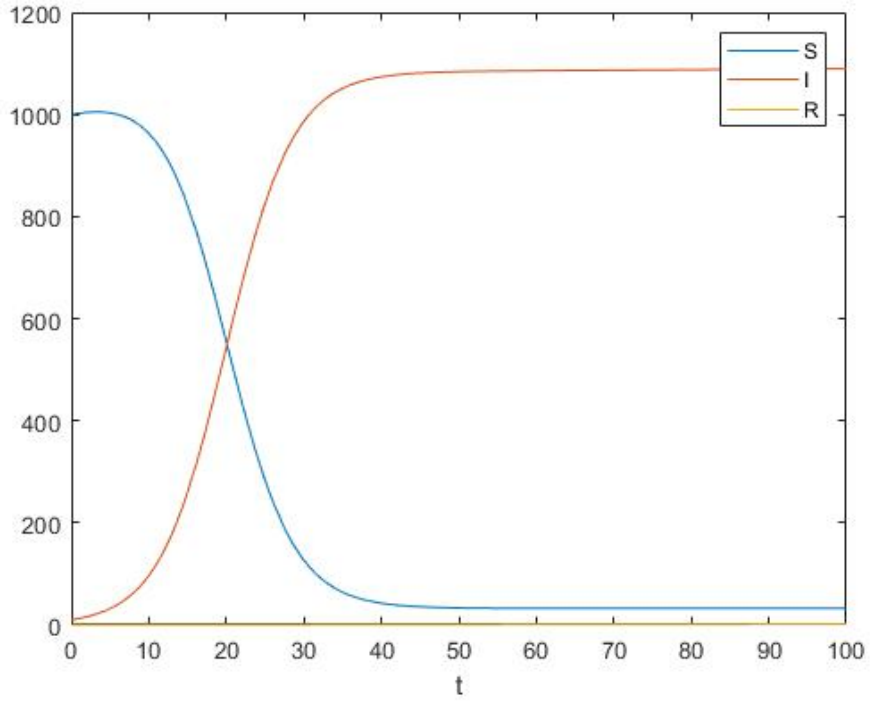


Figure 2: Free evolution of the SIR model

- The rate of death of the infected via $I(\mu + \nu)$. Note how α is another flow out of the infected but is being neglected as it is really small in this scenario.

This makes it so, at any given time interval, there is very little flow between the three state variables, as any input flow is balanced by an external one in both S and I. The dynamics of the R is, again, neglectable because of the choice of α.
Note how, this simulation, as well as all following ones for this model, are run on a 100 weeks time frame.

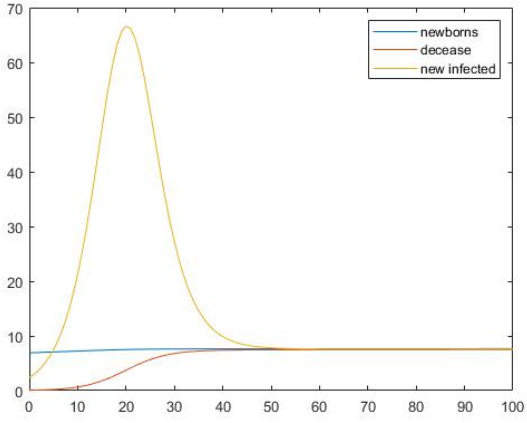


Figure 1: Transfer dynamics between S and I

Another simulation was run with a more "reasonable" value of α. Considering a generic disease with a time of illness such that at every time step, on average, 6.7% of the population would recover. With this new data, a radically different evolution of the model can be observed (Figure 3). Infected start low, then quickly increase until a plateau. After that, the natural death rate, and the ample amount of recovered, make the epidemic slowly recedes. This qualitative analysis can be observed in the SIR model for a wide range of parameters. Actually, the epidemic defined by this set of parameters is a mild

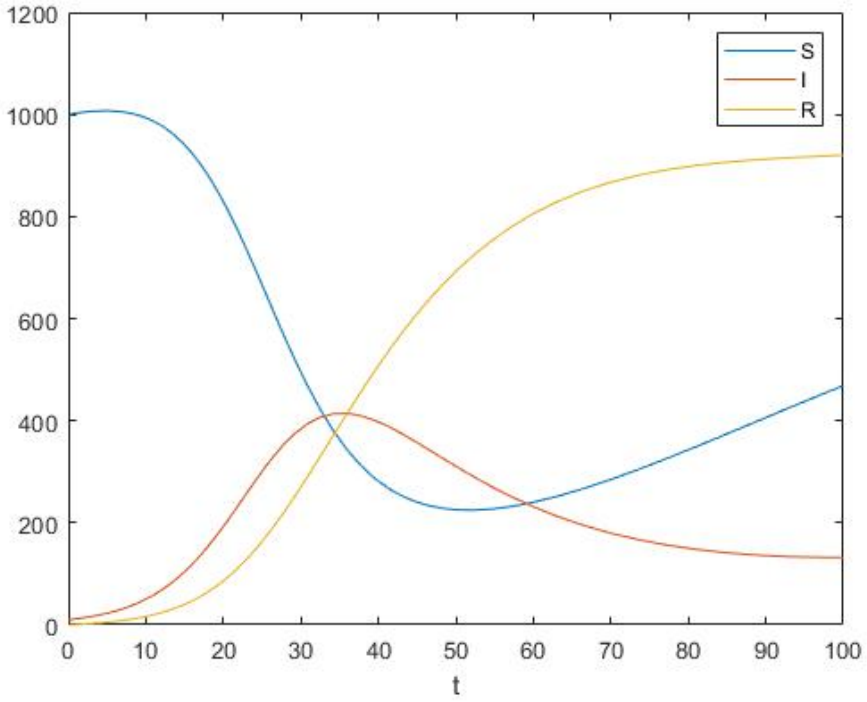


Figure 3: evolution of the system with lowered natural recovery rate

one, not many are infected at any point in time; the transmission factor β is not really high. Looking at the other variables, susceptible individuals get lower and lower at the start of the simulation. Then, when the epidemic is receding, they grow in number from the big amount of newborns, coupled with the effect of the recovered that lose their protection. This effect is particularly strong as recovered are always increasing.
With confidence on the realistic behavior of the model in free evolution, control will now be discussed.

7.2.2 Control

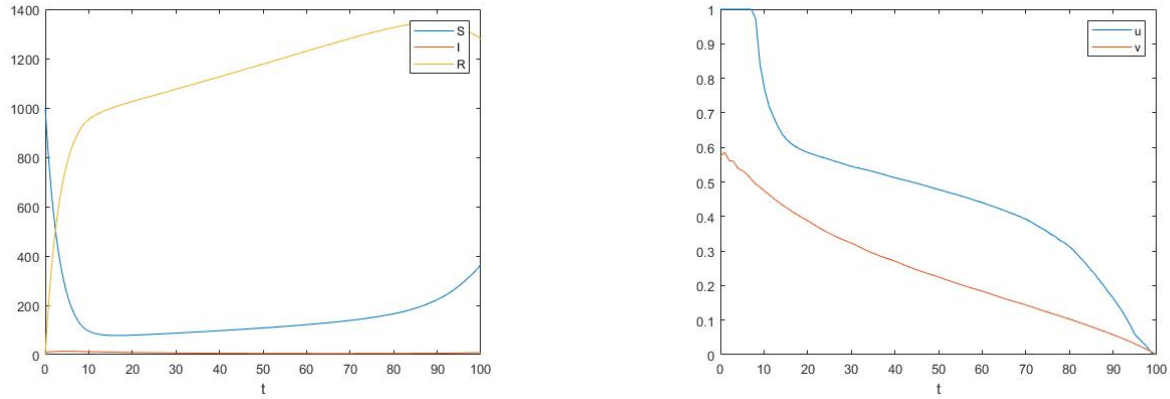
In the control section, to emphasize the effect of the control, and only of the control, α has been lowered once again. In order to pose the problem in a complete way, three more parameters are needed. These have been set (again, following [33]), as

$$a_{cost} = 5 \qquad b_{cost} = 50 \qquad c_{cost} = 300$$

which are the parameters of the cost function. Moreover, in this specific instance, the cost functional has been taken as

$$J = J(u, v) = \int_0^T (aI(t)^2 + bu(t)^2 + cv(t)^2)dt \tag{56}$$

The evolution under the use of feedback is qualitatively completely different. Control starts high in order to limit as much as possible the spread of the disease. Then it slowly tapers off



(a) Evolution with optimal control

(b) Optimal vaccines and healthcare controls

Figure 4: Optimal control evolution

when the number of infected is low enough to justify it. Thus, most individuals are rendered "recovered" by the end of the simulation, while the number of infected is kept low. The population is also allowed to grow as few people decease from the disease.

7.2.3 Identification

Before testing the properties of this model on real world data, a fictional identification procedure has been carried out. Data taken from the previously described simulations has been used for the process. A regressor matrix has been computed as described in Equation 51 as well as a data vector (Equation 6.2). A optimal solution (in the least squares of the error sense) has been found by computing the pseudoinverse of the regressor.
First, this procedure was done for the optimally controlled simulation. Results from estimating all the parameters were, at a first glance look quite good. The estimation algorithm was fed the first 35 datapoints of evolution in the optimal control case. As it can be seen from Figure 5a, the identified model follows the "real" model quite well for most of the simulation. However, this "good" result hides several significant problems.
The first alarm bell rings when computing the conditioning number of the regressor. In fact, in the portrayed simulation, the conditioning number is $1.4367e + 18$ which is definitely too high.

Because of this, when adding even a small amount of noise to the data, the solution provides terrible results. The noise added has been of a random value between 3 and -3 to the susceptible. Considering that the total population is of 1010 individuals, this noise is quite small. In fact, it's too small to be reasonable when simulating the real world. Despite this, the estimated model performs extremely poorly. The susceptible reach $6 \cdot 10^6$ by the end of the simulation.

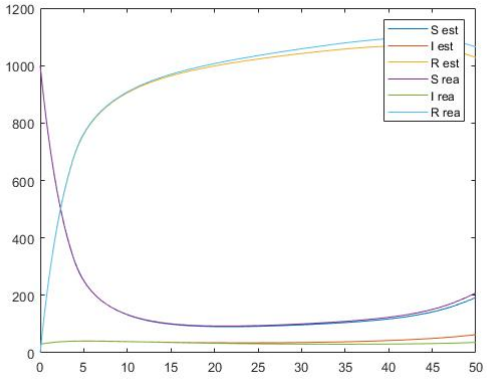
Another important deficit of this estimation, is its feasibility. Even ignoring the lack of robustness with respect to noise, the estimated parameters have been unrealistic. And that is, not only because they were orders of magnitude away from the "true" (assigned) ones, but

because they were in some cases negative. For instance, because of the value of η treatment has the ability of bringing people from the recovered to the infected in the estimated model.

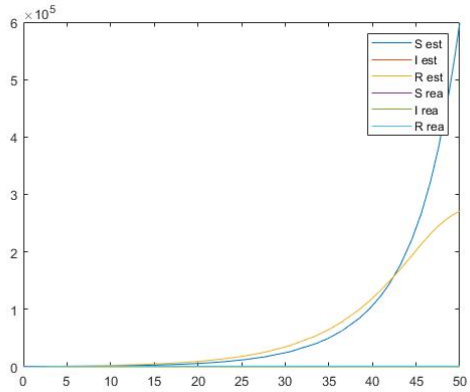
An important consideration about these results should be made. These have tried to estimate a model with an optimal control applied. The evolution of the model, in the optimal case, is far from the free evolution of a similar model. In fact, some of the dynamics, such as the recovered, are severely gimped by the controller. Thus, trying to estimate the parameters in this conditions, may be an error.

ESTIMATED PARAMETERS	VALUE
γ	0.2318
ν	0.0968
β	-1.3184
μ	0.1095
α	0.1095
ρ	-0.1229
κ	0.3301
η	-1.7872

Thus, a new estimation procedure has been tried. This time, no attempt was made at estimating the control effectiveness parameters: the control was set to zero. Moreover, in order to have a complete look at evolutions of the dynamical quantities, the higher value $\alpha = 0.067$ has been chosen. However, this did not solve the estimation problems that the model faced. The conditioning number still remained quite high at $cond(\boldsymbol{A}) = 1.2033e + 04$. Moreover, parameter estimation were, once again, unfeasible.
In light of all this, no attempt has been made at estimating model parameters from real world data. To actually have an useful controller that can be used in a real endemic scenario, a new epidemic model has been designed.



(a) Free evolution of the identified model, compared to the "real" one



(b) Evolution of the estimated model when noise is added to the input

7.3 Simplified model

The final objective of these simulations, was to provide a working optimal controller that could be used in the real world to perfectly handle epidemics. The previously discussed model can't fulfill this objective. In fact, as discussed in Section 6.1, using a complex model

often has important drawbacks. In this particular case, the population varying SIR with 8 parameters proved to be hard to estimate. Because of this, a new, simpler model has to be synthesized. Many of the possible SIR model variations provide ways of simulating evolutions similar to the ones of the one presented above. Thus, another one was defined from the following hypothesis:

- The population was kept constant and the birth rate of the population γ was neglected. Additionally, natural and disease induced death have been ignored ($\mu = \nu = 0$) These assumptions are reasonable in the case of large populations affected by epidemics that do not impact the total number of individuals too much.
- It was considered to be impossible for recovered to become infected again ($\rho = 0$)
- The disease transmission parameter β has been rewritten to include the (now fixed) total population $a = \frac{\beta}{N}$
- The name of the natural healing rate has been changed into $b = \alpha$

Thus, a new epidemic model has been obtained:

$$\begin{cases} \dot{I} = -aSI \\ \dot{I} = aSi - bI - \mu v \\ \dot{R} = bI + \mu v \end{cases} \tag{57}$$

Note how only one input has been considered. This was done in order to better approximate real world conditions of the problem that will be tackled in the next section.

7.4 West Africa Ebola epidemic

The following part of this work, will revolve around simulating and controlling a real world epidemic. In particular, the epidemic of Ebola in West Africa in 2014 has been chosen as a study case. The Western African Ebola virus epidemic (2013–2016) was the most widespread outbreak of Ebola virus disease (EVD) in history—causing major loss of life and socioeconomic disruption in the region, mainly in the countries of Guinea, Liberia, and Sierra Leone. The first cases were recorded in Guinea in December 2013; later, the disease spread to neighboring Liberia and Sierra Leone, with minor outbreaks occurring elsewhere. Small outbreaks occurred in Nigeria and Mali, and isolated cases were recorded in Senegal, the United Kingdom and Italy. In addition, imported cases led to secondary infection of medical workers in the United States and Spain but did not spread further. The number of cases peaked in October 2014 and then began to decline gradually, following the commitment of substantial international resources. As of 8 May 2016, the World Health Organization (WHO) and respective governments reported a total of 28,646 suspected cases and 11,323 deaths (39.5%), though the WHO believes that this substantially understates the magnitude of the outbreak.

The outbreak left about 17,000 survivors of the disease, many of whom report post-recovery symptoms termed post-Ebola syndrome, often severe enough to require medical care for months or even years. An additional cause for concern is the apparent ability of

the virus to "hide" in a recovered survivor's body for an extended period of time and then become active months or years later, either in the same individual or in a sexual partner. In December 2016, the WHO announced that a two-year trial of the rVSV-ZEBOV vaccine appeared to offer protection from the variant of EBOV responsible for the Western Africa outbreak. The vaccine has not yet been given regulatory approval, but it is considered to be effective and is the only prophylactic which offers protection; hence, 300,000 doses have been stockpiled. However, here is currently no antiviral drug licensed by the U.S. Food and Drug Administration (FDA) to treat Ebola in people. Multiple antiviral drugs are being developed and tested. Blood transfusions from survivors and mechanical filtering of blood from patients are also being explored as possible treatments for EVD.

The choice of Ebola was not a random one. Firstly, the disease respects quite closely the assumptions made in the previous sections. To be completely rigorous, it's a bit more aggressive disease than a "constant population model" would allow. However, considering population movement, size of the affected population and size of the epidemic, the assumption is not completely unreasonable. Secondly, the epidemic was quite recent. Thus, it's an actually important real world problem on which plenty of lives depend. Finally, being this a modern epidemic, data is available online with little trouble.

The absence of a second control input in the model defined above, is now clear. At the time of the epidemic, no Ebola vaccine was available. Thus, it's more realistic to have a control only depending on clinical help. Symptoms of Ebola virus disease (Ebola) are treated as they appear. When used early, basic supportive care can significantly improve the chances of survival. These include:

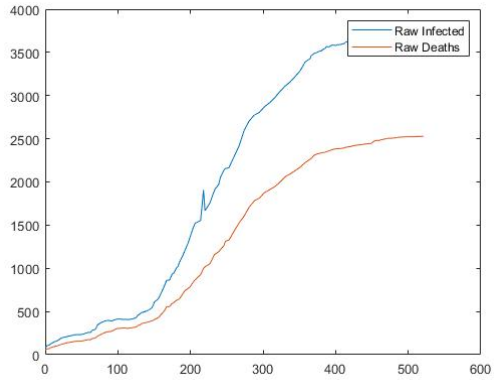
- Providing fluids and electrolytes (body salts) through infusion into the vein (intravenously)
- Offering oxygen therapy to maintain oxygen status.
- Using medication to support blood pressure, reduce vomiting and diarrhea and to manage fever and pain.
- Treating other infections, if they occur.

Those who recover from Ebola develop antibodies that can last 10 years, possibly longer. It is not known if people who recover are immune for life or if they can later become infected with a different species of Ebola virus. Some survivors may have long-term complications, such as joint and vision problems. Thus, in the model, it is a reasonable assumption to consider that no recovered can become susceptible again ($\rho = 0$).

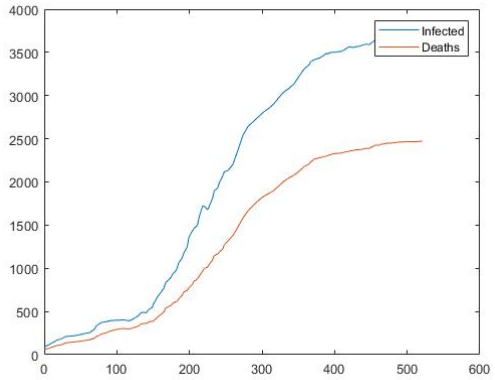
Among the different countries affected by the epidemic, attention will be focused on Guinea. All data described in the following chapters will be referred to that.

7.4.1 Data retrieval

Data concerning Ebola's spread has been made available by the "Centers for Disease Control and Prevention" (CDC). The CDC is a major operating components of the Department of Health and Human Services in the United States. At the time of writing, the data is accessible at the CDC website ("cdc.gov") [35] in a general section dedicated to the Ebola



(a) Raw epidemic data from the WHO



(b) Data after filtering

virus. As stated on the website, the actual data has been reported by the "Whorld Health Organization" (WHO). In particular, it is possible to download an excel sheet containing all case files.

In order to make the data easier to use in MATLAB, the excel sheet has been modified by adding another column of data. This column simply computes dates using days as unit of measurements. This is done by taking the reported data, and subtracting it by the first data of the epidemic.

Using the "import_epidemics_data.m" functions, two vector are created. One contains the reported number of infected, the other the number of deceased. The epidemic in Guinea had its biggest spread for around the first 180 datapoints (from the start of the epidemic up to the end of August 2015). Thus, only the corresponding samples have been considered. The infected data contains a big discontinuous spike at around 210 days from the start of the epidemic. To smoothen it out, both vectors were subjected to a low-pass filter of 0.5 normalized frequency. Higher frequency could not be used as the final part of the dataset is almost constant. Thus, an higher frequency filter would have added inexistent spikes. Therefore, the spike, and other noise discontinuities, were not completely remove. A comparison of the data, before and after filter can be observed in Figure 6a. The visual difference is minimal, but still, the reults on the filtered data have been proven to be satisfactory, and marginally better than the ones on the raw data.

7.4.2 Estimation procedure

The estimation procedure has been carried out as described in Section 6.2. In these context, three really important hypothesis have to be discussed.

First, an estimate of the total amount of population has to be provided from outside. The total population of Guinea at the time of the epidemic was around $7.9M$ individuals. However, the epidemic started, and spread in a demographically complex area in the South of the country. In particular, first reported case was observed in Meliandou, a small village close to the border. Thus, the number of people living in the area was a lot smaller than the entire population of Guinea. Moreover, the epidemic quickly spread across the border moving between the different countries. Estimating the population flow between the countries and

Figure 7: Map of West Africa highlighting the first Ebola case

the population count of the first region are practically complex problems. In this study, the subjected population was assumed to be around 5500 individuals. This is coherent with the number of infected and decesead, as well as reasonable when thinking about a small rural African region.

Secondly, an assumption is needed when inferring actual R and S from the available data. The particular strain of Ebola of the considered epidemic, has a time of disease between 8 and 14 days. Thus, the choice was made to neglect this time frame and instantly link recovered, infected and deceased as

$$R(t) = I(t) - D(t) \tag{58}$$

where $D(t)$ is the number of deceased at time t. This assumptions is clearly not reflective of the actual dynamics of the real world system. There exists a delay between the time in which a patient is infected, and when it recovers or dies. However, over a long time frame, this effect becomes quite small. Moreover, this is an actually conservative hypothesis: it assumes that more people suffer than the real case. Therefore, by controlling a system where this is the case, the same control would fare even better in the real scenario.

Finally, this model was estimated with no assumed control. It is true that, at the very beginning of the epidemic, due to the rural location of the outbreak no modern medical assistance was given to the infected. However, with time, more and more resources were dedicated to fighting the disease. Thus, estimating this model with an assumedly 0 input is not theorically correct. However, two considerations make it more reasonable. First, the epidemic was still developing in a really rural and difficult to reach area. Thus, even with significant economical expenditure, due to logistical constraints, the help provided was far from an optimal and effective one. Secondly, at the time, the virus was a recent mutation of a previous strain. Thus, no knowledge on how to treat it was available. Which made healthcare efforts particularly ineffective. With all of these in mind, the control input can be considered negligible. Or, in a different light, it may be assumed that the optimal controller described in the following section, has to applied on top of the control that was provided

during the epidemic. Nonetheless, the evolution of the estimated model will still be referred to as "free evolution".

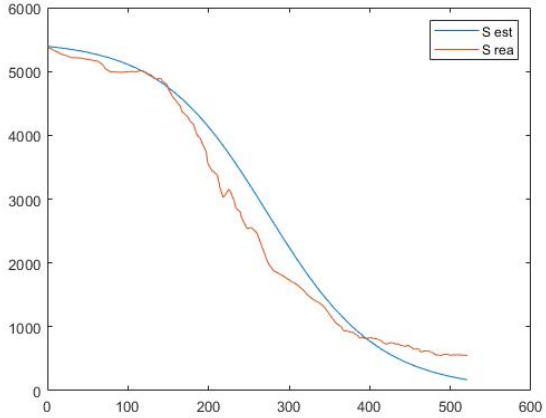
Performing the estimation procedure, a custom regressor has to be computed as

$$\boldsymbol{Y}_{\pi} = \begin{bmatrix} -SI & 0 \\ SI & -I \\ 0 & I \end{bmatrix} \tag{59}$$

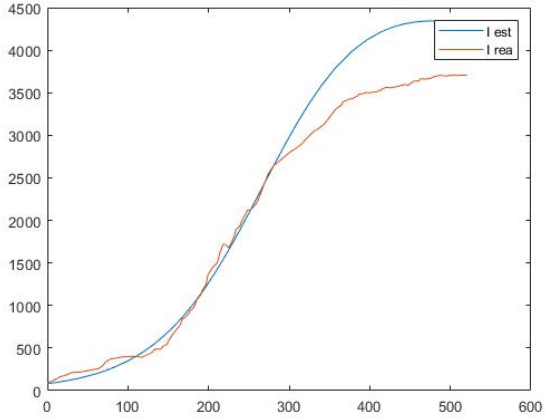
whose conditioning number is 1200 which, considering the order of magnitude of the variables inside, is completely reasonable and far from diverging at infinity. When this, as well as the data vector are computed, the Ordinary Least Squares solution can be computed obtaining the estimate:

$$a = 2.912 \cdot 10^{-6} \qquad b = 0.0008$$

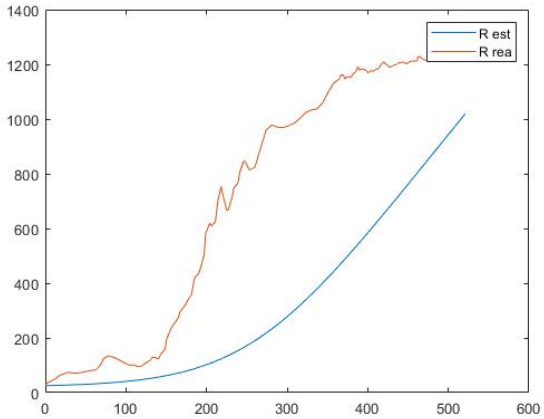
Here, both of the parameters are positive (feasible) and reasonable for an estimated total population of more than 5000 individuals that are all in contact.



(a) Comparison of S



(b) Comparison of I



(c) Comparison of R

Figure 8: Comparison between model prediction and real world data.

Figure 8 reports the difference between model prediction and actual data. Actual data was modified using the hypothesis in Equation 58. It is noticeable how the model does a good job of estimating the evolution of infected and susceptible. The only dynamic that looks far from the actual one is the recovered one. However, the difference, despite being graphically big in the plot, is of a few hundred individuals. Which in a population of several thousands is not too discouraging. This difference is probably the result of externalities that are not included in the current simple model. Moreover, estimating the recovered to be less than they actually are is a conservative estimation. Which makes the model still useful in the sense that it will provide a solution that errs on the side of caution.

Another noticeable mistake, is the higher estimation of the infected in the last part of the simulation. This, again is a conservative estimate that can be tolerated.

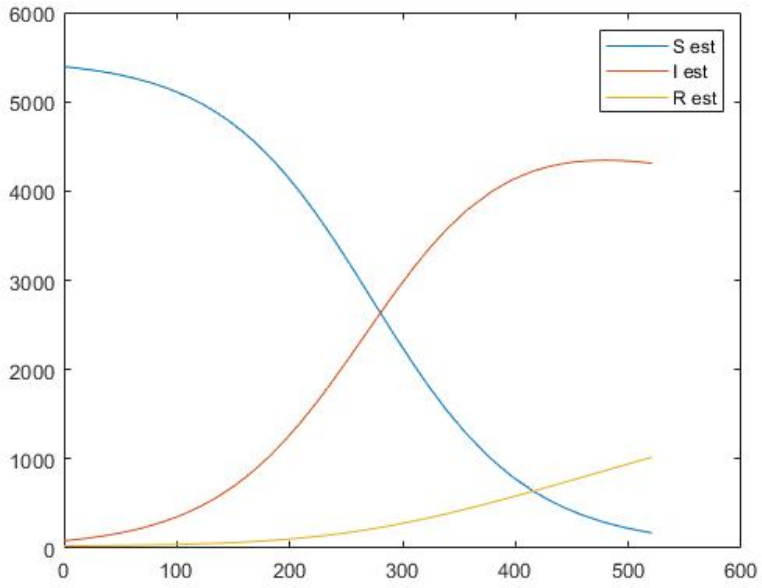


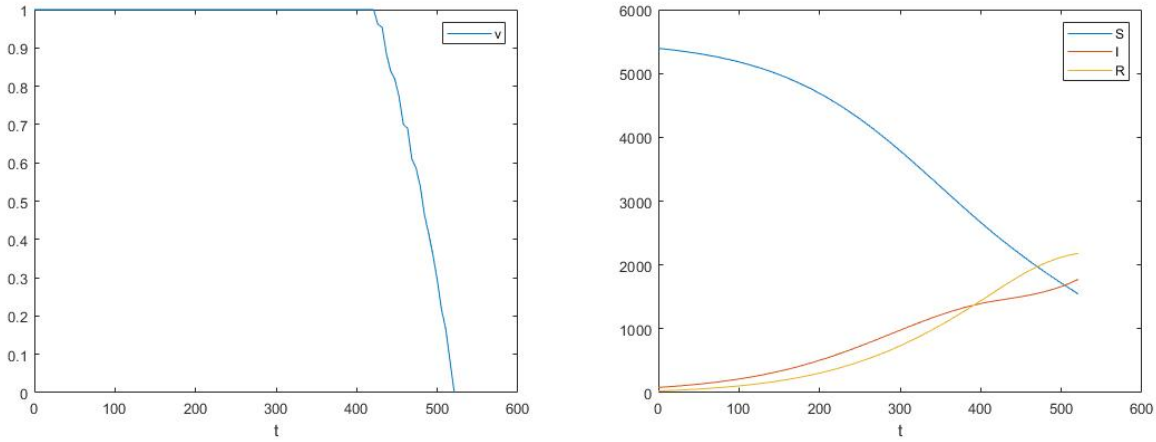
Figure 9: Free evolution of the estimated model

The model free evolution can be conveniently viewed in Figure 9. Just like in the previous simulations, the same qualitative behavior can be observed. It is particularly noticeable in this case, how the natural recovery rate is quite small. During the whole simulation a big part of the population is infected.

7.5 Control of the simplified model

An optimal control problem has been imagined in which healthcare has to be provided in a timely manner during the West Africa epidemic of 2014 in order to minimize a cost functional (with the obvious removal of the vaccination cost). To be more realistic, this controller will not use the possibility of vaccinating the population, as during the epidemic, no Ebola vaccine existed. The starting condition will be taken as the first available data point of 25/03/2014. The simulation will run for the same time as real world data that was

considered (521 days). The same weights will be employed for the cost functional as in the previous simulations. The same forward-backwards algorithm used to control the population varying SIR model was employed in the real world scenario.



(a) Optimal input

(b) Evolution when controlled in the optimal way

Figure 10: Simulation of the optimal solution to the real-worlds estimated model

An important parameter in determining the solution to this problem has not been defined. That is the effectiveness of the control input. In particular, the model described in the previous control simulation, had a nonspecific time frame as an interval. For example, consider the healthcare effectiveness parameter in the context of the population varying SIR model. An effectiveness parameter of $\mu = 0.1$ means that with a control effort of 1, it is possible to heal 10% of the entire infected population at each time stem. This is reasonable when the time interval in not clearly specified: the time interval in the simulation might as well be several months long. However, in the current formulation of the problem, the time interval considered is only one day. And it is obviously far-fetched to think that healthcare could heal 10% of the infected every day. This gets even worse when considering that inside the considered model, some control was already being applied. Thus, a significant reduction of the parameter is necessary to provide reasonable results. In fact, simulating the system with such an high μ results in the epidemic being completely wiped out in a few weeks. Additionally, over a large population, in a rural area, the cost of providing care is orders of magnitude higher than normal. Thus, b_{cost} has also been severely increased increased.

With all this in mind, the following assumptions have been made. With maximal control effort, 0.5% of the infected population can be recovered. This corresponds to 3.5% each week or 15% each month. While maybe slightly optimistic, this estimation shouldn't be too unreasonable. But, to obtain this, in a large population and a difficult environment, the cost index has been increased to $b_{cost} = 5000000$. No variation in a_{cost} has been deemed necessary: a larger population will naturally apply a bigger cost because of the higher number of possible infected.

Results of the simulation can be seen in Figure 10. The control follows a quasi bang-bang pattern. For almost the whole simulation, the control is the upper limit, while at the end it slowly tapers off to 0. The variation in evolution of the system is remarkable. Around 2000

less people are infected, while the number of recovered increases by around 1000 compared to the free evolution.

References

[1] World Health Organization, "Epidemiology", https://www.who.int/topics/epidemiology/en/, 2010.

[2] Field Epidemiology Manual, "Key definitions in infectious diseases epidemiology", https://wiki.ecdc.europa.eu/fem/w/wiki/1201.key-definitions-in-infectious-diseases-epidemiology, 2014.

[3] thebmj, "Epidemiology for the uninitiated", https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-what-epidemiology, 2019.

[4] Oxford Dictionary of English, "Epidemiology", Oxford University Press, 2010.

[5] "Principles of Epidemiology." Key Concepts in Public Health. London: Sage UK, 2009.

[6] Miquel Porta, A Dictionary of Epidemiology (6th ed.). New York: Oxford University Press. ISBN 978-0-19-997673-7.

[7] Centers for Disease Control and Prevention (CDC), "Principles of Epidemiology in Public Health Practice, Third Edition. An introduction to Applied Epidemiology and Biostatistics", Section 7: Analytic Epidemiology, https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section7.html, 2012.

[8] Necdet Sut, "Study Designs in Medicine", Balkan Medical Journal, pp 273-277, 2014.

[9] PennState Eberly College of Science, "Epidemiological Research Methods", STAT 507, 2018.

[10] Rivier University, "What Is the Epidemiologic Triangle?", 2017.

[11] Daley, D. J., Gani, J. (2005). Epidemic Modeling: An Introduction. NY: Cambridge University Press.

[12] Brauer, F.,Castillo-Chávez, C. (2001). Mathematical Models in Population Biology and Epidemiology. NY: Springer.

[13] Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 115 (772): 700.

[14] Yvette Brazier, "All you need to know about flu", MedicalNewsToday, 2017.

[15] Jill Seladi-Schulman, "How Long Do Flu Symptoms Last, and How Long Are You Contagious?", healthline, 2018.

[16] Teri Johnson, "Mathematical Modeling of Disease: Susceptible-Infected-Recovered (SIR) Model", 2009.

[17] Howard (Howie) Weiss, "The SIR model and the Foundations of Public Health", Universitat Autonoma de Barcelona, 17 pp., 2013.

[18] Tuen Wai Ng, "The mathematics of diseases", On modeling Hong Kong's SARS Outbreak, HKU.

[19] European Patients' Academy, "Epidemiologic concepts: incidents and prevalence", https://www.eupati.eu/pharmacoepidemiology/epidemiologic-concepts-incidence-and-prevalence/Prevalence, 2017.

[20] Quinn Morris, "Analysis of a Co-Epidemic Model", Wake Forest University, 2008.

[21] Isabeau Birindelli, Teorema di Poincarè Bendixson, Appunti delle lezioni, 2012.

[22] Peter Roth, "Arithmetica Philosophica", 1608.

[23] Albert Girard, "L'invention nouvelle en l'Algebre", 1629.

[24] Suzanne Lenhart, "Bang-bang and Singular Controls", University of Tennessee, Knoxville.

[25] Zelikin, M. I.; Borisov, V. F. (2005). "Singular Optimal Regimes in Problems of Mathematical Economics". Journal of Mathematical Sciences. 130 (1): 4409–4570 [Theorem 11.1]. doi:10.1007/s10958-005-0350-5.

[26] Terry A. Straeter, "Singular Extremaldoids in Optimal Control Theory and The Calculus of Variations", National Aeronautics and Space Administration, 1970.

[27] MIT OpenCourseWare, "Principles of Optimal Control", http://ocw.mit.edu , 2008.

[28] Nature, "Disease prevention", nature.com.

[29] Urszula Ledzewicz, Heinz Schattler, "Singular controls and chattering acrs in optimal control problems arising in biomedicine", Control and Cybernetics, vol. 38, No. 4B, 2009.

[30] H. Gardner Moyer, "Sufficient conditions for a strong minimum in singular control problems", SIAM J. Control, Vol. 11, No. 4, 1973.

[31] Helena Sofia, Ferreira Rodrigues "Optimal Control and Numerical Optimization Applied to Epidemiological Models", Ph.D. thesis submitted to the University of Aveiro

[32] Helmut Maurer, Maria Do Rosario De Pinho. Optimal Control of Epidemiological SEIR models with L1 Objectives and Control State Constraints.

[33] Urszula Ledzewicz, Mahya Aghaee, Heinz Schattler, "Optimal Control for a SIR Epidemiological Model with Time varying Populations", 2016 IEEE Conference on Control Applications (CCA)

[34] Centers for Disease Control and Prevention, organization https://www.cdc.gov/about/organization/cio.htm

[35] Centers for Disease Control and Prevention, Ebola data https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html

[36] WHO Ebola report https://www.who.int/csr/disease/ebola/en/