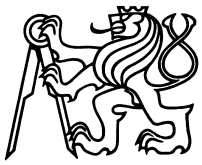
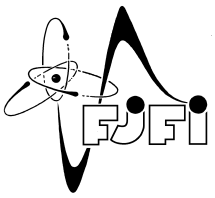


CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Nuclear Sciences and Physical Engineering



Mathematical modeling in epidemiology

Matematické modelování v epidemiologii

Bachelor's Degree Project

Author: **Daria Petrova**

Supervisor: **Ing. Miroslav Kolář, Ph.D.**

Academic year: 2023/2024

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Petrova** Jméno: **Daria** Osobní číslo: **483745**

Fakulta/ústav: **Fakulta jaderná a fyzikálně inženýrská**

Zadávající katedra/ústav: **Katedra matematiky**

Studijní program: **Aplikace přírodních věd**

Studijní obor: **Matematické inženýrství**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Matematické modelování v epidemiologii

Název bakalářské práce anglicky:

Mathematical modeling in epidemiology

Pokyny pro vypracování:

1. Prostudujte vybranou základní literaturu o matematickém modelování v oblasti epidemiologie.
2. Seznamte se s modelem SIR (Susceptible-Infected-Recovered) pro šíření infekčních onemocnění v populaci. Na základě současných výsledků zjistěte možnosti jeho uplatnění.
3. Diskutujte moderní trendy v modelování šíření infekčních onemocnění.
4. Seznamte se se základy teorie a vlastnostmi nelineárních evolučních diferenciálních rovnic.
5. Zahajte práce na jejich vhodném numerickém řešení ve vybraných situacích.

Seznam doporučené literatury:

[1] J. D. Murray, Mathematical Biology. Springer, New York, 2007.
[2] M. J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals. De Gruyter, Princeton, 2011.
[3] H. R. Thieme, Mathematics in Population Biology. Princeton Series in Theoretical and Computational Biology, Princeton University Press, 2003.
[4] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London, Series A 115, 1927, 700-721.
[5] R. Ragonnet, et al., Vaccination Programs for Endemic Infections: Modelling Real versus Apparent Impacts of Vaccine and Infection Characteristics. Sci Rep 5, 15468, 2015.
[6] H. Gion, Y. Saito, S. Yazaki, On a backward bifurcation of an epidemic model with capacities of treatment and vaccination. JSIAM Letters (to appear), 2021.

Jméno a pracoviště vedoucí(ho) bakalářské práce:

Ing. Miroslav Kolář, Ph.D. katedra matematiky FJFI

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **31.10.2023** Termín odevzdání bakalářské práce: **05.08.2024**

Platnost zadání bakalářské práce: **30.09.2025**

Ing. Miroslav Kolář, Ph.D.	prof. Ing. Zuzana Masáková, Ph.D.	doc. Ing. Václav Čuba, Ph.D.
podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Studentka bere na vědomí, že je povinna vypracovat bakalářskou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v bakalářské práci.

1. 12. 2023
Datum převzetí zadání

Podpis studentky

Acknowledgment:
I would like to thank my supervisor Dr. Ing. Miroslav Kolář for his expert guidance, extreme patience and readiness to help with every question I had. His expertise and constructive feedback had a great influence on the content of this work.

Author's declaration:
I declare that this Bachelor's Degree Project is entirely my own work and I have listed all the used sources in the bibliography.

Prague, January 8, 2024 Daria Petrova

Název práce:

Matematické modelování v epidemiologii

Autor: Daria Petrova

Obor: Matematické inženýrství

Zaměření: Matematické modelování

Druh práce: Bakalářská práce

Vedoucí práce: Ing. Miroslav Kolář, Ph.D., Katedra matematiky FJFI ČVUT v Praze Trojanova 13, 120 00 Praha 2

Abstrakt: Tato bakalářská práce se zaměřuje na matematické modely v oblasti epidemiologie, a to konkrétně na SIR model bez demografie i s demografií. V první a druhé částech jsou zmíněné modely odvozeny a jejich vlastností prostudovány. Dále se práce zaměřuje na současné trendy v dané problematice. V poslední části je detailněji rozebrána numerická metoda, která byla použitá k nalezení přibližných řešení daných SIR modelů.

Klíčová slova: analyza stability, epidemie, experimentální řad konvergence, fazové krívky, Rungeova–Kuttova metoda, SIR model

Title:

Mathematical modeling in epidemiology

Author: Daria Petrova

Abstract: This work is focused on mathematical models in the field of epidemiology. As examples, the basic SIR model and SIR model with demography will be studied. In the first and second chapters the models will be derived and their properties will be reviewed. After that the modern trends for these models will be discussed. Finally, the work will focus on the numerical method, which was used to find the approximate solutions for the SIR models.

Key words: epidemic, experimental order of convergence, phase trajectories, Runge-Kutta method, SIR model, stability analysis

Contents

Introduction

Epidemic mathematical models have become a point of interest in last three years due to new epidemic of COVID-19. There are lots of them nowadays with different classes, different parameters and different assumptions. However, all of them have the same origin - the basic SIR model created by W. O. Kermack and A. G. McKendrick in 1927. This work will focus on this model and the variation of this model with demography. We will look into it's properties and stability and will try to find an approximate solution. We will continue with analysis of model's graphs and how change of the parameters will influence the phase trajectories.

One of the goals of this work is also to learn about modern trends and usage of the basic SIR model in solving nowadays problems. To reach this goal we will look into the research articles about COVID-19 and determine whether or not basic SIR model was used in these research papers.

Other two goals are to learn about non-linear evolution differential equations and numerical methods for solving them. The first task will be done during the research of properties of basic SIR model and SIR model with demography. The second task will consist of discovering numerical methods and comparing them to determine which one is more suitable for finding the solution. In addition we will estimate the accuracy of chosen method by calculating the experimental order of convergence.

Chapter 1

SIR model

In this chapter we will be focusing on the SIR model, which is a simple mathematical model for epidemic. The main idea of the model is that the population is divided into three classes: Susceptibles (S), Infected (I) and Recovered (R). We will be discussing required assumptions for the simple SIR model and for the model with demography, then we will derive the basic model using works of Kermack and McKendrick (1927) [2] and Keeling and Rohani (2011) [3]. We will describe the behaviour of each class of the model and will expand the model by adding demographic processes such as birth and natural death.

The main problem can be described as follows: Some number of the infected individuals are introduced into the community of susceptible individuals. The susceptibility varies depending on immunity, hygiene, overall health, environment and other factors. The disease spreads by contact of a susceptible individual with an infected one. Each infected person remains sick until recovery or death. The chances of recovery and the person's infectivity depend on the stage of sickness [2].

To start with basic SIR model, the following assumptions are needed to be made:

- The disease is acute, not chronic. The acute disease is relatively fast (pathogen will be removed after a short period of time) compared to an average lifespan of an individual, which means that all population changes (birth, natural death, migration) can not be taken into consideration. Thus, population remains constant during the epidemic [3].
- The incubation period is short enough to be negligible. A susceptible who contracts the disease is immediately infective [1].
- Every individual has the same probability to come in contact with each other.
- Recovery will lead to a lifelong immunity. Thus, individual can not be infected twice. If the disease is lethal, then dead individuals are still counted as recovered with immunity.
- At the beginning of the epidemic, no individual has an immunity.

These assumptions will restrict our model to simple cases without vaccination, demography or incubation period. Nevertheless, it will give us a model, which will be useful for studying the disease's properties, such as the conditions for the beginning and the end of epidemic, and for observing model's behaviour.

In this work we will be using the deriving from Kermack and McKendrick(1927) [2] and then simplification from Keeling and Rohani (2011) [3]. In SIR models population is divided into three groups: Susceptibles (S), Infected (I), Recovered (R). At the beginning almost all individuals are put into S category, with an exception of few I individuals. Transition from S class to I happens after successful disease

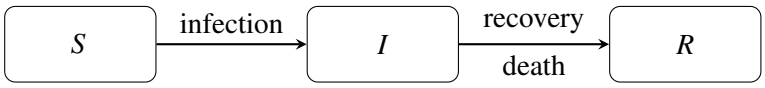


Figure 1.1: Schema of the transition between Susceptible, Infected and Recovered classes

transmission, from I to R after successful (recovery) or unsuccessful (death) fight with the disease. This dynamic is demonstrated in Figure 1.1. The probability of coming into a contact with infected person depends on the number of people in the I class. Therefore, there will be an influence on transmission from S to I.

For purposes of making the model we will denote X as the number of persons in S class, Y as the number of persons in I class and Z the number of persons in R class. As for the number of individuals in the whole population, we will use notation N. Next step will be dividing time into separate intervals. Infection will transmit only during the moment of passing between intervals. There will be no transmission during these intervals. The size of the interval will be taken as the unit of time.

To describe how individuals move to and out of the I class, we need to denote $v_{t,\theta}$ as the number of individuals, who are infected for θ intervals at the time t. If we make a sum through all intervals, we will get the total number who are ill at the time t:

$$Y_t = \sum_{\theta=0}^{t} v_{t,\theta}. \tag{1.1}$$

As expected, $v_{t,0}$ is the number individuals at the beginning of their infection. We will make another denotation v_t for people who were infected at the point $(t-1, t)$. In general $v_t = v_{t,0}$, except the point of origin, where we have assumed that at the beginning of the epidemic we have a few infected individuals Y_0. Thus

$$v_{0,0} = v_0 + Y_0. \tag{1.2}$$

The whole process of infection can be imagined as the scheme, which can be seen in Table 1.1 [2]:

Fresh infections.	Numbers at each stage of illness.	Number ill.
v_3	$v_{3,0} \nearrow v_{3,1} \nearrow v_{3,2} \nearrow v_{3,3}$	Y_3
v_2	$v_{2,0} \nearrow v_{2,1} \nearrow v_{2,2}$	Y_2
v_1	$v_{1,0} \nearrow v_{1,1}$	Y_1
v_0	$v_{0,0}$	Y_0

Table 1.1: Scheme of illness spread. Every $v_{t,\theta}$ will divide at the next stage into people who just became infected and who have been infected for some period.

Transition from the I class to the R class will be described with a help of a new denotation ψ_θ as a removal rate (in our case the sum of recovery and death rates). Now we can express the difference between number of infected at the time t and $t+1$ as follows:

$$v_{t,\theta} - v_{t+1,\theta+1} = \psi_\theta v_{t,\theta}.$$

Thus if we want to know the number of infected individuals at the time t through recursion, we get:

$$v_{t,\theta} = v_{t-1,\theta-1}(1-\psi_{\theta-1}) = v_{t-2,\theta-2}(1-\psi_{\theta-2})(1-\psi_{\theta-1}) = v_{t-\theta,0}B_\theta, \tag{1.3}$$

where B_θ is the product $(1-\psi_0)...(1-\psi_{\theta-2})(1-\psi_{\theta-1})$.

For the transition from the S class to the I class, we will need ϕ_θ as the rate of infectivity at the period θ. Then we have a formula for the v_t (number of individuals who became infected at the point $(t-1,t)$), since the chance of infection must be proportional to the number of infected on the one hand, and to the number of not yet infected on the other [2]:

$$v_t = X_t \sum_{\theta=0}^{t} \phi_\theta v_{t,\theta}.$$

where X_t is the number of uninfected individuals in S class at the time t. We can simplify the sum by making an assumption that $\phi_0 = 0$ (individual can not transmit disease at the moment of infection):

$$v_t = X_t \sum_{\theta=1}^{t} \phi_\theta v_{t,\theta}. \tag{1.4}$$

Obviously, the number of individuals, who have not been exposed to the virus yet, can be found by subtraction of people, who were infected at one period from the whole population. Using equation (1.2):

$$X_t = N - \sum_{x=0}^{t} v_{t,0} = N - \sum_{x=0}^{t} v_t - Y_0. \tag{1.5}$$

If we take Z_t as the number of recovered individuals at time t, then our population can be calculated as a sum of recovered, unaffected and infected people:

$$N = X_t + Y_t + Z_t. \tag{1.6}$$

Now we can rewrite (1.4) using (1.3) and (1.2) with the fact, that $v_t = v_{t,0}$, except at the point of origin:

$$\begin{aligned} v_t = X_t \sum_{\theta=1}^{t} \phi_\theta v_{t,\theta} = X_t \sum_{\theta=1}^{t} \phi_\theta v_{t-\theta,0} B_\theta = X_t(\sum_{\theta=1}^{t-1} \phi_\theta B_\theta v_{t-\theta} + \phi_t B_t v_{0,0}) = \\ = X_t(\sum_{\theta=1}^{t-1} \phi_\theta B_\theta v_{t-\theta} + \phi_t B_t v_0 + \phi_t B_t Y_0) = X_t(\sum_{\theta=1}^{t} A_\theta v_{t-\theta} + A_t Y_0), \end{aligned} \tag{1.7}$$

where $A_\theta = \phi_\theta B_\theta$.

Substitute (1.3) and (1.2) in (1.1) and get an equation for Y_t:

$$Y_t = \sum_{\theta=0}^{t} v_{t,\theta} = \sum_{\theta=0}^{t} B_\theta v_{t-\theta} + B_t Y_0. \tag{1.8}$$

By definition, the difference between unaffected people at the time t and $t-1$ is the number of individuals who became infected at the time t:

$$X_{t-1} - X_t = v_t, \tag{1.9}$$

hence equation (1.7) can be written as:

$$X_t - X_{t-1} = -v_t = -X_t(\sum_{\theta=1}^{t} A_\theta v_{t-\theta} + A_t Y_0). \tag{1.10}$$

Also $Z_{t+1} - Z_t$ is the number of individuals who are removed at the end of the interval of time t (for detailed calculation see [2]):

$$Z_{t+1} - Z_t = \sum_{\theta=1}^{t} \psi_\theta v_{t,\theta} = \sum_{\theta=1}^{t} \psi_\theta B_\theta v_{t-\theta} + \psi_t B_t Y_0 = \sum_{\theta=1}^{t} C_\theta v_{t-\theta} + C_t Y_0, \tag{1.11}$$

where $C_\theta = \psi_\theta B_\theta$. The last equation remaining is for Y_t, which we will get from (1.13):

$$Y_{t+1} - Y_t = X_t[\sum_{\theta=1}^{t} A_\theta v_{t-theta} + A_t Y_0] - [\sum_{\theta=1}^{t} C_\theta v_{t-\theta} + C_t Y_0]. \tag{1.12}$$

For the last step, we will assume that time intervals are very small and take a limit of (1.9) as t approaches $t-1$:

$$v_t = -\lim_{t \to t-1} X_t - X_{t-1} = -\lim_{t \to t-1} \frac{X_t - X_{t-1}}{t - t + 1} = -\frac{dX_t}{dt}$$

according to the definition of the derivative. Repeating the same algorithm with (1.11) and (1.10) and keeping in mind that $A_0 = C_0 = 0$, we get:

$$\begin{aligned} \frac{dX_t}{dt} &= -X_t[\int_0^t A_\theta v_{t-\theta}\, d\theta + A_t Y_0], \\ \frac{dZ_t}{dt} &= \int_0^t C_\theta v_{t-\theta}\, d\theta + C_t Y_0. \end{aligned}$$

Adding equation for Y_t from (1.8) and equation for the whole population, we will get a system of five equations:

$$N = X_t + Y_t + Z_t, \tag{1.13}$$

$$v_t = -\lim_{t \to t-1} X_t - X_{t-1} = -\lim_{t \to t-1} \frac{X_t - X_{t-1}}{t - t + 1} = -\frac{dX_t}{dt}, \tag{1.14}$$

$$\frac{dX_t}{dt} = -X_t[\int_1^t A_\theta v_{t-\theta}\, d\theta + A_t Y_0], \tag{1.15}$$

$$\frac{dZ_t}{dt} = \int_1^t C_\theta v_{t-\theta}\, d\theta + C_t Y_0, \tag{1.16}$$

$$Y_t = \int_0^t B_\theta v_{t-\theta} + B_t Y_0, \tag{1.17}$$

where $B_\theta = e^{-\int_0^\theta \psi_\alpha\, d\alpha}$, $A_\theta = \phi_\theta B_\theta$ and $C_\theta = \psi_\theta B_\theta$.

These equations are not independent and (1.13) is a necessary consequence of (1.15), (1.16), (1.17). Four independent equations (1.14), (1.15), (1.16), (1.17) determine the four functions X, Y, Z and v [2].

However, for further simplification we will make an assumption that the recovery rate ψ and rate of infectivity ϕ are constants; this leads to far more straightforward equations and exponentially distributed infectious periods [3].

Let κ be a number of contacts that one susceptible individual makes per unit of time. To simplify, we will represent the discrete population N by continuous interval [0, 1] and introduce the variables S, I, R as the population density for each class:

$$S = \frac{X}{N},$$
$$I = \frac{Y}{N},$$
$$R = \frac{Z}{N},$$

where N is the number of individuals in the population. Variables S, I, R are fraction, therefore:

$$S + I + R = 1. \tag{1.18}$$

Assuming Infected and Recovered have a linear connection through parameter γ, we get the first equation of SIR model:

$$\frac{dR}{dt} = \gamma I. \tag{1.19}$$

Next step is defining c as a probability of the successful transmission of the infection and δq as a probability that after a contact with an individual from Infected class a susceptible individual became infected. During an infinitesimally small period of time δt the number of contacts with infected individuals will be $\kappa \frac{Y}{N} \delta t$. Then the probability of not catching the disease is:

$$1 - \delta q = (1 - c)^{\kappa \frac{Y}{N} \delta t}.$$

We are interested in the probability of the successful transmission δq. If we define $\beta = -\kappa \ln(1 - c)$, then the equation for δq will transform to:

$$\delta q = 1 - e^{-\beta \frac{Y}{N} \delta t}.$$

Using Taylor expansion, we get:

$$\delta q \approx -\beta \frac{Y}{N} \delta t.$$

We will divide this equation by δt and take the limit of $\frac{\delta q}{\delta t}$ as $\delta t \to 0$

$$\frac{dq}{dt} = \beta \frac{Y}{N}.$$

For the whole Susceptible class we get:

$$\frac{dX}{dt} = \beta X \frac{Y}{N}.$$

Using previous equations, gives us the second differential equation of the basic SIR model without demography:

$$\frac{dS}{dt} = -\beta S I. \tag{1.20}$$

For the third evolution equation (1.18), (1.19), (1.20) are added together:

$$\frac{dI}{dt} = \frac{d(1-S-R)}{dt} = -\frac{dS}{dt} - \frac{dR}{dt} = \beta S I - \gamma I. \tag{1.21}$$

As a result, we get a system of three linear equations:

$$\frac{dS}{dt} = -\beta S I, \tag{1.22}$$

$$\frac{dI}{dt} = \beta S I - \gamma I, \tag{1.23}$$

$$\frac{dR}{dt} = \gamma I. \tag{1.24}$$

These equations are endowed with the initial conditions $S(0) > 0$, $I(0) > 0$, $R(0) = 0$. The typical behavior of these differential equations is shown on Figure 1.2.

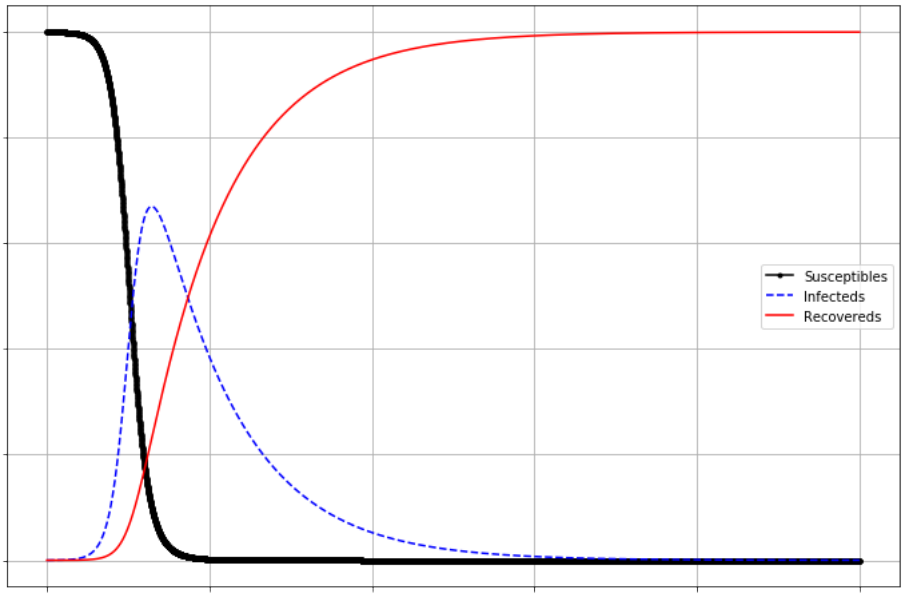


Figure 1.2: The graph of the basic SIR model given by equations (1.22 - 1.24): $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$, $\frac{dR}{dt} = \gamma I$. The figure is plotted with $\beta = 1.428$ and $\frac{1}{\gamma} = 7$ [3].

The basic SIR model was build on the assumption that the disease spread fast so it would not be affected by deaths or births in population. Some important properties could be learned from this model, but if we are interested in exploring the long-term epidemic, we definitely need to add demographic processes. The simplest and most common way to introduce demography into the SIR model is to assume there is a natural host "lifespan", $\frac{1}{\mu}$. Then, the rate at which individuals (in any epidemiological class) suffer natural mortality is given by μ [3]. Some authors assume that mortality influences only Recovered class ([18] Bailey 1975, [18] Keeling et al. 2001a, [18] Brauer 2002), but here the assumption will be from [3] that μ also represents the population's crude birth rate, so the size of population would

not change through time. The simple SIR model changes into the generalized SIR model:

$$\frac{dS}{dt} = \mu - \beta S I - \mu S, \tag{1.25}$$

$$\frac{dI}{dt} = \beta S I - \gamma I - \mu I, \tag{1.26}$$

$$\frac{dR}{dt} = \gamma I - \mu R. \tag{1.27}$$

This change drastically influences the behavior of the SIR curves, which is shown in Figure 1.3

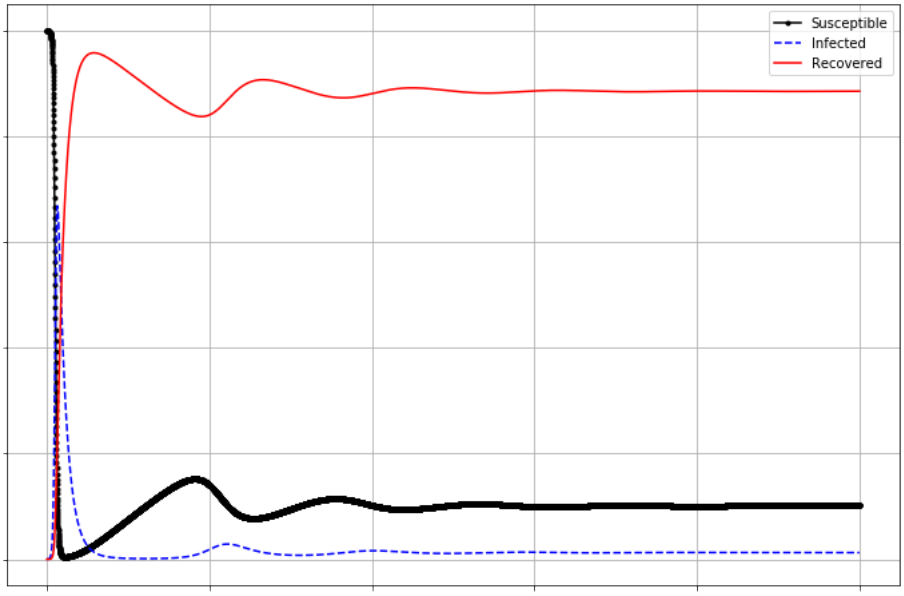


Figure 1.3: The graph of the SIR model with demography given by equations (1.25 - 1.27): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$, $\frac{dR}{dt} = \gamma I - \mu R$. The figure is plotted with $\beta = 5$ and $\frac{1}{\gamma} = 1$ and $\mu = \frac{1}{70}$.

In case of non-constant population size we need to differentiate between frequency- and density-dependent transmission [3]. These two kinds of transmission are based on how we expect the contact structure to change with population size. Frequency-dependent transmission is not influenced by the population size (the number of contacts will stay the same). Density-dependent transmission assumes that the bigger is density, the faster is transmission (contact rate rises) [3]. The first kind of transmission is applicable on human population, the latest on plant and animals [1].

Chapter 2

Mathematical properties of SIR model

In this chapter we will deal with mathematical properties of the basic SIR model and SIR model with demography. We will examine the life cycle of the epidemic, will consider when it has a possibility to start and what is the reason for it's end. With some additional assumptions we will be able to derive an approximate analytical expression for Recovered class for the basic SIR model. Then we will discuss the stability of the model with demography and look into stability analysis. Finally, we will study the phase trajectories for both models and try to determine the reasons of the system's behaviour.

2.1 The SIR Model Without Demography

2.1.1 The Threshold Phenomenon

Despite its simplicity, the SIR model (1.22) - (1.24) does not have an analytical solution; instead we will be looking for its numerical solution (Chapter 3). Nevertheless, the two most important qualitative properties can be studied on this model.

The first is called "The Threshold Phenomenon" by most of the authors. It will appear that for each particular set of infectivity, recovery and death rates, there exists a critical or threshold density of population. If the actual population density is equal to (or below) this threshold value the epidemic will not occur. It will appear also that the size of epidemic increases rapidly as the threshold density is exceeded [2].

We start with rewriting equation (1.23) in the form

$$\frac{dI}{dt} = I(\beta S - \gamma). \tag{2.1}$$

If S(0) is less then $\frac{\gamma}{\beta}$ (threshold value) then $\frac{dI}{dt} < 0$ and the infection "dies out". The inverse $R_0 = \frac{\beta}{\gamma}$ is called the basic reproductive ratio. The definition of this quantity is [8]: "the average number of secondary cases arising from an average primary case in an entirely susceptible population". The dynamics of disease transmission models is often described only by one parameter R_0, not the pair of γ, β parameters. The R_0 is commonly estimated from case data or by fitting the SIR model to the data. However, due to the assumption of neglecting the incubation period, the R_0 approximation can be underestimated. We can try to make R_0 more accurate by adding the incubating period, nevertheless, it will not give us an advantage. The attempts in estimating the incubation period will give the same inaccuracy, if not greater. [8].

With use of R_0 the threshold phenomenon will change to the following: assuming S(0) = 1, a pathogen can invade only if $R_0 > 1$ [3]. The logical explanation for this is that any disease cannot

spread if it cannot successfully transmit to more than one new susceptible [12]. R_0 depends both on the disease and population. Mathematically, basic reproductive ratio represents the rate at which new cases are produced by an infectious person multiplied by the average infectious period.

2.1.2 Epidemic Burnout

The second property is the answer of the question: "When will the epidemic burnout?" First, we will divide equation (1.22) by (1.24):

$$\frac{dS}{dR} = -\frac{\beta}{\gamma}S = -R_0 S. \tag{2.2}$$

Then we will integrate with respect to R, assuming that R(0)=0:

$$\begin{aligned} \int \frac{1}{S}\, dS &= \int -R_0\, dR, \\ \ln S &= -R_0 R + C, \\ S(t) &= e^{-R_0 R(t)} e^C, \end{aligned}$$

solving for t = 0, we get the value of a constant:

$$S(0) = e^C.$$

The final result is:

$$S(t) = S(0)e^{-R_0 R(t)}. \tag{2.3}$$

As the epidemic develops, the Susceptible class declines, but never reaches zero. The meaning of this is that there will be some individuals in the Susceptible class, who escaped the infection. This leads to a very important conclusion: the chain of transmission eventually breaks due to the decline in infectives (not due to a complete lack of susceptibles) [3].

Keeping in mind that the epidemic ends when $I = 0$, and using (1.18), we can get an equation for a long-term behavior of (2.3):

$$\begin{aligned} S(\infty) = 1 - R(\infty) &= S(0)e^{-R(\infty)R_0}, \\ 1 - R(\infty) - S(0)e^{-R(\infty)R_0} &= 0, \end{aligned} \tag{2.4}$$

where $R(\infty)$ is the final portion of recovered individuals, which is equal to the total portion of infected individuals.

The next important thing that we want to get is the expression for the epidemic curve. To obtain it, we will use the expressions (1.18), (2.3) and we will rewrite the expression 1.24:

$$\frac{dR}{dt} = \gamma(1 - S(t) - R(t)) = \gamma(1 - S(0)e^{-R_0 R(t)} - R(t)).$$

This equation does not have analytical solution for all $R_0 R$, but for small values we can use Maclaurin series for exponential function to approximate:

$$\begin{aligned} \frac{dR}{dt} &\approx \gamma\left[1 - S(0)\left(1 - R_0 R(t) + \frac{(R(t)R_0)^2}{2}\right) - R(t)\right] = \\ &= \gamma\left[1 - S(0) + S(0)R_0 R(t) - \frac{S(0)(R(t)R_0)^2}{2} - R(t)\right] = \\ &= \gamma\left[1 - S(0) + (S(0)R_0 - 1)R(t) - \frac{S(0)R_0^2}{2}R^2(t)\right] = \\ &= c + bR(t) + aR^2(t), \end{aligned} \tag{2.5}$$

where

$$c = \gamma(1 - S(0)) = \gamma I(0),$$
$$b = \gamma(S(0)R_0 - 1),$$
$$a = -\frac{\gamma S(0) R_0^2}{2}.$$

This is a separable differential equation, so we will separate the variables and then use partial fraction decomposition. For this we, firstly, need to calculate solutions of:

$$c + bR + aR^2 = 0,$$

$$D = b^2 - 4ac.$$

Discriminant is above zero because $c > 0$ and $a < 0$, so as the result we get a sum of two positive numbers. If we denote R_1 and R_2 as two solutions, we get:

$$R_1 = \frac{-b + \sqrt{D}}{2a} = \frac{1 - S(0)R_0 + \sqrt{(S(0)R_0 - 1)^2 + 2I(0)S(0)R_0^2}}{-S(0)R_0^2},$$

$$R_2 = \frac{-b - \sqrt{D}}{2a} = \frac{1 - S(0)R_0 - \sqrt{(S(0)R_0 - 1)^2 + 2I(0)S(0)R_0^2}}{-S(0)R_0^2}.$$

Then the decomposition will be as follows:

$$\frac{1}{c + bR + aR^2} = \frac{1}{a(R - R_1)(R - R_2)} = \frac{1}{a}\left(\frac{A}{(R - R_1)} + \frac{B}{(R - R_2)}\right) = \frac{1}{a}\frac{A(R - R_2) + B(R - R_1)}{(R - R_1)(R - R_2)},$$

$$A + B = 0 \Rightarrow A = -B,$$

$$-AR_2 - BR_1 = 1 \Rightarrow -AR_2 + AR_1 = 1 \Rightarrow A = \frac{1}{R_1 - R_2}.$$

Now after the separation the left part will be:

$$\int \frac{dR}{c + bR + aR^2} = \frac{1}{a(R_1 - R_2)} \int \left(\frac{1}{(R - R_1)} - \frac{1}{(R - R_2)}\right) dR = \frac{1}{a(R_1 - R_2)} \ln\left|\frac{R - R_1}{R - R_2}\right|.$$

Next question is when the expression $\frac{R-R_1}{R-R_2}$ is more than 0 and when is less. We will start by removing the square root:

$$\frac{R - R_1}{R - R_2} = \frac{R - \frac{-b+\sqrt{D}}{2a}}{R - \frac{-b-\sqrt{D}}{2a}} = \frac{2aR + b - \sqrt{D}}{2aR + b + \sqrt{D}} == \frac{(2aR + b)^2 - D}{(2aR + b + \sqrt{D})^2}.$$

The denominator $(2aR + b + \sqrt{D})^2$ is positive, therefore we can focus on numerator:

$$\begin{aligned}(2aR + b)^2 - D &= 4a^2R^2 + 4aRb + b^2 - b^2 + 4ac = 4a(aR^2 + Rb + c) = \\ &= \frac{-4\gamma S(0)(R_0)^2}{2}\left(-\frac{\gamma S(0)R_0^2}{2}R^2 + R\gamma(S(0)R_0 - 1) + \gamma(1 - S(0))\right) = \\ &= \gamma^2 S(0)(R_0)^2\left[S(0)((R_0R - 1)^2 + 1) - 2 + 2R\right] \approx \\ &\approx \gamma^2 S(0)(R_0)^2\left[2S(0) - 2 + 2R\right].\end{aligned}$$

This gives us the following observation: for $R < 1 - S(0) = I(0)$ the expression $\frac{R-R_1}{R-R_2}$ is less than 0. This situation characterizes the beginning of the epidemic. For the end of the epidemic we get $R \geq 1 - S(0) = I(0)$ and the expression $\frac{R-R_1}{R-R_2}$ is greater than 0 or ist equal to 0. The final form of equation is:

$$\frac{\ln\left(\frac{R-R_1}{R-R_2}\right)}{a(R_1 - R_2)} = t + C, \ for\ R(t) \geq I(0),$$
$$\frac{\ln\left(-\frac{R-R_1}{R-R_2}\right)}{a(R_1 - R_2)} = t + C, \ for\ R(t) < I(0),$$

where C is integration constant. Solving in terms of R, we get:

$$\ln\left(\pm\frac{R - R_1}{R - R_2}\right) = (t + C)a(R_1 - R_2),$$
$$\pm\frac{R - R_1}{R - R_2} = e^{(t+C)a(R_1-R_2)},$$

finding constant by calculating the equation for t = 0:

$$\pm\frac{R_1}{R_2} = e^{Ca(R_1-R_2)}.$$

As the result:

$$\pm\frac{R - R_1}{R - R_2} = \pm e^{ta(R_1-R_2)}\frac{R_1}{R_2},$$
$$\frac{R - R_2 + (R_2 - R_1)}{R - R_2} = e^{ta(R_1-R_2)}\frac{R_1}{R_2},$$
$$1 + \frac{R_2 - R_1}{R - R_2} = e^{ta(R_1-R_2)}\frac{R_1}{R_2},$$
$$\frac{R - R_2}{R_2 - R_1} = \frac{1}{e^{ta(R_1-R_2)}\frac{R_1}{R_2} - 1},$$
$$R = \frac{R_2 - R_1}{e^{ta(R_1-R_2)}\frac{R_1}{R_2} - 1} + R_2 = \frac{R_1(e^{ta(R_1-R_2)} - 1)}{e^{ta(R_1-R_2)}\frac{R_1}{R_2} - 1}.$$

After reverse substitution:

$$R = \frac{1}{S(0)R_0^2}\frac{(S(0)R_0 - 1 + \zeta)(e^{t\gamma\zeta} - 1)}{e^{t\gamma\zeta}\frac{S(0)R_0-1+\zeta}{S(0)R_0-1-\zeta} - 1} = \frac{1}{S(0)R_0^2}\frac{(S(0)R_0 - 1 + \zeta)(e^{t\gamma\zeta} - 1)}{e^{t\gamma\zeta}\frac{S(0)R_0-1+\zeta+2\zeta}{S(0)R_0-1-\zeta} - 1} =$$
$$= \frac{1}{S(0)R_0^2}\frac{(S(0)R_0 - 1 + \zeta)(e^{t\gamma\zeta} - 1)}{e^{t\gamma\zeta} - 1 + \frac{e^{t\gamma\zeta}2\zeta}{S(0)R_0-1+\zeta}} =$$
$$= \frac{1}{S(0)R_0^2}\frac{(S(0)R_0 - 1 + \zeta)(e^{t\gamma\zeta} - 1)(S(0)R_0 - 1 - \zeta)}{(e^{t\gamma\zeta} - 1)(S(0)R_0 - 1) - \zeta(e^{t\gamma\zeta} - 1) + e^{t\gamma\zeta}2\zeta} =$$
$$= \frac{1}{S(0)R_0^2}\frac{((S(0)R_0 - 1)^2\zeta^2)(e^{t\gamma\zeta} - 1)}{(e^{t\gamma\zeta} - 1)(S(0)R_0 - 1) + \zeta(e^{t\gamma\zeta} + 1)} =$$
$$= \frac{1}{S(0)R_0^2}\frac{((S(0)R_0 - 1)^2 - \zeta^2)(e^{t\gamma\zeta} - 1)}{(\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t)(S(0)R_0 - 1) + 1)\zeta(e^{t\gamma\zeta} + 1)} =$$
$$= \frac{1}{S(0)R_0^2}\frac{((S(0)R_0 - 1)^2 - \zeta^2)\tanh(\frac{1}{2}\zeta\gamma t)}{(\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t)(S(0)R_0 - 1) + 1)} =$$
$$= \frac{\zeta}{S(0)R_0^2}\frac{((S(0)R_0 - 1)^2\frac{1}{\zeta^2}\tanh(\frac{1}{2}\zeta\gamma t) - \tanh(\frac{1}{2}\zeta\gamma t)}{(\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t)(S(0)R_0 - 1) + 1)} =$$
$$= \frac{\zeta}{S(0)R_0^2}\frac{\frac{S(0)R_0-1}{\zeta}((S(0)R_0 - 1)\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t) + 1) - \frac{S(0)R_0-1}{\zeta} - \tanh(\frac{1}{2}\zeta\gamma t)}{(\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t)(S(0)R_0 - 1) + 1)} =$$
$$= \frac{1}{S(0)R_0^2}\left(S(0)R_0 - 1 - \frac{(S(0)R_0 - 1) + \zeta\tanh(\frac{1}{2}\zeta\gamma t)}{(\frac{1}{\zeta}\tanh(\frac{1}{2}\zeta\gamma t)(S(0)R_0 - 1) + 1)}\right) =$$
$$= \frac{1}{S(0)R_0^2}\left(S(0)R_0 - 1 + \zeta\tanh(\frac{1}{2}\zeta\gamma t - \phi)\right),$$

where $\zeta = \sqrt{(S(0)R_0 - 1)^2 + 2I(0)S(0)R_0^2}$ and $\phi = \tanh^{-1}\left(\frac{1}{\alpha}(S(0)R_0 - 1)\right)$.

The final result:

$$R(t) = \frac{1}{S(0)R_0^2}\left[S(0)R_0 - 1 + \zeta\tanh\left(\frac{1}{2}\zeta\delta t - \phi\right)\right], \tag{2.6}$$

where $\phi = \tanh^{-1}\left(\frac{1}{\alpha}(S(0)R_0 - 1)\right)$, $\zeta = \sqrt{(S(0)R_0 - 1)^2 + 2I(0)S(0)R_0^2}$.

To verify the correctness of the result, we will use another method to solve equation (2.5). Let's start with substitution $R(t) = u(t) - \frac{b}{2a}$ [14]:

$$\frac{du}{dt} = a\left(u - \frac{b}{2a}\right)^2 + b - \frac{b^2}{2a} + c = au^2 - \frac{b^2}{2a} + c = au^2 - \frac{D}{4a},$$

where $D = b^2 - 4ac$. Then we will substitute $u = w\frac{\sqrt{D}}{2a}$:

$$\frac{\sqrt{D}}{2a}\frac{dw}{dt} = \frac{du}{dt} = au^2 - \frac{D}{4a} = w^2\frac{D}{4a} - \frac{D}{4a},$$

$$\frac{\sqrt{D}}{2a}\frac{dw}{dt} = (w^2 - 1)\frac{D}{4a},$$

$$\frac{dw}{w^2-1} = \frac{\sqrt{D}}{2}dt.$$

Solving differential equation:

$$\tanh^{-1} w = -\frac{\sqrt{D}}{2}t + C,$$

where C is a constant. After applying reverse substitution:

$$u(t) = \frac{\sqrt{D}}{2a}\tanh\left(-\frac{\sqrt{D}}{2} + C\right).$$

$$R(t) = -\frac{b}{2a} + \frac{\sqrt{D}}{2a}\tanh\left(-\frac{\sqrt{D}}{2} + C\right),$$

Constant C can be found from initial condition $R(0) = 0$:

$$0 = -\frac{b}{2a} + \frac{\sqrt{D}}{2a}\tanh C,$$

$$\tanh C = \frac{b}{\sqrt{D}},$$

$$C = \tanh^{-1}\left(\frac{b}{\sqrt{D}}\right).$$

With this result and reverse substitution for a, b, c we get:

$$\begin{aligned} R(t) &= -\frac{b}{2a} + \frac{\sqrt{D}}{2a}\tanh\left(-\frac{\sqrt{D}}{2}\right) + \tanh^{-1}\left(\frac{b}{\sqrt{D}}\right) = \\ &= \frac{S(0)R_0 - 1}{S(0)R_0^2} - \frac{\zeta}{S(0)R_0^2}\tanh\left(-\frac{\gamma\zeta t}{2}\right) + \tanh^{-1}\left(\frac{S(0)R_0 - 1}{\zeta}\right) = \\ &= \frac{1}{S(0)R_0^2}\left(S(0)R_0 - 1 + \zeta\tanh\left(-\frac{\gamma\zeta t}{2} + \phi\right)\right), \end{aligned}$$

where

$$\phi = \tanh^{-1}\left(\frac{S(0)R_0 - 1}{\zeta}\right),$$

$$\zeta = \sqrt{(S(0)R_0 - 1)^2 + 2I(0)S(0)R_0^2}.$$

The final result

$$R(t) = \frac{1}{S(0)R_0^2}\left(S(0)R_0 - 1 + \zeta\tanh\left(-\frac{\gamma\zeta t}{2} + \phi\right)\right)$$

was reached by assuming that R_0R is small. This condition is most likely to be met at the beginning at the epidemic and will not be very accurate for highly infectious diseases [3]. Another important curve is an epidemic curve. Expression for it can be obtained by differentiating $R(t)$ (2.6) in terms of time t:

$$\frac{dR}{dt} = \frac{\delta\zeta^2}{2S(0)R_0^2}\operatorname{sech}^2\left(\frac{1}{2}\zeta\delta t - \phi\right). \tag{2.7}$$

For any specific epidemic γ and ϕ can be estimated from a data. The derivation also requires R_0R to be small.

The example of $R(t)$ can be seen in Figure 2.1. The coefficients were taken from the Bombay plague epidemic of 1905-1906. The epidemic was not severe compared to the population size. This allowed to find the best fit for three parameters and get the result for an epidemic curve [1]:

$$\frac{dR}{dt} = 890 sech^2(0.2t - 3.4),$$

which can be seen in Figure 2.2

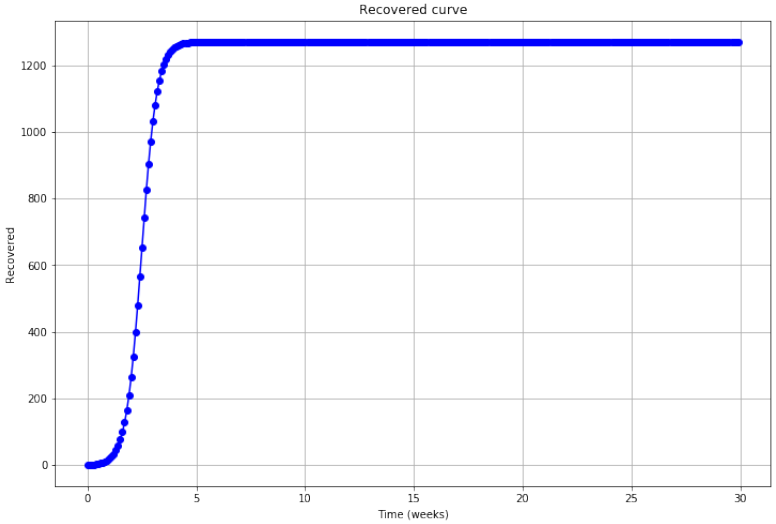


Figure 2.1: Example of R(t) approximation (2.6) $R(t) = \frac{1}{S(0)R_0^2}\left(S(0)R_0 - 1 + \zeta \tanh\left(-\frac{\delta\zeta t}{2} + \phi\right)\right)$ with coefficients derived from the data of the plague in Bombay 1905.

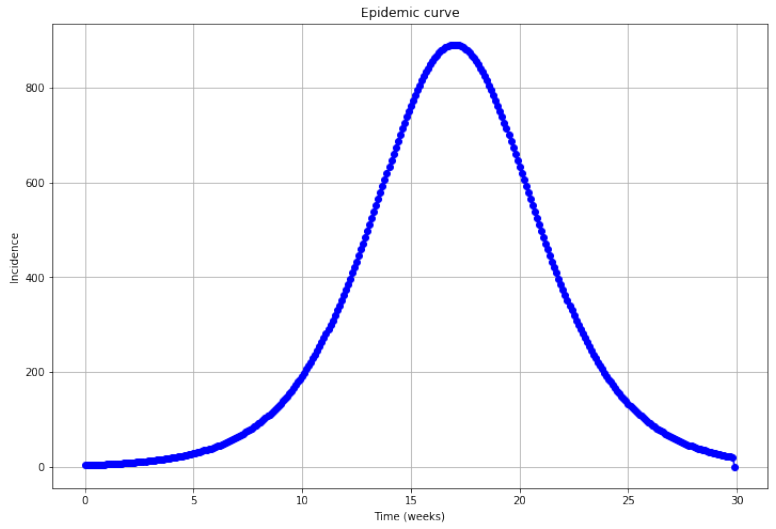


Figure 2.2: Example of epidemic curve (2.7) given by equation $\frac{dR}{dt} = 890\text{sech}^2(0.2t - 3.4)$ derived from the data of the plague in Bombay 1905 [3].

2.2 The SIR Model With Demography

2.2.1 The Equilibrium State

The changes in host demographic can lead to the long term disease persistence [3]. To tell what is going to happen, we can look at the system at equilibrium state with

$$\frac{dS}{dt} = \frac{dI}{dt} = \frac{dR}{dt} = 0.$$

First equilibrium is trivial and is called the disease-free equilibrium. This is the situation when the pathogen has died out and in the long term all population consists of Susceptibles. Mathematically it is the point $(S, I, R) = (1, 0, 0)$.

Other equilibrium is the endemic equilibrium. In epidemiology the term endemic means that infection is constantly present in population or is maintained on the same level, without extra infection being brought. Therefore, we start by setting the equation for Infected (1.26) to zero:

$$I(\beta S - (\gamma + \mu)) = 0,$$

from where we can see that $I = 0$ or $\beta S - \gamma - \mu = 0$. First one is the disease-free equilibrium. The second one is for endemic, we get condition:

$$S = \frac{\gamma + \mu}{\beta},$$

which is the inverse of R_0. Then we substitute the result for S into (1.25):

$$\mu - \beta\frac{1}{R_0}I - \mu\frac{1}{R_0} = 0, \tag{2.8}$$

$$\frac{\beta}{R_0}I = \mu(1 - \frac{1}{R_0}), \tag{2.9}$$

$$I = \frac{\mu}{\beta}(R_0 - 1). \tag{2.10}$$

Using that the sum of S, I and R is 1 (1.18), we can get an expression for R:

$$R = 1 - S - I = 1 - \frac{1}{R_0} - \frac{\mu}{\beta}(R_0 - 1).$$

Therefore, the point of the endemic equilibrium is:

$$(S, I, R) = \left(\frac{1}{R_0}, \frac{\mu}{\beta}(R_0 - 1), 1 - \frac{1}{R_0} - \frac{\mu}{\beta}(R_0 - 1)\right). \tag{2.11}$$

2.2.2 Stability Properties

The study of an equilibrium cannot be complete without stability analysis. Consider a general vector field:

$$\dot{x} = f(x, t),\ x \in \mathbf{R^n}, \tag{2.12}$$
$$x(t_0) = x_{ini}.$$

Let $\phi = \phi(t; t_0, x_{ini})$ be any solution of 2.12, $\|.\|$ is a norm on $\mathbf{R^n}$. Then we can define stability:

Definition 1. *The solution ϕ is stable if for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) > 0$ such that, for any other solution, $y(t)$, of 2.12 satisfying $\|\phi(t_0) - y(t_0)\| < \delta$, then $\|\phi(t) - y(t)\| < \epsilon$ for $t > t_0$, $t_0 \in \mathbf{R}$ [22].*

And asymptotic stability:

Definition 2. *The solution ϕ is asymptotically stable if it is stable and for any other solution, $y(t)$, there exist a constant $b > 0$ such that, if $\|\phi(t_0) - y(t_0)\| < b$, then $\lim_{t\to\infty} \|\phi(t) - y(t)\| = 0$ [22].*

In order to determine the stability of solution ϕ we can study the behaviour of other solutions near it:

$$x = \phi(t) + y(t).$$

Substituting x in (2.12) we get:

$$\dot{x} = \dot{\phi} + \dot{y} = f(\phi(t), t) + \dot{y}.$$

Also we can use Taylor expansion for $f(x, t)$ around $\phi(t)$ and get:

$$\dot{x} = f(\phi(t)) + Df(\phi(t))(x - \phi(t)) + O(|(x - \phi(t))^2|) = f(\phi(t)) + Df(\phi(t))y + O(|y^2|).$$

Aligning these two equations:

$$\dot{y} = Df(\phi(t))y + O(|y^2|).$$

Thus, if we want to describe the behavior of the solutions close to ϕ, we need to study the following linear system:

$$\dot{y} = Df(\phi(t))y.$$

To tell if the solution ϕ is stable we will use Poincaré-Lyapunov theorem:

Theorem 1. *Suppose all of the eigenvalues of $Df(\phi(t))$ have negative real parts. Then the equilibrium solution $x = \phi(t)$ of the nonlinear vector field 2.12 is asymptotically stable [22].*

Returning to the SIR model, the Df will be:

$$Df = \begin{pmatrix} -\beta I - \mu & -\beta S & 0 \\ \beta I & \beta S - (\mu + \gamma) & 0 \\ 0 & \gamma & -\mu \end{pmatrix}.$$

Next step is to obtain the characteristic polynomial:

$$(-\beta I - \mu - \lambda)(\beta S - (\mu + \gamma) - \lambda)(-\mu - \lambda) + (\beta I)(\beta S)(-\mu - \lambda) = 0.$$

The first eigenvalue is $\lambda_1 = -\mu$, which is negative. For the others we need to take in consideration our equilibriums. First one is the disease-free equilibrium, after the substitution ($S = 0, I = 0$) we get:

$$(-\mu - \lambda)(\beta - (\mu + \gamma - \lambda)).$$

As the result, $\lambda_2 = -\mu$ and $\lambda_3 = \beta - (\mu + \gamma)$. For this equilibrium to be stable (for all eigenvalues to be negative) this criterion must be met:

$$\beta < \mu + \gamma,$$

which translates into ensuring $R_0 < 1$ [3].

For the endemic equilibrium ($S = \frac{\gamma+\mu}{\beta}, I = \frac{\mu}{\beta}(R_0 - 1)$) we get the following expression:

$$(-\mu(R_0 - 1) - \mu - \lambda)(\gamma + \mu - (\mu + \gamma) - \lambda) + \mu(R_0 - 1)(\gamma + \mu) = 0,$$

$$(-\mu(R_0 - 1) - \mu - \lambda)(-\lambda) + \mu(R_0 - 1)(\gamma + \mu) = 0,$$
$$\lambda^2 + \mu R_0 \lambda + \mu(R_0 - 1)(\gamma + \mu) = 0,$$

the solution of which is:

$$\lambda_{2,3} = -\frac{\mu R_0 \pm \sqrt{(\mu R_0)^2 - \frac{4}{AG}}}{2},$$

where the term $A = \frac{1}{\mu(R_0-1)}$ denotes the mean age at infection and $G = \frac{1}{\mu+\gamma}$ determines the typical period of a host's infectivity [3].

To continue with these eigenvalues, we can notice that $(\mu R_0)^2 = (\mu\frac{\beta}{\gamma+\mu})^2$ is often small enough to ignore, hence we can approximate:

$$\lambda_{2,3} \approx -\frac{\mu R_0}{2} \pm \frac{i}{\sqrt{AG}}.$$

Therefore, the endemic equilibrium is feasible only when R_0 is greater than one, but it is always stable. The fact that dominant eigenvalues are complex tells us that equilibrium is approached via oscillatory dynamics. The simple SIR without demography is an excellent example of a damped oscillator, the amplitude of fluctuation declines over time.

2.2.3 Phase Trajectories

The following set of figures is representing phase trajectories of models with and without demography for different parameters β, γ and μ.

We will start with the basic SIR without demography. In Figure 2.3 we see a typical example of phase trajectories for simple SIR model. The trajectories are set for different initial conditions of $I(0)$: 0.9 (the highest one), 0.8, 0.6, 0.5, 0.3, 0.03 (the lowest one). The values for $S(0)$ are found by using the fact that since $R(0) = 0$ then $S + I = 1$, so $S = 1 - I$. The trajectories start on the line $S + I = 1$ and remain within a triangle since $0 < S + I < 1$. An epidemic exists if $I(t) > I(0)$ for any $t > 0$ [1]. The full picture of the all three curves' behaviour can be found in three dimensional representation of phase trajectories (Figure 2.5).

Results for different parameters β and γ can be seen in Figure 2.4. For $\beta > \gamma$ (Figures 2.4, a and 2.4, c) trajectories are curves or polygonal chains and for $\beta < \gamma$ (Figures 2.4, b and 2.4, d) they are lines. In the first case, the more is the difference the longer are segments and the higher polygonal chains are. If the β (probability of catching an infection) is too high, the speed of spread will rise giving us the sharp increase in Infected and decrease of Susceptibles. Combining with the fact that the speed of recovery γ is slow, Infected will get a sharp rise with a small decrease leading to higher curve position and longer segments of no decline (Figure 2.4, c). For the second case, the increase in difference leads to decrease in tilt angle. The probability of catching a disease now is very small and the speed of recovery from illness is high giving us almost no decrease in Susceptibles and sharp decrease of Infected. As the result we see smaller and smaller tilt angle the bigger the recovery speed is (Figure 2.4, d). More of this behaviour can be seen on phase trajectories in three dimensions (Figure 2.6).

For the SIR with demography, we are getting a third parameter - natural mortality μ. It gives us an oscillatory dynamic which is also seen in phase trajectories towards the end of a trajectory (Figure 2.7). The third dimensional view on oscillation is captured in Figure 2.10. The change in this parameter (Figure 2.8) gives us higher and lower position of oscillation. For higher values of μ the oscillation point moves higher. If the mortality rate is high, we will see the fast decrease of the population. Both Infected and Susceptibles are dependent on the population size, so the decrease in population will move the end points of the curves to a higher level, making oscillation appear higher (Figure 2.8, b). More confirmation can be found when studying phase trajectories in three dimensions (Figure 2.11).

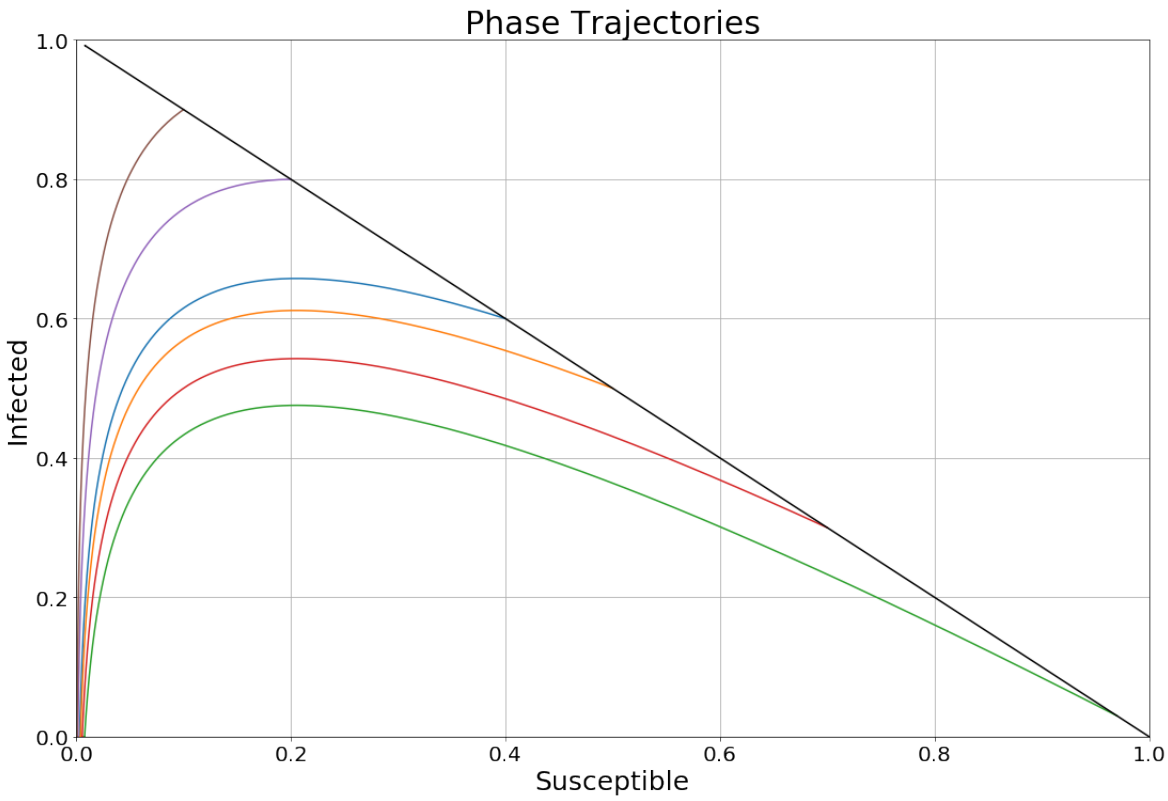


Figure 2.3: Phase trajectories of Susceptible and Infected given by equations (1.22), (1.23): $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$ with parameters $\beta = 0.18$, $\gamma = 0.037$.

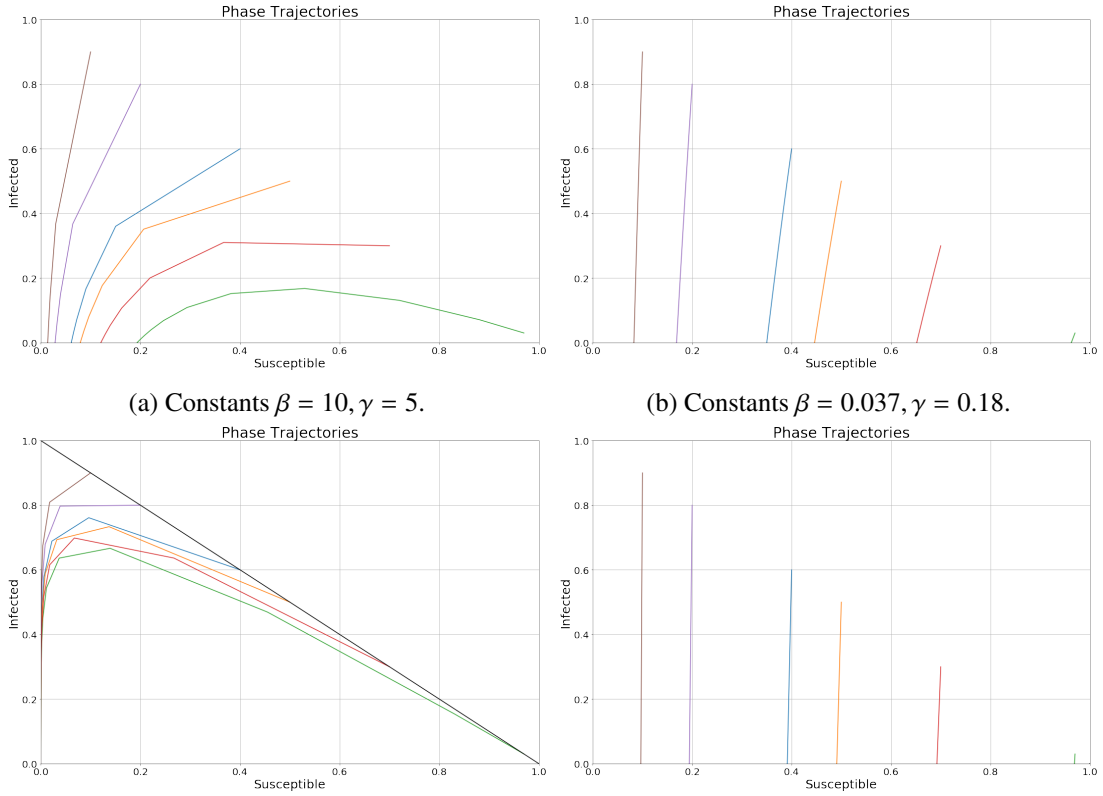
For the case of no epidemic ($\beta = 0.037$ and $\gamma = 0.18$) this oscillation turns into a slight curve (Figure 2.9). It shows that even with the mortality the decrease in Infected is still faster than decrease in Susceptibles. In Figure (Figure 2.9, b) the speed of population's decrease is so high, that despite the absence of epidemic, the speed of Infected and Susceptibles decrease is almost the same. The three dimensional representation of these situations can be found in Figure 2.12.

To sum up, for the model without demography the chances of epidemic occurrence are higher for the cases when $\beta > \gamma$ (Figures 2.4, a and 2.4, c), which is in accordance with the condition $R_0 > 1$. And we can see, that the end of epidemic happens not due to lack of Susceptibles (Figures 2.4, b and 2.4, d), which also corresponds with our theoretic findings.

For the model with demography we also count the influence of parameter μ. If the $\beta > \gamma + \mu$ the pathogen will be present in population long-term (2.8). This characterizes the endemic equilibrium. For the disease-free equilibrium we see no oscillation no matter how big is parameter μ (Figure 2.9).

2.2.4 Modern trends

Nowadays there are a lot of extensions of the simple SIR model. The SI model for illnesses with mortality risk, the SIS model for cases without immunity, the SIR model for immunity that lasts for a limited time, the SEIR model for diseases with latent period and more. Nevertheless, even the simple SIR model is used today in epidemiology researches. One of the examples is the research of COVID-19 in Algeria [9] (Fig. 2.13). We can use basic SIR to predict the disease spread in this situation because of the following reasons:



(a) Constants $\beta = 10, \gamma = 5$.

(b) Constants $\beta = 0.037, \gamma = 0.18$.

(c) Constants $\beta = 10, \gamma = 1$.

(d) Constants $\beta = 0.18, \gamma = 5$.

Figure 2.4: Phase trajectories given by equations (1.22), (1.23): $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$ with different parameters β and γ.

- No vaccination, curfews, lockdowns and social distancing were there in the time of research.
- Although Algeria's health-care system ranks among the best in Africa, it lags far behind standards of wealthier countries [19].
- The population is concentrated at northern regions with high density so the number of contacts between people is high.
- Lifecycle of the COVID-19 is relatively short so we can ignore the demography.

Another good use of the model it to determine parameter for infectivity and the basic reproductive ratio [6]. This case is represented in research of the COVID-19 in Italy [10] (Figure 2.14),

Recent trend in epidemiology is to not only customize the model with various specific parameters but also to study the behavioral responses of individuals to the infection's spread. This has led to the birth of a new branch of mathematical epidemiology, which is called the behavioral epidemiology of infectious diseases [16]. It takes into consideration lockdowns, restrictions, social distancing and other governmental practises to determine whether or not they are effective. It is done by including a new function, which is called the control function. Then the task is to optimize this function to prevent spread of the disease or to slow it down. The other function that is added to the model is the information index. This index describes the information regarding the status of the disease in the community [20].

For example, in order to compare of how different countries manage epidemics, Spatial SIR model was introduced. The name comes from adding the spatial dimension to SIR model. It was very important

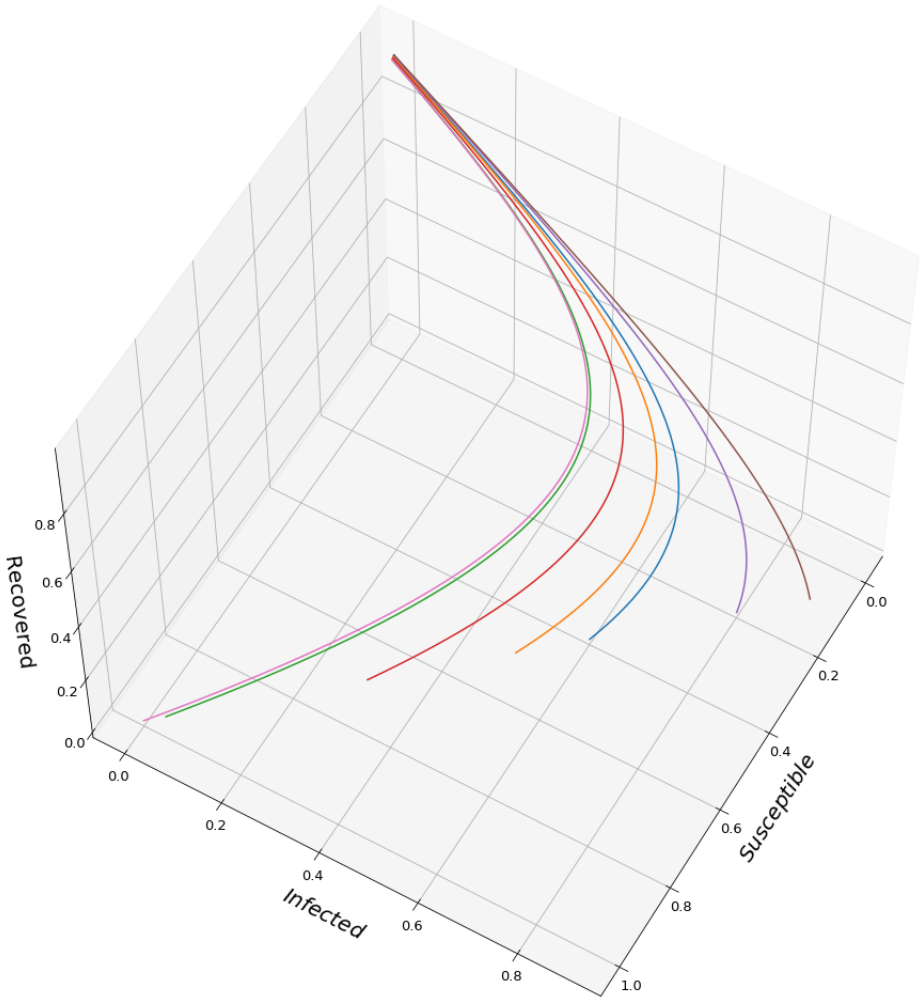
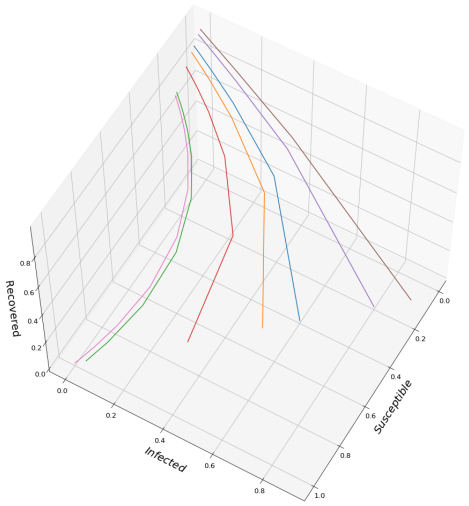


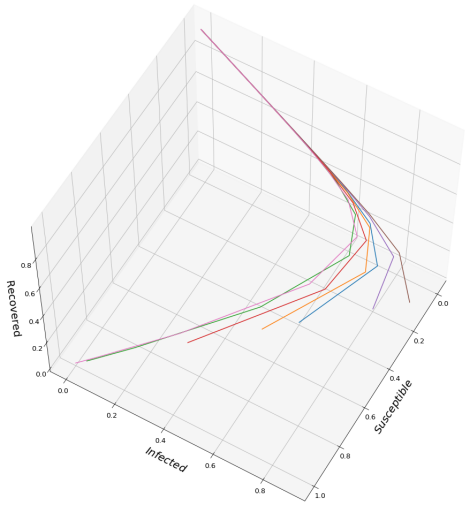
Figure 2.5: Phase trajectories given by equations (1.22), (1.23), (1.24): $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$, $\frac{dR}{dt} = \gamma I$.

in researching the spread of SARS-CoC-2 under lockdown restriction. It allowed to study the epidemic locally and count the local herd immunities to reach more accurate result [24]. Another example is Distance-contagion model. It is used specifically to estimate the effectiveness of the social distancing. The infectivity rate there is depended on the average dynamic distance that individuals usually keep from each other [25].

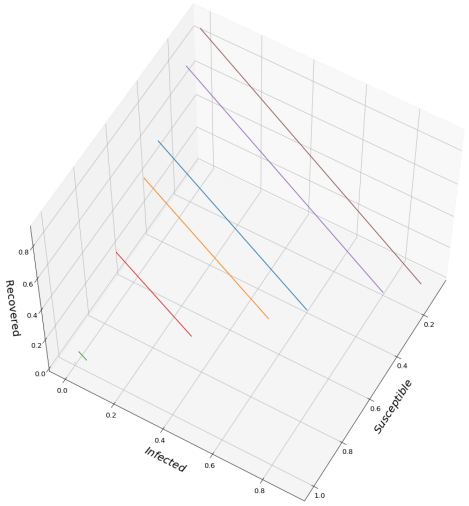
The other trend is to use these epidemiological models not only for biological diseases, but also for spread of computer viruses, online social networks and for the economic epidemiology [21]. It can be used to predict the spread of gossips, to research viral market strategies, to understand the spread of information on complex networks.



(a) Constants $\beta = 10, \gamma = 5$.



(b) Constants $\beta = 10, \gamma = 1$.



(c) Constants $\beta = 0.18, \gamma = 5$.

Figure 2.6: Phase trajectories given by equations (1.22), (1.23), (1.24): $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$, $\frac{dR}{dt} = \gamma I$ with different parameters β and γ.

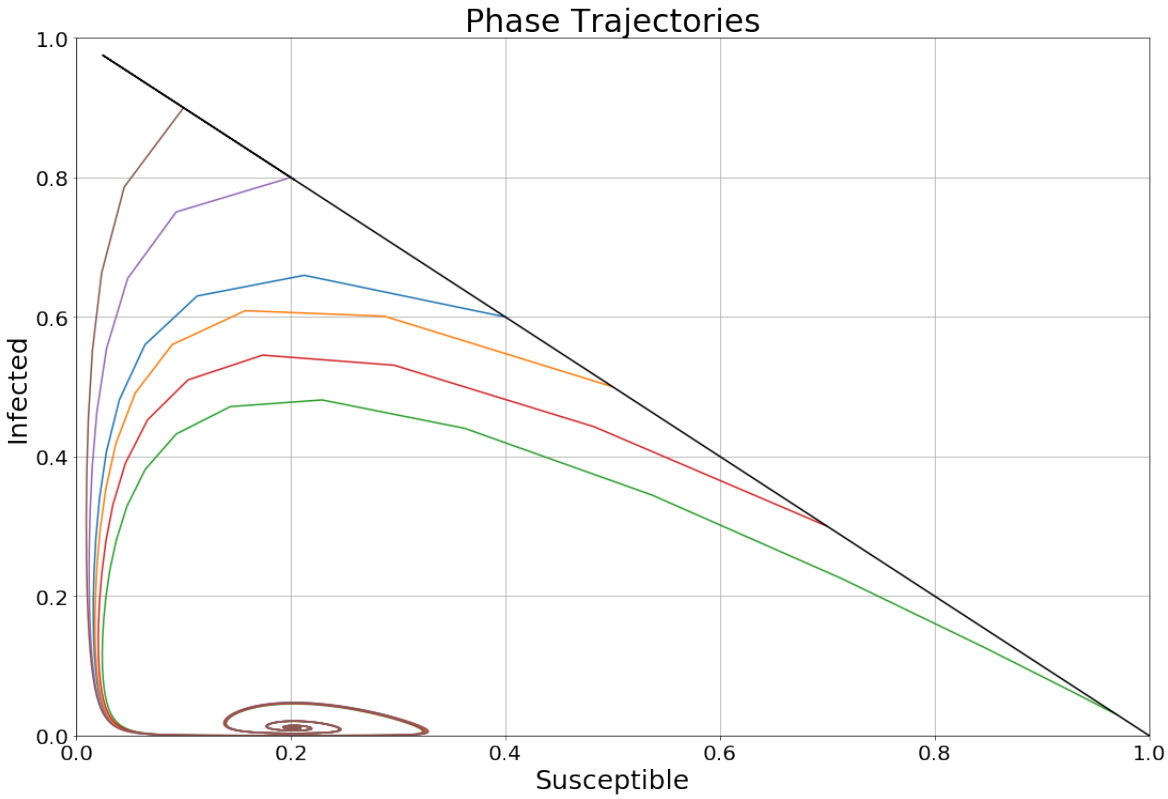
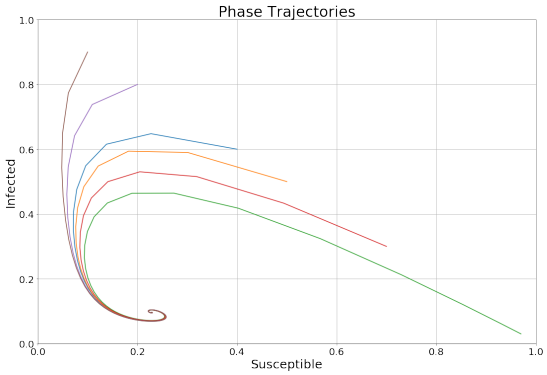
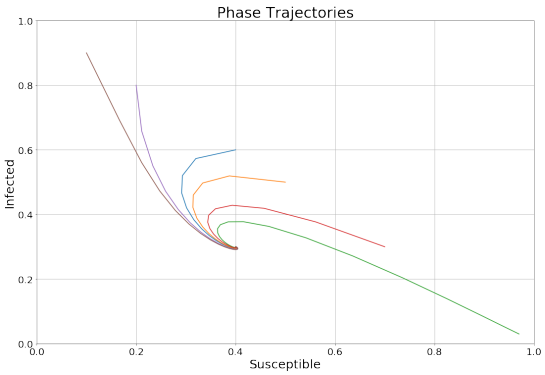


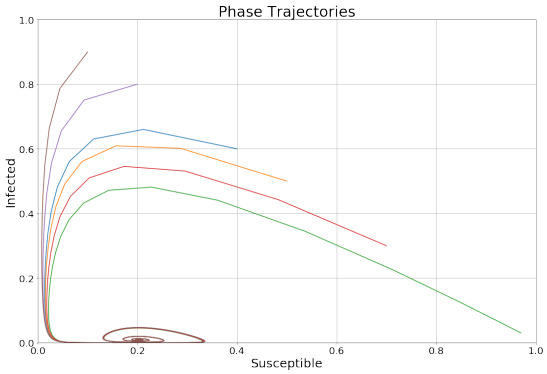
Figure 2.7: Phase trajectories for the SIR model with demography given by equations (1.25), (1.26): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$ with parameters $\beta = 5$, $\gamma = 1$, $\mu = \frac{1}{70}$.



(a) Constant $\mu = \frac{1}{7}$.

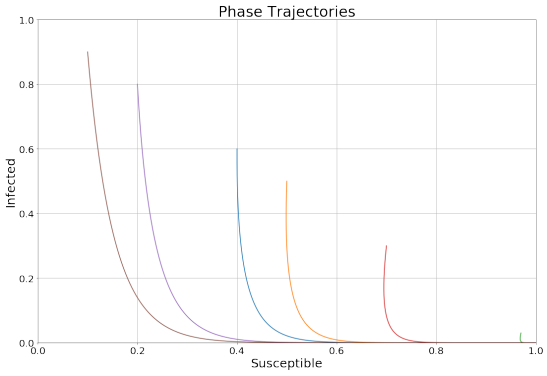


(b) Constant $\mu = 1$.

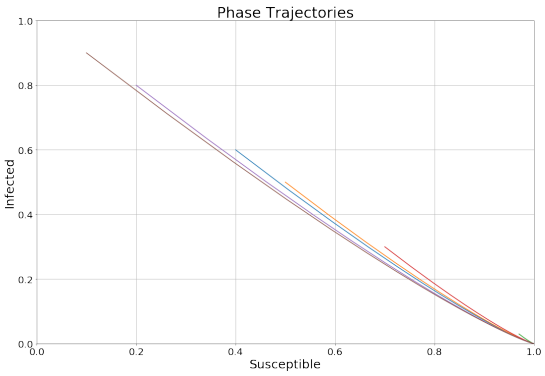


(c) Constant $\mu = \frac{1}{90}$.

Figure 2.8: Phase trajectories given by equations (1.25), (1.26): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$ with different parameter μ for same $\beta = 5$ and $\gamma = 1$.



(a) Constant $\mu = \frac{1}{7}$



(b) Constant $\mu = 1$.

Figure 2.9: Phase trajectories given by equations (1.25), (1.26): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$ with different parameter μ for same $\beta = 0.037$ and $\gamma = 0.18$.

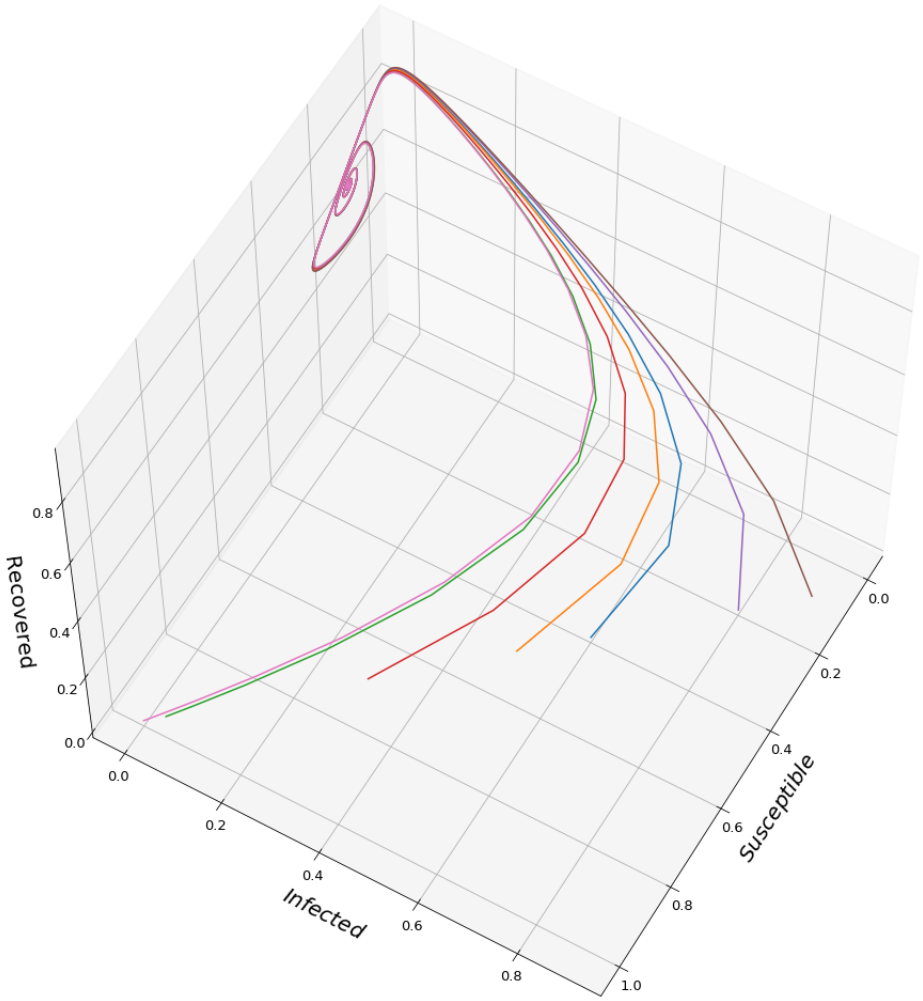
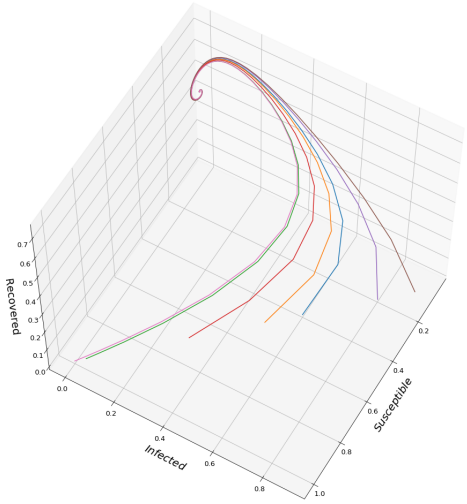
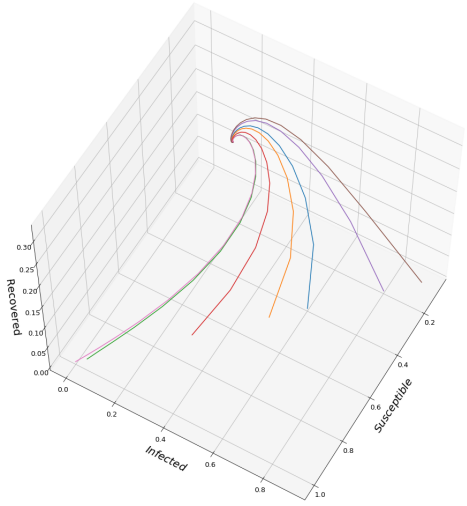


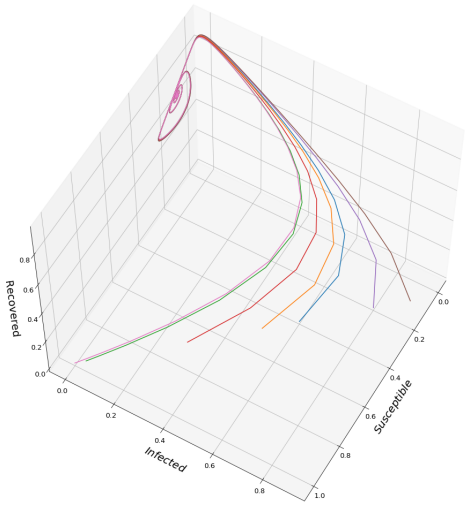
Figure 2.10: Phase trajectories given by equations (1.25), (1.26), (1.27): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$, $\frac{dR}{dt} = \gamma I - \mu R$ with parameters $\beta = 5$, $\gamma = 1$, $\mu = \frac{1}{70}$.



(a) Constant $\mu = \frac{1}{7}$.

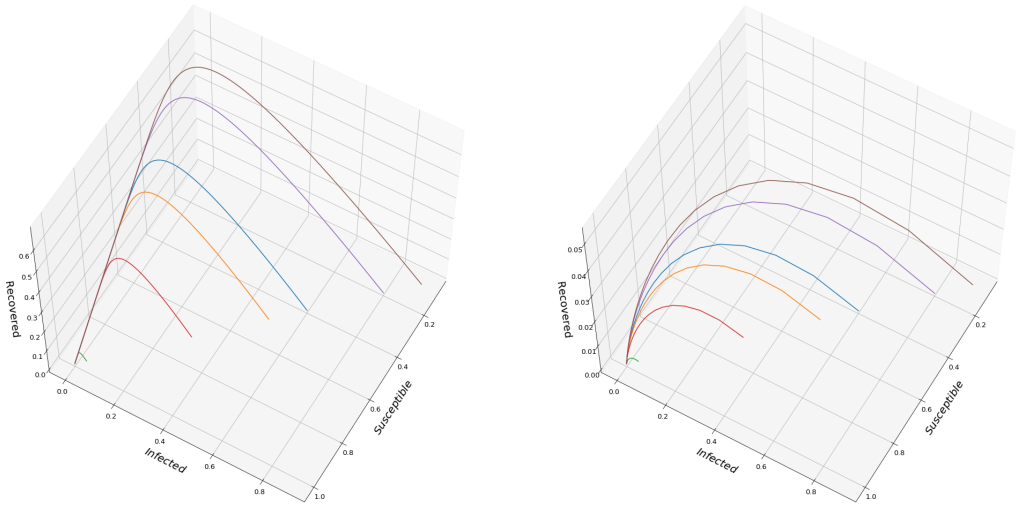


(b) Constant $\mu = 1$.



(c) Constant $\mu = \frac{1}{90}$.

Figure 2.11: Phase trajectories in 3 dimensions given by equations (1.25), (1.26), (1.27): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$, $\frac{dR}{dt} = \gamma I - \mu R$ with different parameter μ for same $\beta = 5$ and $\gamma = 1$.



(a) Constant $\mu = \frac{1}{7}$. (b) Constant $\mu = 1$.

Figure 2.12: Phase trajectories in 3 dimensions given by equations (1.25), (1.26), (1.27): $\frac{dS}{dt} = \mu - \beta S I - \mu S$, $\frac{dI}{dt} = \beta S I - \gamma I - \mu I$, $\frac{dR}{dt} = \gamma I - \mu R$ with different parameter μ for same $\beta = 0.037$ and $\gamma = 0.18$.

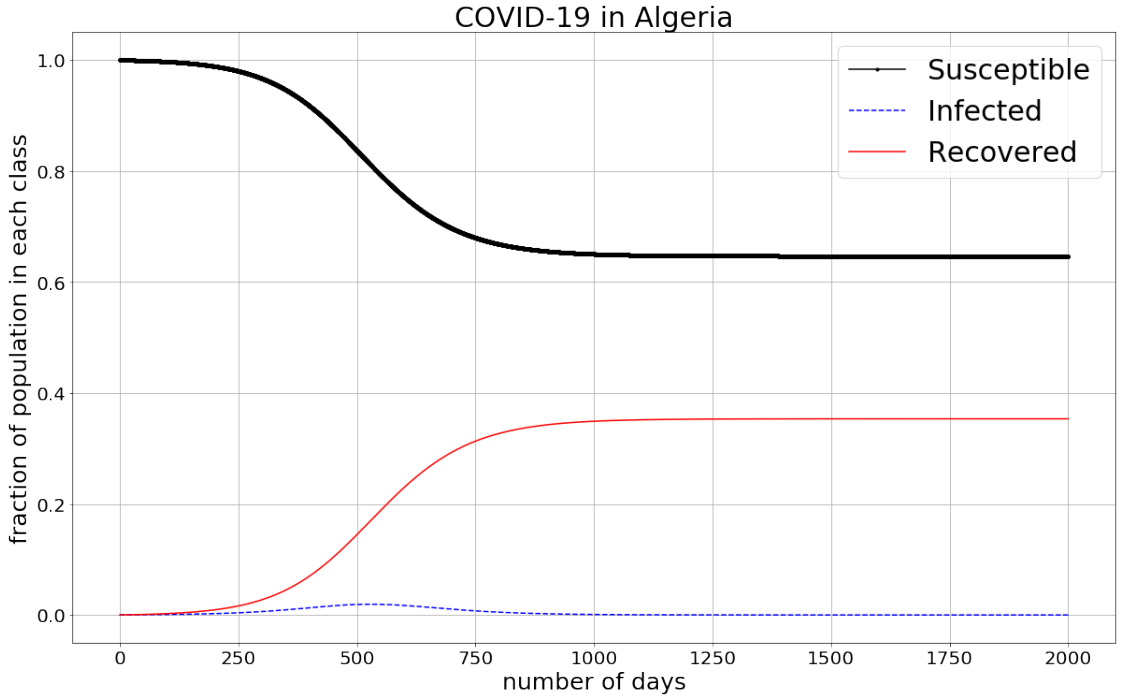


Figure 2.13: The SIR model for epidemic of COVID-19 in Algeria with coefficients $\beta = 0.0561215$ and $\gamma = 0.0455331$ [9].

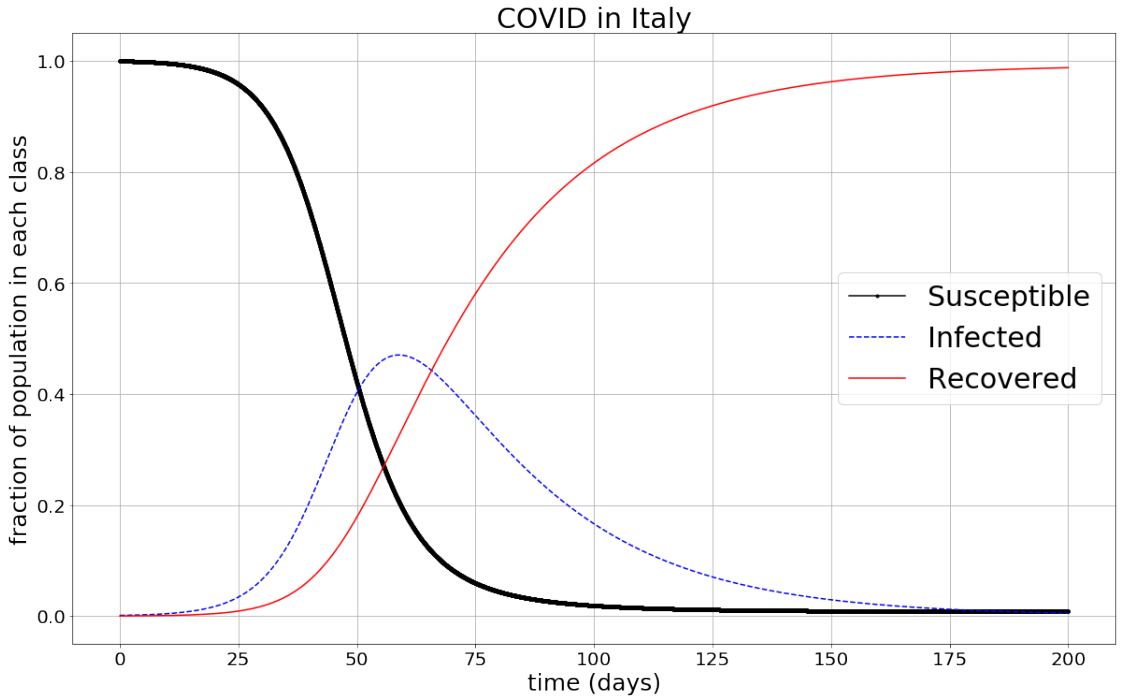


Figure 2.14: The SIR model for epidemic of COVID-19 in Italy with coefficients $\beta = 0.18$ and $\gamma = 0.037$ [10].

Chapter 3

Numerical methods

Consider having a system of non-linear differential equations without analytical solution:

$$\frac{dy}{dt} = f(t, y), \tag{3.1}$$
$$y(t_0) = y_0.$$

Solving (3.1) is only possible with numerical methods. The most basic and simple method to use is Euler's method, where approximation of y_n, solution of y(t) at the point t_n, can be find from recursion [13]

$$y_0 = \eta,$$
$$y_{n+1} = y_n + hf(t_n, y_n), n = 0, 1, ..., m_n \in \langle a, b \rangle,$$

where η is the initial condition, h is a constant (integration step). This method was implemented by myself in Python[15]:

Algorithm 1 Euler's method

Require: $y[0] = t_0$ for vector y
Ensure:
  for i from the beginning to the end of the vector y **do**
    $y[i + 1] \leftarrow y[i] + h * f(t[i], y[i])$
  end for

Euler's method is very dependent on the integration step. Figure 3.1 is comparing graphs for $h = 1$ and $h = 0.1$. The other method used for solving differential equations is Runge-Kutta method. Recursion for the solution is [13]:

$$y_{n+1} = y_n + \frac{1}{6}h * (k_1 + 2 * k_2 + 2 * k_3 + k_4),$$
$$k_1 = f(t_n, y_n),$$
$$k_2 = f(t_n + \frac{1}{2}h, y_n + \frac{1}{2}h * k_1),$$
$$k_3 = f(t_n + \frac{1}{2}h, y_n + \frac{1}{2}h * k_2),$$
$$k_4 = f(t_n + h, y_n + h * k_2),$$
$$n = 0, 1, ..., m_n \in \langle a, b \rangle.$$

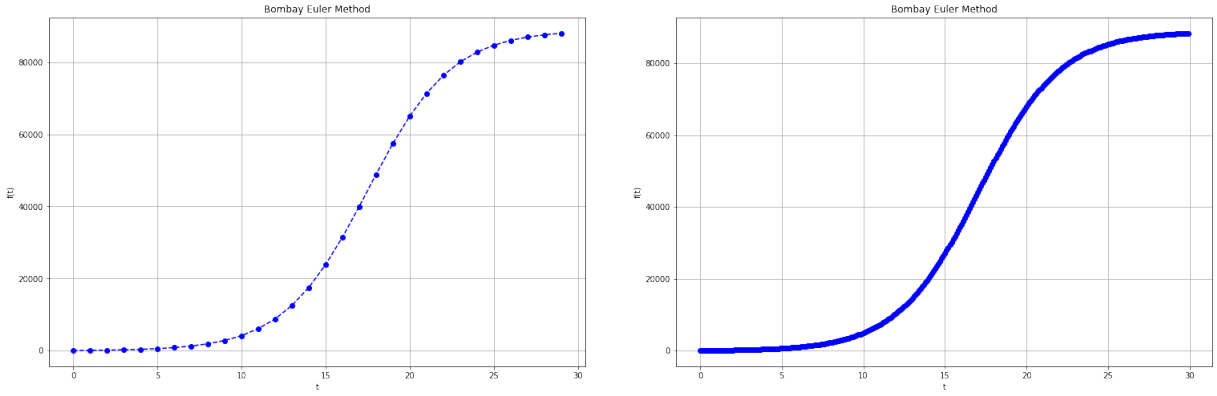


Figure 3.1: Graphical solution for the function $f(t) = \frac{8900}{\cosh^2(0.2t-3.4)}$ with integration step 1 and 0.1.

Algorithm 2 Runge-Kutta method

Require: $y[0] = t_0$ for vector y
Ensure:
for i from the beginning to the end of the vector y **do**
$h \leftarrow t[i+1] - t[i]$
$k1 \leftarrow f(y[i], t[i])$
$k2 \leftarrow f(y[i] + k1 * h/2, t[i] + h/2)$
$k3 \leftarrow f(y[i] + k2 * h/2, t[i] + h/2)$
$k4 \leftarrow f(y[i] + k3 * h, t[i] + h)$
$y[i+1] \leftarrow y[i] + (h/6) * (k1 + 2 * k2 + 2 * k3 + k4)$
end for

In order to choose one of them for solving SIR model, I compared the graphs of the Runge-Kutta solution, the Euler solution and the exact solution. The verification task was the following: I took the following equation:

$$\begin{aligned} \frac{dy}{dt} &= e^{-t}, \\ y(0) &= -1. \end{aligned} \tag{3.2}$$

. Then I applied both methods on this function and plotted them with the exact solution of integration $Y(t) = -e^{-t} - 1$. In Figure 3.2 is seen that the Runge-Kutta method is more accurate than Euler. That is why my choice for numerical method is Runge-Kutta.

One of the important thing to know about a numerical method is how fast it converges to the exact solution. In order to determine it for Runge-Kutta method I calculated the experimental order of convergence (EOC) [13]. To complete this task, firstly, I needed to find the value of error. For comparison I took errors in maximum, absolute value and Euclidean norms. The error is calculated by finding the difference between the numerical and exact solution. If we denote exact solution as y, numerical solution as $\tilde{y}$, interval as T, the number of points of numerical solution as M, then we get the following formulas for calculating errors (here we denote ψ as an error):

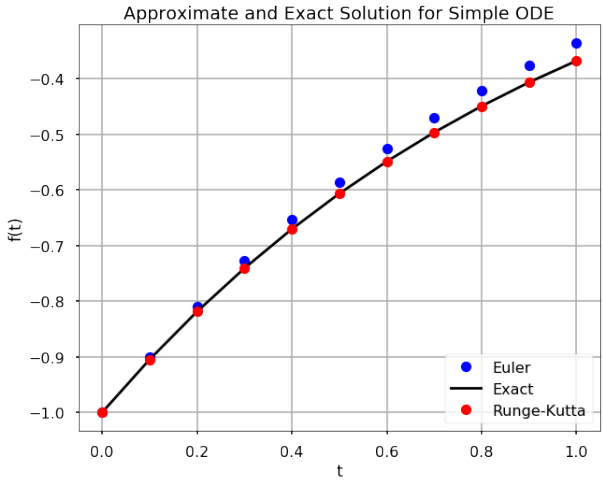


Figure 3.2: Comparison of Runge-Kutta and Euler solutions for the equation (3.2) : $\frac{dy}{dt} = e^{-t}, y(0) = -1$, the exact solution is calculated for $Y(t) = -e^{-t} - 1$.

$$\psi_{max} = \max_{i=1,..,M}(|y_i - \tilde{y}(i\Delta t)|), \tag{3.3}$$

$$\psi_1 = \frac{\sum_{i=1}^{M}|y_i - \tilde{y}(i\Delta t)|\Delta t}{T}, \tag{3.4}$$

$$\psi_2 = \frac{(\sum_{i=1}^{M}(y_i - \tilde{y}(i\Delta t))^2\Delta t)^{\frac{1}{2}}}{T}. \tag{3.5}$$

The (1.22) SIR model does not have the exact solution, so I took numerical solution [23] for more points and replaced the exact solution with it. For numerical solution the number of points was maximum 6000 and for replacement it was 60000. Then I implemented the norms calculations in Python.

Algorithm 3 Error in maximum norm calculation

Require: $errmax = 0$
Ensure:
for k from the beginning to the end of the vectors S, I, R **do**
$a = abs(S(k) - \tilde{S}(k))$
$b = abs(I(k) - \tilde{I}(k))$
$c = abs(R(k) - \tilde{R}(k))$
$errmax = max(max(max(errmax, a), b), c)$
end for

Algorithm 4 Error in discrete L_1 norm calculation

Require: $err = 0$
for k from the beginning to the end of the vectors S, I, R **do**
$a = (abs(S(k) - \tilde{S}(k)) + abs(I(k) - \tilde{I}(k)) + abs(R(k) - \tilde{R}(k))) * \Delta t$
$err = \frac{err+a}{T}$
end for

Algorithm 5 Error in discrete L_2 norm calculation

Require: $err = 0$
  for k from the beginning to the end of the vectors S, I, R **do**
    $a = pow(pow(S(k) - \tilde{S}(k), 2) + pow(I(k) - \tilde{I}(k), 2) + pow(R(k) - \tilde{R}(k), 2), 1/2) * \Delta t$
    $err = \frac{err+a}{T}$
  end for

If $\Delta t = \frac{T}{M}$ then the equation for order of convergence (α) is:

$$\alpha = \frac{\ln\left(\frac{err(\Delta t_{N1})}{err(\Delta t_{N2})}\right)}{\ln\left(\frac{\Delta t_{N1}}{\Delta t_{N2}}\right)}. \tag{3.6}$$

We can replace the expression in denominator: $\frac{\Delta t_{N1}}{\Delta t_{N2}} = \frac{M2}{M1}$. Then the expression (3.7) will become:

$$\alpha = \frac{\ln \frac{err(\Delta t_{N1})}{err(\Delta t_{N2})}}{\ln \frac{M2}{M1}}. \tag{3.7}$$

Now we have two distributions of interval T: M_1 and M_2. To compare the solutions we need to choose one of them and interpolate or extrapolate the other [13]. For accuracy the biggest of them was chosen and the smallest was extrapolated to match the biggest. The extrapolation is the process of filling the missing points with data using available values. In order to simplify calculations I implemented the linear extrapolation using function InterpolatedUnivariateSpline from the Python library scipy.interpolate.

After implementation I calculated EOC for the simple SIR model without demography (1.18, 1.19, 1.20). The results can be found in table below.

Experimental order of convergence			
Number of points	L_{max} norm	L_1 norm	L_2 norm
100	-	-	-
200	2.169864	1.252807	1.253235
300	2.080329	1.152308	1.152739
400	2.036691	1.112664	1.112913
500	2.040279	1.09112	1.091249
600	2.007817	1.077364	1.077429
1000	2.01226	1.223652	1.223831
2000	2.003693	1.339686	1.339751
6000	2.000776	1.648976	1.649005

Table 3.1: Experimental order of convergence for SIR without demography: $\frac{dS}{dt} = -\beta S I$, $\frac{dI}{dt} = \beta S I - \gamma I$, $\frac{dR}{dt} = \gamma I$.

Runge-Kutta method has the order 4. But for the system of non-linear equations, which cannot be analytically solved, we should not expect that EOC will converge to 4. Nevertheless, if for Runge-Kutta the EOC converges to a number bigger than 1, the method is more accurate than Euler method, which order is 1. As we can see in Figure 3.1, EOC for every norm converges to the number greater than 1. Moreover, for L_{max} norm it converges to 2. This fact proves that Runge-Kutta for SIR model is more accurate than Euler method.

Conclusion

The purpose of this work was to learn how mathematical models can be used in epidemioligy. The purpose was reached by studying to SIR models: the basic one and the model with demography. Firstly, some restrictions on the model were established. We decided that for the simplicity of the model we need to focus on acute fast diseases with short incubation period. Our population should be homogenous and recovery from the disease should give a lifelong immunity. Then we denoted some important and widely used parameters in epidemiology, such as the recovery rate γ, the infectivity rate β and the basic reproductive ratio R_0. Next step was discovering the model of three non-linear differential equations for Susceptibles, Infected and Recovered. Then, by dismissing the "fast disease" assumption, we added the natural host lifespan to the model, thus the model with demography was obtained.

The second chapter was devoted to models' properties. We tried and succeeded in calculating the approximate solution for the basic SIR model. We used two different methods to determine that the obtained solution was correct. Unfortunately, we were obligated to add the additional assumption (small Recovered percentage) to obtain it. This result helped us to study the properties of the basic SIR model. We determined the conditions for the beginning of the epidemic, discussed the role of the basic reproductive ratio and then established the conditions for the epidemic to "die out". For the SIR with demography we discovered it's oscillatory dynamic and tried to study the stability of it's solution. We defined stability and asymptotic stability and used the Poincaré-Lyapunov theorem to find if the model with demography is asymptotically stable. We continued with studying the phase trajectories for both models in two and three dimensional spaces. We discussed the influence of the recovery rate, infective rate and reverse to host lifespan on the phase trajectories. The change in the parameters result in appearance of the curves or polygonal chains and in changing size of oscillation.

We ended the chapter by discussing some modern trends, especially behavioral epidemiology, in SIR models and proved that the basic model is still used by researches to conduct an analysis on modern diseases.

In the last chapter we tried to solve SIR model with numerical methods for solving differential equations. We studied two methods: Euler's and Runge-Kutta's and compared them. Comparison showed that Runge-Kutta will be a better choice for solving SIR model. Then we accessed the speed of convergence of the solution found by Runge-Kutta method to the exact solution. For this objective we calculated the experimental order of convergence. Because of non-linear character of the SIR equations, the convergence to the exact solution is slower than expected for Runge-Kutta method.

For the future work, more restriction of the model can be discarded and study on SEIR, SI, SIS models can be executed. Additionally, the detailed research can be conducted on the behavioral SIRs.

Bibliography

[1] J. D. Murray: *Mathematical Biology*. Springer, New York, 2007.

[2] W. O. Kermack, A. G. McKendrick, *A contribution to the mathematical theory of epidemics*. Proceedings of the Royal Society of London, Series A 115, 1927, 700-721.

[3] M. J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. De Gruyter, Princeton, 2011.

[4] H. R. Thieme, *Mathematics in Population Biology. Princeton Series in Theoretical and Computational Biology*, Princeton University Press, 2003.

[5] R. Ragonnet, et al., *Vaccination Programs for Endemic Infections: Modelling Real versus Apparent Impacts of Vaccine and Infection Characteristics*. Sci Rep 5, 15468, 2015.

[6] H. Gion, Y. Saito, S. Yazaki, *On a backward bifurcation of an epidemic model with capacities of treatment and vaccination*. JSIAM Letters (to appear), 2021.

[7] M. A. Nowak, R. M. May, *Virus dynamics. Mathematical principles of immunology and virology*. Oxford University Press, 2000.

[8] J.A.P. Heesterbeek, Odo Diekmann, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Son, 2000.

[9] Mohamed Lounis, Dilip Kumar Bagal, *Estimation of SIR model's parameters of COVID-19 in Algeria*. SpringerOpen, 2020.

[10] Manuel De la Sen, Asier Ibeas, *On an Sir Epidemic Model for the COVID-19 Pandemic and the Logistic Equation*. Hindawi, 2020.

[11] Gabriel G. Katul, Assaad Mrad, Sara Bonetti, Gabriele Manoli, Anthony J. Parolari, *Global convergence of COVID-19 basic reproduction number and estimation from early-time SIR dynamics*. PLOS ONE, 2020.

[12] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp & W.M. Getz, *Superspreading and the effect of individual variation on disease emergence*. Nature 438, 2005, 355-359.

[13] E.Vitásek, *Numerické metody*. Nakladatelství Technické Literatury, 1987.

[14] Ali Y. Rohedi ,*Introducing closed form solution of Riccati differential equation of constant coefficient for solving some model of chemical processes*. AIP Conference Proceedings 2349, 020038 (2021) https://doi.org/10.1063/5.0054521.

[15] Qingkai Kong, Timmy Siauw, Alexandre Bayen, *Python Programming and Numerical Methods: A Guide for Engineers and Scientists*. Academic Press, 2020.

[16] Alberto d'Onofrio, Piero Manfredi, *Behavioral SIR models with incidence-based social-distancing*. Chaos, Solitons & Fractals 159, 2022.

[17] Majee, S., Adak, S., Jana, S. et al., *Complex dynamics of a fractional-order SIR system in the context of COVID-19*. J. Appl. Math. Comput. 68, 4051–4074, 2022.

[18] Zhao-Wei Tong, Yu-Pei Lv, Rahim Ud Din, Ibrahim Mahariq, Gul Rahmat, *Global transmission dynamic of SIR model in the time of SARS-CoV-2*. Results in Physics 25, 2021.

[19] Selma Nihel Klouche-Djedid, Jaffer Shah, Maya Khodor, Salah Eddine Oussama Kacimi, Sheikh Mohammed Shariful Islam, Hani Aiash, *Algeria's response to COVID-19: an ongoing journey*. The Lancet Respiratory Medicine, 2021.

[20] Sileshi Sintayehu Sharbayta, Bruno Buonomo, Alberto d'Onofrio, Tadesse Abdi, *'Period doubling' induced by optimal control in a behavioral SIR epidemic model*. Chaos, Solitons & Fractals 161, 2022.

[21] Helena Sofia Rodrigues, *Application of SIR epidemiological model: new trends*. International journal of applied mathematics and informatics 10, 2016.

[22] Stephen Wiggins, *Introduction to applied nonlinear dynamical systems and chaos*. Springer, New York, 2003.

[23] M. Blasik, M. Klimek, *Exact and numerical solutions of sequential fractional differential equations of order in (1,2)*. Proceedings of the 13th International Carpathian Control Conference (ICCC), 2012.

[24] Alberto Bisin, Andrea Moro, *Learning Epidemiology by Doing: The Empirical Implications of a Spatial-SIR Model with Behavioral Responses*. NBER Working Paper No. 27590, 2020.

[25] Cabrera, M., Córdova-Lepe, F., Gutiérrez-Jara, J.P. et al., *An SIR-type epidemiological model that integrates social distancing as a dynamic law based on point prevalence and socio-behavioral factors*. Sci Rep 11, 2021.