Julius-Maximilians-Universität Würzburg
Institut für Mathematik

Lehrstuhl für
Dynamische Systeme und Kontrolltheorie

Stability of Switched Epidemiological Models

Stabilität geschalteter epidemiologischer Modelle

Masterarbeit

von
Sebastian Pröll

Julius-Maximilians-Universität Würzburg
Institut für Mathematik

Lehrstuhl für
Dynamische Systeme und Kontrolltheorie

Stability of Switched Epidemiological Models

Stabilität geschalteter epidemiologischer Modelle

Masterarbeit

eingereicht
im
Oktober 2013
von
Sebastian Pröll
Matr.-Nr. 1593307

betreut durch
Prof. Dr. Fabian Wirth

Zusammenfassung

In der vorliegenden Arbeit werden Möglichkeiten aufgezeigt, wie man die Ausbreitung von Infektionskrankheiten mit Hilfe von mathematischen Modellen beschreiben kann. Anhand solcher Modelle möchte man mehr über die Dynamik von Epidemien lernen und vorhersagen, wie sich eine gegebene Infektionskrankheit innerhalb einer Population ausbreitet.
Zunächst werden gewöhnliche Differentialgleichungen verwendet, um grundlegende epidemiologische Modelle aufzustellen. Hierbei unterscheidet man sogenannte SIR und SIS Modelle, je nachdem ob die betrachtete Krankheit einem Individuum nach seiner Heilung Immunität verleiht oder nicht. Charakteristisch für Infektionskrankheiten sind Parameter wie die Infektionsrate oder die Heilungsrate. Sie geben an, wie ansteckend eine Krankheit ist bzw. wie schnell eine Person nach einer Erkrankung wieder gesund wird. Im Allgemeinen sind diese Parameter abhängig von bestimmten Bevölkerungsgruppen und verändern sich mit der Zeit. Daher werden am Ende des zweiten Kapitels Modelle entwickelt, die die Betrachtung mehrerer Bevölkerungsgruppen zulassen. Zeitvariante Parameter werden durch die Verwendung geschalteter Systeme berücksichtigt.
Bei der Untersuchung solcher Systeme ist derjenige Zustand von besonderem Interesse, bei dem innerhalb der Bevölkerung keine Infizierten auftreten, die gesamte Bevölkerung also von der betrachteten Krankheit frei bleibt. Es stellt sich die Frage, unter welchen Bedingungen sich dieser Zustand nach einer Infizierung der Bevölkerung im Laufe der Zeit von selbst einstellt. Mathematisch gesehen untersucht man die triviale Ruhelage des Systems, bei der keine Infizierten existieren, auf Stabilität.
Für die Stabilitätsanalyse sind einige mathematische Begriffe und Aussagen notwendig, die im zweiten Kapitel bereitgestellt werden. Grundlegend ist die Theorie gewöhnlicher Differentialgleichungen, einschließlich der Stabilitätstheorie von Lyapunov. Darüberhinaus kommen wichtige Erkenntnisse aus den Gebieten Konvexe und Nichtglatte Analysis, Positive Systeme und Differentialinklusionen.
Ausgestattet mit diesen Hilfsmitteln werden im vierten Kapitel Sätze bewiesen, die hinreichende Bedingungen dafür angegeben, dass die triviale Ruhelage in geschalteten SIS, SIR und SIRS Systemen asymptotisch stabil ist.

Contents

1 Introduction

Infectious diseases seem to be related to mankind ever. In 430 BC, typhoid fever killed a quarter of the population of Athens within four years. The number of casualties caused by Black Death during the 14th century is estimated at 20 to 30 million Europeans in six years. Until the 18th century, more than 100 plague epidemics swept across Europe. On the American continent, the Native people had to suffer from smallpox, measles and influenza European conquerors brought to the New World. Smallpox killed 150 thousand people only in the Aztec city-state Tenochtitlan, located in today's Mexico, in the 1520s. The Third Pandemic starting in China in the 19th century killed over 12 million people in China and India. The Spanish flu in 1918-19 became a worldwide pandemic which killed some 50 million people.

For a long time the mechanisms of spread of diseases were unknown. In the middle of the 16th century, a doctor from Verona named Girolamo Fracastero was first to propose a theory about contagion that postulated spread from person to person. It took until the 19th century when Louis Pasteur proved that certain diseases are caused by infectious agents and Robert Koch formulated criteria to determine an infectious disease. This new microbiological knowledge was seminal to control epidemics. After Dr. John Snow – one of the fathers of modern epidemiology – discovered that cholera was transmitted by the fecal contamination of water, first steps were made to establish hygiene in western cities. Sewers were built and private bathrooms and flush toilets became popular. Vaccination is another important way to defeat communicable diseases. Several successful vaccines were introduced in the twentieth century, including those against diphtheria, measles, mumps, and rubella. Major achievements included the development of the polio vaccine in the 1950s and the eradication of smallpox during the 1960s and 1970s.

Still, there remain challenging problems for today's civilization. According to the WHO, more than one fourth of all deaths worldwide were caused by infectious diseases in 2002. To date there are no vaccines available against HIV and malaria. The growing world population and highly developed infrastructure are ideal conditions for proliferation of pathogens. Industrial livestock farming, including the massive use of antibiotics, induces the generation of drug-resistant microbes. Some widespread animal diseases in the past two decades were mad cow disease (BSE), foot-and-mouth disease, bird flu and swine flu.

All this keeps specialists from different fields busy and they are interested in different aspects of diseases. Medical doctors and veterinary clinicians primarily wish to know how to treat human patients or animals and are therefore most concernced about the infection's pathophysiology (i.e. how it affects the organism) or clinical symptoms. Microbiologists, on the other hand, focus on the natural history of the causative organism and ask about the etiological agent (a virus, bacterium, protozoan, fungus, or prion) as well as about the optimal conditions for its growth. Finally, epidemiologists are most interested in features which determine patterns of disease and its transmission.

The main reason for studying infectious diseases is to prevent an outbreak, respectively to improve control and ultimately to eradicate the infection from the population. Several forms of control measure exist such as vaccination or quarantine. The former measure prevents susceptible individuals within a population from getting infected by a pathogen whereas the latter one tries to

isolate already infected individuals such that they cannot spread the pathogen and infect further persons.
However, resources are limited and not all measures are always appropriate or available. In case of an outbreak of a previously unknown disease, vaccines need time to be developed, produced and distributed. This could take too much time to protect the population from being infected. For some diseases vaccines do not exist. And sometimes, even if they do, side effects cause people not to accept vaccines. Medical isolation of infected persons are a logistical challenge. It can be established for a small group of patients, but not for large areas. In consequence, it is of great importance to deploy any measure as efficiently as possible and to optimize the use of resources. Detailed knowledge about how infectious disease agents behave and how diseases spread is needed to support the decision-making process. Here, mathematical models which describe the evolution of an epidemic in time can help in two ways. On one hand they enhance *insight* to the dynamics of a disease. On the other hand they can *predict* its progress, provided that the model is carefully adapted to the respective circumstances.
The systematic investigation of spread of epidemics using mathematical language began around hundred years ago. The models described in the article [22] of Kermack and McKendrick in 1927 have become standard in mathematical epidemiology. A landmark is also the book of Bailey [7], published in 1957. If someone is interested in the field, the paper [17] from Hethcote could serve as a starting point. It contains a comprehensive list of references for further study. A highlight is the book [21] of Keeling and Rohani, since it deals carefully with real-world problems and does not regard mathematics isolated but puts it into a wider context. Some ideas mentioned before and in the following are taken from this book.
Mathematical models have to fulfill many requirements that often conflict with each other. Models should be *transparent* to be accessible to analytical and numerical tools and to bring a better understanding. They should be *flexible* to be applied to diverse situations. And finally, they should be *accurate* to provide a reliable prediction. Anyone who starts with a modeling process is advised to keep the following words by Keeling and Rohani [21, Chapter 1.4] in mind.

> By definition, all models are “wrong,” in the sense that even the most complex will make some simplifying assumptions. It is, therefore, difficult to express definitely which model is “right,” though naturally we are interested in developing models that capture the essential features of a system. Ultimately, we are faced with a rather subjective measure of the *usefulness* of any model.

We shall put these thoughts into the context of communicable diseases. There are numerous elements of chance that influence the spread of the disease. It is unpredictable how many contacts each individual has at each day, which of these contacts are sufficient to transmit pathogens and how each individual’s immune system reacts to the invasion of pathogens. Some people get sick after having had contact to an infected person, others don’t. Other highly variable parameters like temperature, wind, climate have an effect on the activity of a pathogen. Thus, we “will never be able to predict the precise course of an epidemic, or which people will be infected. The best that we can hope for is

models that provide confidence intervals on the epidemic behavior and determine the risk of infection for various groups of hosts."[1]
Now, since we are aware of the limitations that arise while modeling infectious diseases, we shall emphasize the benefits that come along with their theoretical treatment. Nowadays we have access to extensive data related to communicable diseases. The WHO publishes the World Health Statistics report[2] every year that contains the number of cases of the most relevant diseases in a number of countries. The national health authorities themselves collect data and cooperate with the WHO to improve statistics and to specify characteristics of certain infectious diseases. Mathematical models profit from these statistics in two ways. First, the data helps to determine key numbers like the rate of infectivity or the rate of recovery that are essential when applying models to a present disease. Without these parameters, models would remain useless. Secondly, statistics that trace the progress of a disease are the only way to validate a model. If the number of infective individuals within a population predicted by a mathematical model during time coincides with the actual numbers given by a statistic, then it can be said that the model is reliable (at least for the specific situation). In conclusion, statistical data make mathematical models be an authentic instrument in the prediction of epidemic spread.
A major advantage of epidemiological models is also the possibility to pass through an epidemic virtually. With modern computers it is possible to solve differential equations numerically and hence simulate and visualize even high complex behaviors. By changing the parameters it may be studied how they effect progress of an epidemic. This is highly valuable since such experiments cannot be conducted in reality and it increases insight to epidemiological phenomena.
The purpose of this thesis is to show how modern mathematical techniques can be deployed to analyze epidemiological models. In Chapter 2, we introduce some standard models of mathematical epidemiology and improve them in several steps. Thus they hopefully are flexible enough and generate the desired accuracy to fit practical situations. In Chapter 3, we present the mathematical background that is needed to investigate the epidemiological models. The notions and theorems are interesting for themselves and we take our time to discuss them carefully. In Chapter 4 we present the main result of the thesis and proof certain stability properties of epidemiological models.

[1] [21, Chapter 1.6]
[2] http://www.who.int/gho/publications/world_health_statistics/en/index.html

2 Epidemiological Models

The progress of an infectious disease is defined qualitatively in terms of the level of pathogen within the host. As a host we denominate an individual that is able to carry a microparasite, be it a human, an animal or a plant. Initially the host is free of the pathogen but able to get infected by it, thus the host is *susceptible*. At the time of infection the host is exposed to some infectious material, maybe by direct physical contact with an infectious individual, breathing in infectious organisms, touching contaminated surfaces, etc. The following stage is full of uncertainties. Whether the pathogen succeeds in settling within the host or not depends on the amount of microparasites incorporated and the ability of the host's immune system to fight the invader. Assume the pathogen gained a foothold and grows latently within the host over time but is not yet emitted to the environment. The host is then in the *exposed* phase. When the pathogen has reached a level where it is transmitted, the host is *infectious*. Finally, once the individual's immune system has defeated the parasite and is no longer infectious, the host is referred to as *recovered*. The different stages of infection are illustrated in Figure 1 which is taken from the book [21, p. 4].
This fundamental classification as susceptible (S), exposed (E), infectious (I), or recovered (R) depends only on the host's ability to transmit the pathogen. It is irrelevant whether the host is showing symptoms. An individual who feels perfectly healthy can excrete large amounts of pathogen. Conversely, a patient who feels sick might already have eliminated the parasite. This situation is reflected in Figure 1 by the overlapping of incubation and diseased period with exposed, infectious and recovered period. The incubation period is the time elapsing between the receipt of infection and the appearance of symptoms. The subsequent diseased period is the time where the patient is showing symptoms and feels sick.

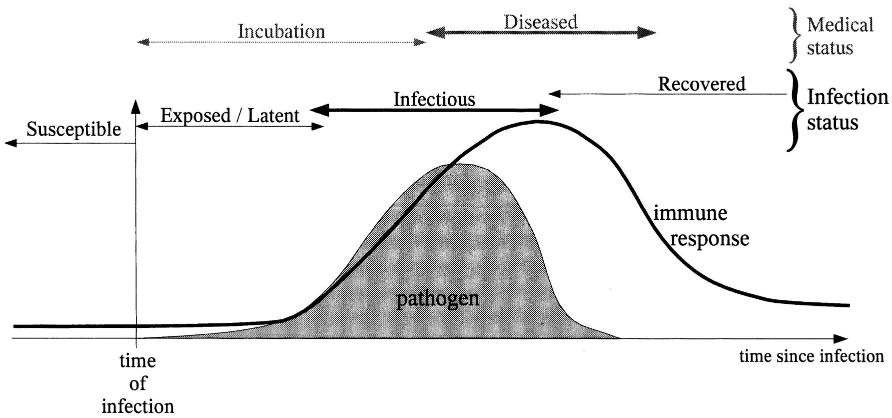

Figure 1: A sketch of the progress of infection, showing the dynamics of the pathogen (gray area) and the host immune response (black line) as well as labeling the various infection classes. Note that the medical status does not necessarily coincide with the infection status.

2.1 An SIR Model

We will now consider a whole population. Each individual belongs at each time to exactly one of these four categories S, E, I, R, depending on its infection status. The basic question is, how does the number of infective individuals $I(t)$ evolve in time?
Let $S(t)$ be the number of susceptibles at time t and N the total population size, which we assume to be constant over time. We suppose further that infectives and susceptibles are homogeneously mixing. Let β be the average number of contacts of one person per unit time, which are sufficient for transmission. β is called *contact rate* or *infection rate*. Each person may have many more contacts, but only those contacts are relevant that distribute the disease. Then the number of new infectives in a small time interval dt around t is given by

$$\beta \frac{I(t)S(t)}{N} dt. \tag{2.1}$$

This relation has proved to be reasonable (see [17, Chapter 2.1]) and can be justified by the following consideration. If one person meets β other persons in a certain time interval, maybe one day, then – due to a homogeneous density of infected individuals within the population – the person meets approximately $\beta \cdot I(t)/N$ infectives the day. As the number of infectives changes over time, the above number will only be valid over short time intervals dt. Since the number of contacts is supposed to be equal for each susceptible, the number of contacts of *all* susceptibles with infectives within a small interval dt is given by $S(t)\beta\frac{I(t)}{N}dt$. This gives the above relation.
As shown in Figure 1, each infected person recovers after a certain time. We assume that the average infectious period of an individual is $1/\gamma$. Thus over a small time dt,

$$\gamma I(t)dt \tag{2.2}$$

infectives recover from the disease. The number γ is called *recovery rate*. A combination of (2.1) and (2.2) yields the differential equation

$$\dot{I}(t) = \frac{dI(t)}{dt} = \beta \frac{I(t)S(t)}{N} - \gamma I(t) \tag{2.3}$$

which describes the evolution of infectives in our simple model. It will accompany us for the rest of the thesis. At (2.3) we see that the exposed class does not influence the interaction between susceptibles and infectives. We will therefore omit this class in our model and only consider the remaining compartments S, I and R. Each individual will be classified inter one of the three categories in the following. The mechanism shown in Figure 1 tells us that susceptibles that have been infected are moving to the class I and infectives that are cured from the disease are moving to the class R and remain there for the rest of the time, since they have acquired permanent immunity. This flow pattern is illustrated in Figure 2. Acronyms for epidemiological models are often based on the flow patterns between the compartments they incorporate, such that we will call the model that characterizes the flow pattern of Figure 2 an SIR model.
Differential equations for $S(t)$ and $R(t)$ are immediately achieved by the one for $I(t)$ and the flow pattern. If the number of infectives increases by the value

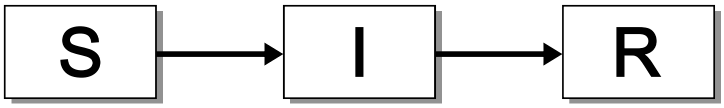

Figure 2: flow pattern of an SIR model

given in (2.1), then the number of susceptibles has to decrease by the same value. Hence

$$\dot{S}(t) = \frac{dS(t)}{dt} = -\beta \frac{I(t)S(t)}{N}.$$

The same argument holds for $R(t)$. This number increases by $\gamma I(t)dt$ as the number of infectives decreases over a given interval dt. Thus

$$\dot{R}(t) = \frac{dR(t)}{dt} = \gamma I(t).$$

Our first basic epidemiological model is now fully described by the three preceding equations. Such a model is called *deterministic*, because if we know β, γ and the initial numbers $S(0)$, $I(0)$, $R(0)$ of susceptible, infective and recovered individuals at the beginning of an epidemic, then the functions $S(\cdot)$, $I(\cdot)$, $R(\cdot)$ are determined for every $t \geq 0$. Mathematically spoken, the initial value problem has a unique solution. We will revisit the correspondent result in Section 3.5.1.
Up to now we have made many assumptions to build our model. We assumed that the parameters β and γ do not change over time, that the population size is constant and that infectives are homogeneously mixing with other individuals. In reality, this will not be the case. As already said in the introduction, the spread will rather be dependent on multiple random events. These stochastic elements should be considered in a model that claims to be realistic. As Bailey says in his book [7, Chapter 4]:

> Any mathematical picture of the behaviour of an epidemic that attempts to be at all realistic in detail, e. g. specifying the number of new cases that will occur in a given short interval of time, must inevitably involve the use of probability concepts.

Nevertheless, we will hold on to deterministic models. For a sufficiently large population size this is reasonable, since the relative magnitude of stochastic fluctuations reduces as the number of cases increases. Moreover, we are only at the beginning of our modeling process. Several improvements will be done in the following to make models more realistic.

2.2 An SIR Model with Demography

The model just introduced is suitable for diseases that pass through a population in a rather short period of time. If we are interested in the long-term behavior of a disease, demographic processes should be respected. We will therefore extend our equations by a term that considers births and deaths.
The number of children born in our population in a unit time interval is given by the product μN of a parameter μ and the population size N. All newborns are susceptible to the disease by assumption, such that they join the susceptible

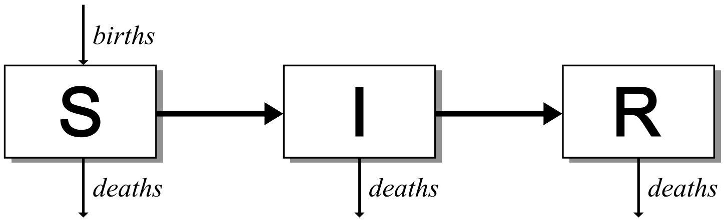

Figure 3: flow pattern of an SIR model with demography

class in our model, see Figure 3. If N is constant in time, then there have to be as many deaths as births per time. People die at any stage of the infection process so that the quantities S, I, R reduce by the numbers μS, μI, μR, respectively. The parameter μ is called *birth/death rate*. The inverse $1/\mu$ has the meaning of an average life span of individuals. Note that the deaths reflect individuals reaching their average life span and are not caused by the disease. Up to now we assumed that there is no mortality risk and everybody gets healthy again after an infection. However, some infections in real life are not that harmless and individuals have a significant risk of dying. An SIR model with mortality has been established in [21, Chapter 2.2]. There, the authors introduce a probability that reflects the mortality risk. The problem with mortality is that infection actively removes individuals from the population. It is therefore no longer admissible to assume a constant population size. We shall exclude this difficulty and suppose that no or at most a tiny percentage dies of the disease so that a constant population size is still reasonable. The SIR model with demography is now given by the following differential equations.

$$\begin{aligned}
\dot{S}(t) &= \mu N - \mu S(t) - \beta \frac{S(t)I(t)}{N}, \qquad && S(0) = S_0, \\
\dot{I}(t) &= \beta \frac{S(t)I(t)}{N} - \gamma I(t) - \mu I(t), \qquad && I(0) = I_0, \\
\dot{R}(t) &= \gamma I(t) - \mu R(t), \qquad && R(0) = R_0.
\end{aligned} \tag{2.4}$$

We introduce other types of epidemiological models in the next section. They all contain terms of demographic dynamics. Whenever the parameter μ appears in the equations, births and deaths are considered in the model.
Some models in mathematical epidemiology consider another class, denoted by M, such that they have the flow pattern M→S→(E)→I→R. This additional class accomodates the following situation. If a mother has been infected, then some antibodies are transferred to her unborn infant. The antibodies remain in the newborn for a certain time such that the infant has temporary immunity to an infection. The class M contains these infants with passive immunity. After the maternal antibodies disappear from the body, the infant moves to the susceptible class S. Infants who do not have any passive immunity, because their mothers were never infected, enter directly to the class S. We will not pursue this detailed point of view.
In Section 2.1 we asked how we could describe the evolution of an infectious disease mathematically. A simple answer has now been given by the differential equations (2.4). The subsequent question is, if we know the parameters β, γ, μ

and the initial values S_0, I_0, R_0, how will $I(t)$ behave in the long run? The good news is that the system (2.4) is already well understood and we are able to present a comprehensive answer to this question.
For this, we modify the above equations a bit. At first, note that since $R(t) = N - S(t) - I(t)$ holds for all t, we can drop the equation for the recovered individuals and only consider the remaining two equations. Dividing the latter by N and setting $s(t) := S(t)/N$ and $i(t) := I(t)/N$, we receive

$$\begin{aligned} \frac{ds(t)}{dt} &= \mu - \mu s(t) - \beta s(t)i(t), \qquad s(0) = s_0, \\ \frac{di(t)}{dt} &= \beta s(t)i(t) - (\gamma + \mu)i(t), \qquad i(0) = i_0. \end{aligned} \tag{2.5}$$

By definition, the components $s(t)$ and $i(t)$ of the solution to problem (2.5) should be less or equal than 1 for all times t. In other words, if we consider

$$\Sigma := \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0 \text{ and } x_1 + x_2 \leq 1\}$$

then any solution with $(s_0, i_0) \in \Sigma$ should remain in Σ for all $t \geq 0$. This is fortunately the case and therefore (2.5) is a well posed problem, see [17, p. 608] and references therein. We are now able to give the following theorem which can be found in [17, Theorem 2.2].

Theorem 2.1. *Let $(s(t), i(t))$ be a solution to (2.5) in Σ. Define $\sigma := \beta/(\gamma+\mu)$.*

(i) If $\sigma \leq 1$ or $i_0 = 0$, then solution paths starting in Σ approach the so called disease-free equilibrium *given by $(s, i) = (1, 0)$, i. e.*

$$\lim_{t \to \infty} (s(t), i(t)) = (1, 0).$$

(ii) If $\sigma > 1$, then all solution paths with $i_0 > 0$ approach the so called endemic equilibrium *given by $(s_e, i_e) = (1/\sigma, \mu(\sigma - 1)/\beta)$.*

In Figure 4 we see how the progress of an infection could evolve. The plot shows the fractions $S(t)/N$, $I(t)/N$, $R(t)/N$ over a short time period of 30 days where parameters were chosen to be $\beta = 2$, $\gamma = 0.15$, $\mu = 0.000035$ and initial values $s_0 = 0.9$, $i_0 = 0.1$.

2.3 SIS and SIRS Epidemiological Models

SIR models fit diseases that confer lifelong immunity once an infective has been recovered. Examples that match this condition usually are measles, mumps, rubella, chicken pox or poliomyelitis. Others, for instance sexually transmitted infections like gonorrhea, syphilis or herpes, do not confer immunity. That is, an infected person who has been recovered can again be infected by the same disease. The theoretical consequence of this behavior is that there exists no compartment R and infectives that have been

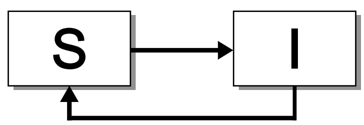

Figure 5: SIS flow pattern

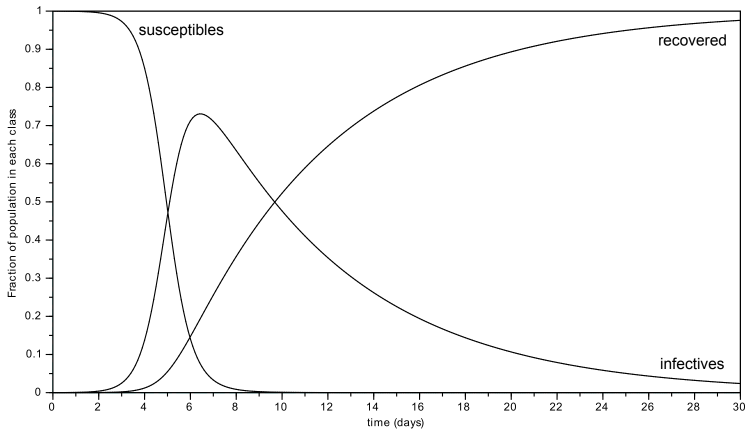

Figure 4: The plot shows the fractions $S(t)/N$, $I(t)/N$ and $R(t)/N$ as functions of time.

recovered join again the S compartment, as shown in Figure 5. Differential equations for the SIS model would be

$$\dot{S} = \mu N - \mu S - \beta \frac{SI}{N} + \gamma I,$$
$$\dot{I} = \beta \frac{SI}{N} - \gamma I - \mu I.$$

Here and in future equations we omit the argument t. It should be clear now that S, I and R, if existent, are functions of time.
The SIR and SIS models represent behavioral extremes where immunity is either lifelong or does not exist at all. One could imagine that there exists an intermediate behavior where immunity lasts for a limited period before waning such that an individual joins again the susceptible class. For this, we would have to introduce a parameter ω that represents the rate at which immunity vanishes. The inverse $1/\omega$ would then signify the average period where a once infected individual stays immune. An SIRS flow pattern is shown in Figure 6 and the corresponding differential equations would be

$$\dot{S} = \mu N - \mu S - \beta \frac{SI}{N} + \omega R,$$
$$\dot{I} = \beta \frac{SI}{N} - \gamma I - \mu I,$$
$$\dot{R} = \gamma I - \mu R - \omega R.$$

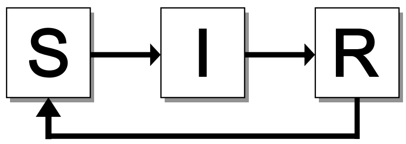

Figure 6: SIRS flow pattern

2.4 Models with Multiple Subgroups

In the previous epidemiological models we assumed that the rate at which individuals get infected, recover and die are equal all over the whole population. Especially for the risk of infection this will not hold in reality. Think of sexually transmitted diseases. The risk for a prostitute to get infected is surely higher than for some other population groups due to a higher contact rate. It is therefore desired to separate high and low risk groups. Mathematically this means that we split the population of size N into several subgroups of size N_i, $i = 1, \dots, n$, with $N_1 + \cdots + N_n = N$. We will develop a multigroup model for the SIR type. It should be clear how to adapt the idea for SIS and SIRS models. As before we assume subgroups to have constant size, i. e. $N_i = S_i(t) + I_i(t) + R_i(t)$ for all $i = 1, \dots, n$. Each subgroup obtains now its own parameters γ_i and μ_i, that reflect individual recovery and birth/death rates. The situation for the contact rate is a little more intricate. Susceptibles from the subgroup i can be infected by individuals from the own subgroup as well as from all other subgroups. Consequently, we have to introduce n^2 parameters β_{ij}, $i, j = 1, \dots, n$, where β_{ij} is the rate at which susceptibles of group i get infected by infectives of group j. Hence we receive the differential equations

$$\begin{aligned} \dot{S}_i &= \mu_i N_i - \mu_i S_i - \sum_{j=1}^{n} \beta_{ij} \frac{S_i I_j}{N_i}, \\ \dot{I}_i &= \sum_{j=1}^{n} \beta_{ij} \frac{S_i I_j}{N_i} - \gamma_i I_i - \mu_i I_i, \\ \dot{R}_i &= \gamma_i I_i - \mu_i R_i. \end{aligned} \tag{2.6}$$

For the remaining part of the text, γ_i and μ_i will always be positive for all $i \in \{1, \dots, n\}$ and all the β_{ij} are nonnegative. Since we have constant population size in each subgroup, we can describe the susceptibles via the equation $S_i(t) = N_i - I_i(t) - R_i(t)$. This allows us to drop the differential equations for the susceptibles S_i and only consider the infectives and recovered in each subgroup.

$$\begin{aligned} \dot{I}_i &= \sum_{j=1}^{n} \beta_{ij} \frac{(N_i - I_i - R_i) I_j}{N_i} - (\gamma_i + \mu_i) I_i, \\ \dot{R}_i &= \gamma_i I_i - \mu_i R_i. \end{aligned} \tag{2.7}$$

It is convenient to divide equations (2.7) by N_i and define new state variables

$$x_i := \frac{I_i(t)}{N_i}, \qquad x_{n+i} := \frac{R_i(t)}{N_i}, \quad i = 1, \dots, n.$$

The vector $x \in \mathbb{R}^{2n}$ consists of the 'infective' components $i = 1, \dots, n$ and the 'recovered' components $i = n + 1, \dots, 2n$. For an arbitrary vector $x \in \mathbb{R}^{2n}$, it will be very useful to set

$$x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}, \text{ where } x^1 := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, x^2 := \begin{pmatrix} x_{n+1} \\ \vdots \\ x_{2n} \end{pmatrix} \in \mathbb{R}^n.$$

Moreover, $\alpha_i := \gamma_i + \mu_i$ and $b_{ij} := \beta_{ij} N_j / N_i$. Our model gets $2n$-dimensional and has the shape

$$\begin{aligned} \dot{x}_i &= (1 - x_i - x_{n+i}) \sum_{j=1}^{n} b_{ij} x_j \ - \alpha_i x_i, \\ \dot{x}_{n+i} &= \gamma_i x_i - \mu_i x_{n+i}. \end{aligned} \qquad i = 1, \dots, n. \tag{2.8}$$

These $2n$ equations are easier to handle if we formulate them in matrix form. Therefore we set

$$\Gamma := \begin{pmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_n \end{pmatrix}, \quad M := \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{pmatrix}, \quad B := \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix}$$

and as a shorthand $D := \Gamma + M$. Here and in the following, for an arbitrary vector $z \in \mathbb{R}^n$, we define

$$\operatorname{diag}(z) := \begin{pmatrix} z_1 & & 0 \\ & \ddots & \\ 0 & & z_n \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Then the multigroup SIR disease model becomes

$$\dot{x} = \begin{pmatrix} -D + B & 0 \\ \Gamma & -M \end{pmatrix} x \ + \ \begin{pmatrix} -\operatorname{diag}(x^1 + x^2) B x^1 \\ 0 \end{pmatrix} =: f(x). \tag{2.9}$$

2.5 Switched Epidemiological Models

Imagine the parameters which characterize the disease of interest are not temporally constant but change in time. This holds for example for an influenza epidemic. The influenza virus is most active during the winter months when it is cold. In contrast, it is not heat resistant such that it disappears in the summer months.[3] The result would be a lower contact rate in the summer than in the winter.

We consider different conditions for the spread of a disease by introducing more than just one of each parameter matrix B, Γ, M. For this, define triples (B_j, Γ_j, M_j), $j = 1, \dots, m$, where Γ_j and M_j are $n \times n$ diagonal matrices with positive entries along the diagonal, and $B_j \in \mathbb{R}^{n \times n}$ has nonnegative entries throughout for all j. For convenience, $D_j := \Gamma_j + M_j$. Now $f_j : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is determined by equation (2.9), if we replace B, Γ, M by B_j, Γ_j, M_j, respectively. Let $\sigma : [0, \infty) \to \{1, \dots, m\}$ be an arbitrary function. We call σ a *switching signal.* The switched version of our SIR model will then be denoted by

$$\dot{x}(t) = f_{\sigma(t)}(x(t)) \tag{2.10}$$

where

$$f_{\sigma(t)}(x) := \begin{pmatrix} -D_{\sigma(t)} + B_{\sigma(t)} & 0 \\ \Gamma_{\sigma(t)} & -M_{\sigma(t)} \end{pmatrix} x \ - \ \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_{\sigma(t)} x^1 \\ 0 \end{pmatrix}.$$

[3] http://de.wikibooks.org/wiki/Medizinische_Mikrobiologie:_Orthomyxoviridae

We shall explain switched epidemiological models with our influenza example. Assume we have two maps f_1 and f_2 which describe different conditions for the spread of the virus, for instance f_1 corresponds to high infection rates (in the winter) and f_2 to low infection rates (in the summer). A model that simulates the progress of the influenza epidemic over a whole year could be built by setting up a differential equation

$$\dot{x}(t) = \begin{cases} f_1(x(t)), & 0 \leq t < t_1, \\ f_2(x(t)), & t_1 \leq t \leq t_2, \end{cases} \tag{2.11}$$

where the interval $[0, t_1)$ represents the first half of the one-year cycle and $[t_1, t_2]$ the second half. The above differential equation can also be described in a concise way if we introduce a switching signal $\sigma : [0, t_2] \to \{1, 2\}$, $\sigma(t) = 1$ if $0 \leq t < t_1$ and $\sigma(t) = 2$ if $t_1 \leq t \leq t_2$. Then

$$\dot{x} = f_{\sigma(t)}(x)$$

is equal to (2.11). The time t_1 is the point where the system switches from f_1 to f_2. This is why we call systems of the form (2.10) *switched systems*.
In the following we establish switched systems also for SIS and SIRS models. The SIS system is fully described via the equations for the infectives due to $S_i(t) = N_i - I_i(t)$. We define the state variables $y_i := I_i(t)/N_i$, $i = 1, \dots, n$ and build the state vector $y := (y_1, \dots, y_n)^T$. Then the differential equations

$$\dot{y}_i(t) = (1 - y_i(t)) \sum_{k=1}^{n} b_{ik} y_k(t) \ - \alpha_i y_i(t), \quad i = 1, \dots, n,$$

give rise to a multigroup SIS model. Keep in mind that $1 - y_i(t) = S_i(t)/N_i$. In matrix form, the above equations equal

$$\dot{y} = (-D + B)y - \operatorname{diag}(y)By.$$

The switched multigroup SIS model is then given by

$$\dot{y} = (-D_{\sigma(t)} + B_{\sigma(t)})y - \operatorname{diag}(y)B_{\sigma(t)}y. \tag{2.12}$$

The switched multigroup SIRS model is similar to (2.10), but we have to include the parameters $\omega_1, \dots, \omega_n$ that represent the loss-of-immunity rates in each subgroup. Therefore, $\Omega := \operatorname{diag}(\omega_1, \dots, \omega_n)$ and in the switched case, $\Omega_j, j = 1, \dots, m$ are $n \times n$ diagonal matrices with nonnegative entries on the diagonal. Now the switched multigroup SIRS model is given by

$$\dot{x} = \begin{pmatrix} -D_{\sigma(t)} + B_{\sigma(t)} & 0 \\ \Gamma_{\sigma(t)} & -M_{\sigma(t)} - \Omega_{\sigma(t)} \end{pmatrix} x - \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_{\sigma(t)} x^1 \\ 0 \end{pmatrix}. \tag{2.13}$$

In Theorem 2.1 we saw that under certain conditions the solution $(s(t), i(t))$ to the simple SIR model (2.5) converges to the *disease-free equilibrium* $(1, 0)$ as $t \to \infty$. The aim of the thesis is to indicate conditions which assure an analogous statement for switched multigroup SIR, SIS and SIRS models. The question is already answered for switched SIS models in the article [1] and we will widely use the results presented there to establish conditions for SIR and SIRS models. Various mathematical notions and instruments are needed to approach this problem. We shall provide all necessary definitions and theorems in the next chapter.

3 Mathematical Tools

We will need a wide range of mathematical tools to analyze the epidemiological systems (2.10), (2.12) and (2.13). The purpose of the present chapter is to equip the reader with all these tools. Since it will not always be clear for what purpose we introduce some new notion or prove some theorem, we shall provide the reader with a map that shows the connections between the various mathematical terms and how they are used later to prove certain stability properties of epidemiological models. The reader is invited to look at this map from time to time for not getting lost.

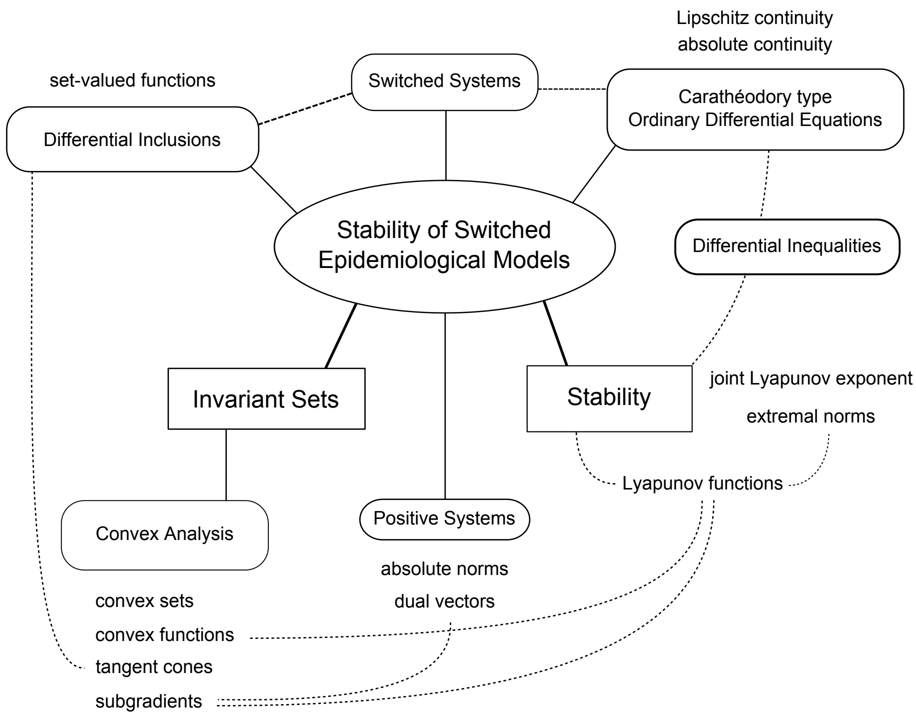

Figure 7: The map shows the interplay of most of the definitions that will be introduced.

3.1 Preliminaries

Troughout the text we will use the following notations. The symbols $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$ denote the natural, real, nonnegative real numbers, respectively. The positive orthant in $\mathbb{R}^n$, that is the subset of vectors $x \in \mathbb{R}^n$ with $x_i \geq 0$ for all $i = 1, \ldots, n$, is denoted by $\mathbb{R}^n_+$. Let X be a vector space over $\mathbb{R}$. For two subsets $A, B \subset X$, $x \in X$ and $\lambda \in \mathbb{R}$ we define,

$$A \pm B := \{a \pm b \mid a \in A, b \in B\},$$
$$x + A := \{x\} + A$$

and

$$\lambda A := \{\lambda \cdot a \mid a \in A\}.$$

If $\|\cdot\| : X \to \mathbb{R}$ is a norm, then

$$\mathcal{B}_{\|\cdot\|} := \{x \in X \mid \|x\| \leq 1\}$$

denotes the closed unit ball in X with respect to the norm $\|\cdot\|$. If the norm is clear from the context, we will omit the index. The closed ball around an arbitrary $x \in X$ with radius $r > 0$ would then be expressed by $x + r\mathcal{B}$. On $\mathbb{R}^{n\times n}$ we introduce the operator norm

$$\|A\| := \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|\leq 1} \|Ax\| = \sup_{\|x\|=1} \|Ax\|, \tag{3.1}$$

induced by a given vector norm on $\mathbb{R}^n$. For the equalities, see [36, p. 46] for instance. The supremum is in fact a maximum, i. e. there always exists a vector $x^0 \in \mathbb{R}^n \setminus \{0\}$ with $\|A\| = \|Ax^0\|/\|x^0\|$. To see this, consider the right hand side of (3.1) and mind that since an arbitrary norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a continuous function it admits maximum and minimum on the compact set $\{x \in \mathbb{R}^n \mid \|x\| = 1\}$. An operator norm is always *submultiplicative*, i. e. for any $A, B \in \mathbb{R}^{n\times n}$ the inequality $\|AB\| \leq \|A\|\|B\|$ holds.
A matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ is called *diagonal*, if $a_{ij} = 0$ for all $i \neq j$. The standard inner product on $\mathbb{R}^n$ is denoted by $\langle\cdot,\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, i. e. for $x, y \in \mathbb{R}^n$ we have $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$. Sometimes we simply write $\langle x, y\rangle = x \cdot y = xy$.
Given a nonempty set $M \subset X$, the terms $\operatorname{int} M$, $\operatorname{cl} M$, $\operatorname{bd} M$ denote the interior, the closure, the boundary of M, respectively. A *neighborhood* of M is a set $N \subset X$, which contains an open set U with $M \subset U$. The symbol $\mathcal{P}(X)$ denotes the power set of X, that is the set of all subsets of X.
We introduce the *distance function* $d : \mathbb{R}^n \to \mathbb{R}$ which measures the distance of an element $x \in \mathbb{R}^n$ to a nonempty set $A \subset \mathbb{R}^n$ in a given norm $\|\cdot\|$.

$$d(x, A) := \inf\{\|x - a\| \mid a \in A\}.$$

If A is closed, then there exist elements $a \in A$ for which $d(x, A) = \|x - a\|$. In this case, $\pi(x, A)$ denotes the set of all those points in A which satisfy the latter equality.
When we write $\lambda_k \searrow \lambda$, $(\lambda_k \nearrow \lambda)$, we mean a sequence $(\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}$, $\lambda_k > \lambda$, $(\lambda_k < \lambda)$ for all $k \in \mathbb{N}$ and $\lim_{k\to\infty} \lambda_k = \lambda$.
The reader should be familiar with differential calculus and ordinary differential equations, including stability theory. We will make use of elementary results in Lebesgue measure and integration theory. When we speak of a *measurable* or *integrable* function, we always mean that with respect to the Lebesgue measure.

3.2 Continuity Notions

Let $f : D \to \mathbb{R}^m$ be a function defined on an open subset $D \subset \mathbb{R}^n$. It is well known what we mean if we say that f is a continuous or differentiable function. In contrast, the notion of *Lipschitz continuity* differs in the literature, in particular when the addendum *local* or *global* is used. So we give a precise definition of Lipschitz continuity and show the relation between it and differentiability, continuity respectively.

Definition 3.1. *Let $D \subset \mathbb{R}^n$ and $f : D \to \mathbb{R}^m$. $\|\cdot\|$ denote arbitrary norms on $\mathbb{R}^n, \mathbb{R}^m$ respectively. We say that f is* Lipschitz continuous[4] *if there exists a constant $L > 0$ such that*

$$\|f(x) - f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in D. \tag{3.2}$$

f is called Lipschitz continuous on a set *$M \subset D$, if the restriction $f|_M$ of f to the set M is Lipschitz continuous.*
We say that the map f is locally Lipschitz continuous *if for every $x \in D$ there exists a neighborhood U of x such that f is Lipschitz continuous on $U \cap D$.*

Every locally Lipschitz continuous function is continuous: Let $x \in D$ be arbitrary. There exists a neighborhood U of x and a constant $L > 0$ such that (3.2) holds for all $y \in U \cap D$. Given $\varepsilon > 0$, choose $0 < \delta \leq \varepsilon/L$ but at least so small that $x + \delta\mathcal{B} \subset U$. Then for all $y \in (x + \delta\mathcal{B}) \cap D$ it holds that

$$\|f(x) - f(y)\| \leq L\|x - y\| < L \cdot \delta \leq \varepsilon.$$

Examples for Lipschitz continuous functions are linear maps between vector spaces and norms. As we will see later, the distance function is also Lipschitz continuous. In contrast, the real function $f(x) = x^2$ is not Lipschitz continuous, but only locally Lipschitz continuous. Fortunately, the following proposition holds.

Proposition 3.2. *Let $D \subset \mathbb{R}^n$ and $f : D \to \mathbb{R}^m$ be locally Lipschitz continuous. Then f is Lipschitz continuous on every compact set $K \subset D$.*

Proof. See [5, Satz 2.4.5]. □

A practical criterion to verify local Lipschitz continuity of a function states the following

Proposition 3.3. *Let $D \subset \mathbb{R}^n$ be open. Assume the map $f : D \to \mathbb{R}^m$ possesses partial derivatives in every component of x and each of the functions $\frac{\partial f_j}{\partial x_i}(x)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, is continuous in D. Then f is locally Lipschitz continuous in D.*

Proof. See [5, Satz 2.4.6]. □

The modulus function $f(x) = |x|$ shows us that (locally) Lipschitz continuous function do not have to be differentiable. Nevertheless, the famous theorem of Rademacher states that this is “almost” true.

Theorem 3.4 (Rademacher)**.** *Let $f : D \to \mathbb{R}^m$ be a locally Lipschitz continuous function on an open subset $D \subset \mathbb{R}^n$. Then f is almost everywhere differentiable, that is, the set of points $x \in D$ where f is not differentiable has Lebesgue measure zero.*

Proof. [16, Chapter 3, Theorem 2]. □

Proposition 3.5. *Let $A \subset \mathbb{R}^n$ be nonempty. Then the distance function $d(x, A)$ is Lipschitz continuous with constant $L = 1$, i.e. let $x, y \in \mathbb{R}^n$, then*

$$|d(x, A) - d(y, A)| \leq \|x - y\|.$$

[4]This property is sometimes called *global Lipschitz continuity.*

Proof. [5] Fix $x, y \in \mathbb{R}^n$ and let $z \in A$ be arbitrary. By definition of the distance function and the triangle inequality we have

$$d(x, A) \leq \|x - z\| \leq \|x - y\| + \|y - z\|.$$

Since $z \in A$ was arbitrary, it holds that

$$d(x, A) - \|x - y\| \leq \inf\{\|y - z\| \mid z \in A\} = d(y, A).$$

Equivalently,

$$d(x, A) - d(y, A) \leq \|x - y\|.$$

Permuting x and y we get $d(y, A) - d(x, A) \leq \|x - y\|$ and finally

$$|d(x, A) - d(y, A)| \leq \|x - y\|.$$

□

We now introduce the notion of Lipschitz continuity for set-valued functions. A *set-valued function* $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$ is a map where $F(x)$ is a subset of $\mathbb{R}^m$ for every $x \in \mathbb{R}^n$.

Definition 3.6. *Given two norms on $\mathbb{R}^n, \mathbb{R}^m$ respectively, a set-valued function $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^m)$ is said to be Lipschitz continuous on a set $K \subset \mathbb{R}^n$, if there exists $L > 0$ such that*

$$F(x) \subset F(y) + L\|x - y\|\mathcal{B} \quad \text{for all } x, y \in K. \tag{3.3}$$

Here, $\mathcal{B}$ denotes the closed unit ball on $\mathbb{R}^m$.

For a vector-valued map $f : \mathbb{R}^n \to \mathbb{R}^m$, the above definition of Lipschitz continuity coincides with the one given in Definition 3.1. Now we want to present a simple fact which will be of use in later chapters.

Proposition 3.7. *Assume we have a family of Lipschitz continuous functions $f_i : D \subset \mathbb{R}^n \to \mathbb{R}^m$, $i = 1, \ldots, k$. Then the set-valued function $F : D \to \mathcal{P}(\mathbb{R}^m)$, $F(x) := \{f_i(x) \mid i = 1, \ldots, k\}$, is Lipschitz continuous in the sense of Definition 3.6.*

Proof. Let $l_i > 0$ be the constant that satisfies the Lipschitz condition (3.2) for f_i, $i = 1, \ldots, k$. Set $L := \max\{l_i \mid i = 1, \ldots, k\}$, then

$$\|f_i(x) - f_i(y)\| \leq L\|x - y\| \quad \forall i = 1, \ldots, k.$$

This is equivalent to $f_i(x) \in f_i(y) + L\|x - y\|\mathcal{B}$ or $\{f_i(x)\} \subset \{f_i(y)\} + L\|x - y\|\mathcal{B}$ for all $i = 1, \ldots, k$, respectively. Therefore the inclusion

$$F(x) \subset F(y) + L\|x - y\|\mathcal{B}$$

holds for all $x, y \in D$, and thus, $F(x)$ satisfies condition (3.3). □

Besides Lipschitz continuity, the notion of *absolute continuity* plays a crucial role in the theory of ordinary differential equations. Therefore we will introduce its definition and talk about some properties of absolutely continuous functions.

[5] [5, Anhang C, Hilfssatz C.1]

Definition 3.8. *Let* $[a,b] \subset \mathbb{R}$ *be a compact interval. A function* $f : [a,b] \to \mathbb{R}^n$ *is said to be* absolutely continuous, *if for all* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that for all collections*

$$a \leq \alpha_1 < \beta_1 \leq \alpha_2 < \beta_2 \leq \cdots \leq \alpha_m < \beta_m \leq b$$
$$\text{with} \quad \sum_{i=1}^{m} (\beta_i - \alpha_i) < \delta \tag{3.4}$$

it holds that

$$\sum_{i=1}^{m} \|f(\beta_i) - f(\alpha_i)\| < \varepsilon.$$

Obviously, absolutely continuous functions are continuous (choose $m = 1$ in the definition). Further, Lipschitz continuous functions are absolutely continuous.

Proposition 3.9.

(i) Let $f : [a,b] \to \mathbb{R}^n$ *be locally Lipschitz continuous. Then* f *is absolutely continuous.*

(ii) Let $f : [a,b] \to \mathbb{R}^n$ *be absolutely continuous and* $g : \mathbb{R}^n \to \mathbb{R}^p$ *be locally Lipschitz continuous. Then* $g \circ f : [a,b] \to \mathbb{R}^p$ *is absolutely continuous.*

Proof. By Proposition 3.2, f is in fact Lipschitz continuous on $[a,b]$ with a constant $L > 0$. Given $\varepsilon > 0$, set $\delta = \varepsilon / L$. Now

$$\sum_{i=1}^{m} \|f(\beta_i) - f(\alpha_i)\| \leq L \sum_{i=1}^{m} (\beta_i - \alpha_i) < L\delta = \varepsilon,$$

and the claim is true.
For the second statement, note that $f([a,b]) = K \subset \mathbb{R}^n$ is a compact set, since $[a,b]$ is compact and f is continuous. Hence for $g(\cdot)$ there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$ for all $x, y \in K$. For an arbitrary $\varepsilon > 0$, set $\tilde{\varepsilon} = \varepsilon / L$. Then there exists $\delta > 0$ such that for all collections (3.4) we have $\sum_{i=1}^{m} \|f(\beta_i) - f(\alpha_i)\| < \tilde{\varepsilon}$. Thus,

$$\sum_{i=1}^{m} \big\|g(f(\beta_i)) - g(f(\alpha_i))\big\| \leq L \sum_{i=1}^{m} \big\|f(\beta_i) - f(\alpha_i)\big\| < L\tilde{\varepsilon} = \varepsilon,$$

which was the assertion. □

Absolutely continuous functions emerge to be the main ingredient for the fundamental theorem of calculus for Lebesgue measurable functions.

Theorem 3.10. *A function* $f : [a,b] \to \mathbb{R}^n$ *is absolutely continuous if and only if there exists an integrable function* $\varphi(\cdot)$ *on* $[a,b]$ *such that*

$$f(t) - f(a) = \int_a^t \varphi(s)ds \quad \text{for all } t \in [a,b].$$

In this case, f *is almost everywhere differentiable on* (a,b) *and its derivative* f' *satisfies* $f'(t) = \varphi(t)$ *almost everywhere.*

Proof. Follows from [15, Kapitel VII, Satz 4.14]. □

For simplicity, we will identify $f'(t) \equiv \varphi(t)$ for all $t \in (a,b)$ and therefore use the notation $f'(t)$ even if f is not differentiable in t.

3.3 Positive Systems

If we consider the systems of differential equations introduced in Chapter 2 we see that solution trajectories to these systems describe the evolution of the respective quantities *susceptibles, infectives, recovered* in time. In reality, these quantites are always nonnegative and we would ask that each component of a solution trajectory is nonnegative for every $t \geq 0$ as well if we claimed that our epidemiological models make sense. Systems with this property arise in many practical applications and there exists a mathematical notion adapted to this situation.

A system is called *positive*, if given an initial state $x(0) \in \mathbb{R}^n_+$, the state variables are nonnegative for any time $t > 0$. As we will see in Chapter 4, solution trajectories to our epidemiological systems always remain inside certain subsets of the positive orthant $\mathbb{R}^n_+$. In the following we shall have a closer look at several notions that are of help when investigating positive systems. We recommend the book [20] of Horn and Johnson for further information about the different types of norms we establish below as well as about nonnegative vectors and matrices.

Let $x = (x_i) \in \mathbb{R}^n$. We say that the vector x is *nonnegative* and write

$$x \geq 0 \quad :\Leftrightarrow \quad x_i \geq 0 \quad \text{for all} \quad i = 1, \dots, n.$$

Analogously, a matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ is *nonnegative*,

$$A \geq 0 \quad :\Leftrightarrow \quad a_{ij} \geq 0 \quad \text{for all} \quad i, j = 1, \dots, n.$$

For two vectors x, y we define $x \geq y \Leftrightarrow x - y \geq 0$. For $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$, we define $|x|$ via $|x|_i := |x_i|$ and $|A|$ via $|A|_{ij} := |a_{ij}|$ for all $i, j = 1, \dots, n$. Some interesting facts about nonnegative vectors and matrices are given in the following

Lemma 3.11. *Let $A \in \mathbb{R}^{n \times n}$ and $x, y \in \mathbb{R}^n$. Then*

(i) $|x + y| \leq |x| + |y|$.

(ii) $|Ax| \leq |A||x|$.

(iii) $|x| \leq |y| \Rightarrow A|x| \leq A|y|$ *for every* $A \geq 0$.

(iv) If $Ax \geq 0$ *for all* $x \geq 0$, *then* $A \geq 0$.

Proof. (i) This is the triangle equality $|x_i + y_i| \leq |x_i| + |y_i|$ applied for all i.
(ii) Consider the ith entry of $|Ax|$,

$$|Ax|_i = \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \sum_{j=1}^{n} |a_{ij}||x_j| = [|A||x|]_i.$$

(iii) The ith component of the difference $A|y| - A|x|$ is given by

$$[A|y|]_i - [A|x|]_i = \sum_{j=1}^{n} a_{ij}|y_j| - \sum_{j=1}^{n} a_{ij}|x_j| = \sum_{j=1}^{n} a_{ij}(|y_j| - |x_j|) \geq 0.$$

(iv) Suppose there exists an index $i_0 j_0$ such that $a_{i_0 j_0} < 0$. Then define the vector $x \in \mathbb{R}^n$ by $x_j = 0$ if $j \neq j_0$ and $x_{j_0} = 1$. Now

$$[Ax]_{i_0} = \sum_{j=1}^{n} a_{i_0 j} x_j = a_{i_0 j_0} < 0,$$

which contradicts the assumption. □

It will be useful to study norms that are adapted to the nonnegative setting.

Definition 3.12. *A norm $\|\cdot\|$ on $\mathbb{R}^n$ is called*

(i) absolute *if for any $x \in \mathbb{R}^n$ the equality $\|x\| = \||x|\|$ holds.*

(ii) monotone *if $|x| \leq |y|$ implies $\|x\| \leq \|y\|$ for all $x, y \in \mathbb{R}^n$.*

These two properties are not unrelated, as the following theorem states which is taken from [20, Theorem 5.5.10].

Theorem 3.13. *A norm $\|\cdot\|$ on $\mathbb{R}^n$ is absolute if and only if it is monotone.*

Another type of norm is obtained by the concept of duality. Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, its *dual norm* $\|\cdot\|^*$ is defined by

$$\|y\|^* := \max\{\langle x, y\rangle \mid \|x\| \leq 1\} = \max\{\langle x, y\rangle \mid \|x\| = 1\}, \quad y \in \mathbb{R}^n, \tag{3.5}$$

where $\langle\cdot,\cdot\rangle$ is the standard inner product on $\mathbb{R}^n$. The dual norm is really a norm, see [20, Section 5.4, p. 275f]. To see the last equality in (3.5), assume $\|y\|^* > 0$ and the maximizing vector $\tilde{x}$, for which $\|y\|^* = \langle\tilde{x}, y\rangle$, has norm $0 < \|\tilde{x}\| < 1$. Then there exists $\lambda > 1$ such that $\|\lambda\tilde{x}\| = 1$. This implies $\langle\lambda\tilde{x}, y\rangle > \langle\tilde{x}, y\rangle > 0$, a contradiction to maximality. It is interesting to know that a norm is absolute if and only if its dual norm is, see [1, Section 2] and references therein. From (3.5) we deduce for all $x, y \in \mathbb{R}^n$ the inequality

$$\langle x, y\rangle \leq \|x\|\|y\|^*. \tag{3.6}$$

For $x = 0$ the inequality is clear. Let $x \neq 0$, then

$$\left\langle \frac{x}{\|x\|}, y \right\rangle \leq \max\{\langle z, y\rangle \mid \|z\| = 1\} = \|y\|^*. \tag{3.7}$$

Multiplying (3.7) with $\|x\|$ gives (3.6). For a given $x \in \mathbb{R}^n$ we define the set

$$\mathcal{D}(x) := \{y \in \mathbb{R}^n \mid \|y\|^* \leq 1 \ \wedge \ \langle x, y\rangle = \|x\|\}. \tag{3.8}$$

$\mathcal{D}(x)$ is called the *dual of x with respect to* $\|\cdot\|$ and its elements are the *dual vectors to x*. From inequality (3.6) follows that

$$\mathcal{D}(x) := \begin{cases} \{y \in \mathbb{R}^n \mid \|y\|^* = 1 \ \wedge \ \langle x, y\rangle = \|x\|\} & \text{if } x \neq 0, \\ \mathcal{B}_{\|\cdot\|^*} & \text{if } x = 0. \end{cases} \tag{3.9}$$

Sometimes, dual vectors to x are defined via the equation

$$\langle x, y\rangle = \|x\|\|y\|^*. \tag{3.10}$$

By (3.9) we see that every $y \in \mathcal{D}(x)$ satisfies the latter equation. Conversely, if a pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ satisfies (3.10) and $y \neq 0$, then $(\|y\|^*)^{-1} \cdot y$ is in the dual of x. The following lemmas show some properties of dual vectors, if the norm in consideration is absolute.

Lemma 3.14. *Let $\|\cdot\|$ be an absolute norm on $\mathbb{R}^n$ and $x \in \mathbb{R}^n_+$. If $y \in \mathbb{R}^n$ satisfies* (3.10) *and $x_i > 0$, then $y_i \geq 0$.*

Proof. [6] By (3.10) and the assumption we have

$$\|x\|\|y\|^* = \langle x, y\rangle = \sum_{x_i>0} x_i y_i \leq \sum_{x_i>0} x_i|y_i| = \langle x, |y|\rangle \leq \|x\|\||y|\|^*.$$

The last inequality is (3.6). As $\|\cdot\|^*$ is absolute, we have $\|y\|^* = \||y|\|^*$ and so equality throughout. This implies that all summands in $\sum_{x_i>0} x_i y_i$ are nonnegative, which is the assertion. □

Lemma 3.15. *Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ be an absolute norm and $x \in \mathbb{R}^n_+$. If $y \in \mathbb{R}^n$ satisfies* (3.10)*, then so does $|y|$.*

Proof. Let $I \subseteq \{1, \ldots, n\}$ be the index set such that $y_i < 0 \Leftrightarrow i \in I$. If $I = \emptyset$, then there is nothing to show, so assume $I \neq \emptyset$. By Lemma 3.14 necessarily $x_i = 0$ for all $i \in I$. We define the diagonal matrix $P = (p_{ij}) \in \mathbb{R}^{n\times n}$ by $p_{ii} = -1$ if $i \in I$, $p_{ii} = 1$ else. Then $|y| = Py$ and $Px = x$. Now

$$\|x\|\||y|\|^* = \|Px\|\|Py\|^* = \|x\|\|y\|^* = \langle x, y\rangle = \langle Px, Py\rangle = \langle x, |y|\rangle,$$

since $\|\cdot\|$, $\|\cdot\|^*$ are absolute norms and P is symmetric with $P^2 = I$. So $|y|$ satisfies (3.10). □

3.4 Convex Analysis

The idea of convexity is outstanding in the theory of optimization. By using tools of convex analysis it is possible to solve an extremum problem even if the function related to the problem is not differentiable, providing that it is a *convex function*. For the latter class of functions a calculus similar to that of differential theory has been established that will be of great help for us, when we analyze properties of epidemiological models in Chapter 4. The book [31] of Rockafellar has become standard in convex analysis and the reader will find comprehensive information about the field in this book. We start with the basic notion of a *convex set*.

Definition 3.16. *A set $K \subset \mathbb{R}^n$ is called* convex *if for all $x, y \in K$, $\lambda \in [0, 1]$ it holds that*

$$\lambda x + (1 - \lambda)y \in K.$$

For an arbitrary set $M \subset \mathbb{R}^n$ the convex hull $\operatorname{conv} M$ *is defined as the smallest convex set that contains M, i. e.*

$$\operatorname{conv} M := \bigcap_{\substack{M\subset K,\\ K convex}} K \qquad (3.11)$$

The definition says that for any two points x, y in a convex set, the line segment between x and y also belongs to the set. Figure 8 presents some convex sets in the plane $\mathbb{R}^2$ as well as some sets which are not convex. Some facts about convex sets are presented in the following

[6] See [1, Lemma 2.1].

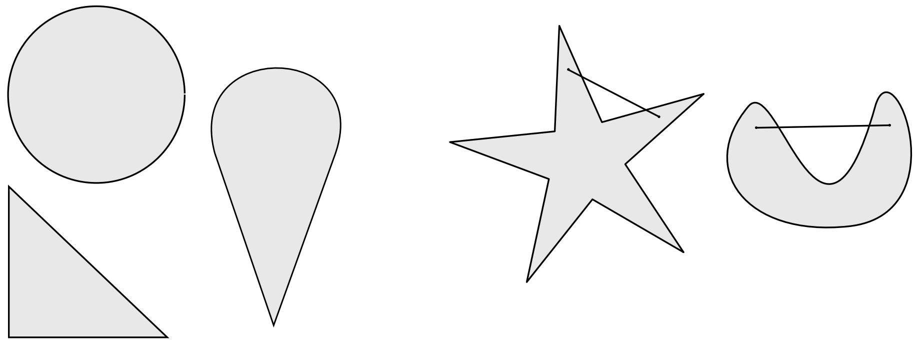

Figure 8: Examples of convex sets (left) and nonconvex sets (right).

Proposition 3.17.

(i) An intersection of an arbitrary collection of convex sets is convex.

(ii) If K_1 and K_2 are convex and $\lambda \in \mathbb{R}$, then $K_1 + K_2$ and λK_1 are convex sets.

(iii) If $M \subset \mathbb{R}^n$ is a compact set, then $\operatorname{conv} M$ is compact.

Proof. (i) Let $x, y \in \bigcap_{i \in \mathcal{I}} K_i$ where $\mathcal{I}$ is an arbitrary index set and each K_i is convex. Then $x, y \in K_i$ for all $i \in \mathcal{I}$. For $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda) y$ belongs to every K_i due to convexity of K_i. Hence $\lambda x + (1 - \lambda) y \in \bigcap_{i \in \mathcal{I}} K_i$.
(ii) can be found in [31, §3], (iii) is [33, Corollary 1.1]. □

Assertion (i) in the above proposition assures that $\operatorname{conv} M$ defined by (3.11) is really a convex set.

Example 3.18. For an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^n$ and $x \in \mathbb{R}^n$, $r > 0$, the closed ball around x with radius r is a convex set. By Proposition 3.17(ii) it suffices to show that $\mathcal{B}_{\|\cdot\|}$ is convex. Let $x, y \in \mathbb{R}^n$ with $\|x\|, \|y\| \leq 1$ and $\lambda \in [0, 1]$. Then

$$\|\lambda x + (1 - \lambda) y\| \leq \lambda \|x\| + (1 - \lambda) \|y\| \leq \lambda \cdot 1 + (1 - \lambda) \cdot 1 = 1,$$

such that $\lambda x + (1 - \lambda) y \in \mathcal{B}$.

We now introduce the *tangent cone*. That is, a set related to a convex set and a given point within it. It may surprise but the tangent cone with its properties will help us later to characterize the behavior of solution trajectories of differential equations. We first explain what a cone is.

Definition 3.19. *A set $C \subset \mathbb{R}^n$ is called a* cone *if for all $x \in K$ and all $\lambda > 0$ it holds that $\lambda x \in K$.*

Definition 3.20. *Let $K \subset \mathbb{R}^n$ be a convex set and $x \in K$. The set*

$$T_K(x) := \operatorname{cl} \left\{ \frac{y - x}{\lambda} \;\middle|\; y \in K, \lambda > 0 \right\} = \operatorname{cl} \bigcup_{\lambda > 0} \frac{K - x}{\lambda}$$

is called the tangent cone to K at x.

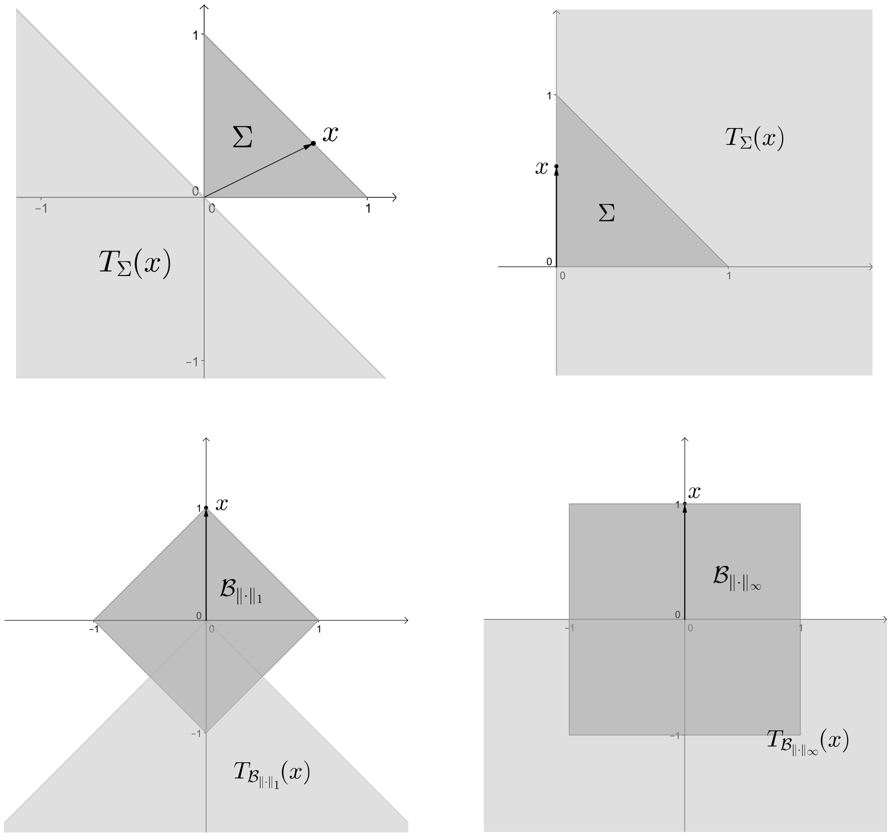

Figure 9: The figure shows the sets Σ, $\mathcal{B}_{\|\cdot\|_1}$ and $\mathcal{B}_{\|\cdot\|_\infty}$ (dark gray), and the respective tangent cones at given points x (light gray).

Tangent cones to several sets at given points x are illustrated in Figure 9. The considered sets are $\Sigma = \{x \in \mathbb{R}^2_+ \mid x_1 + x_2 \leq 1\}$, $\mathcal{B}_{\|\cdot\|_1} := \{x \in \mathbb{R}^2 \mid |x_1| + |x_2| \leq 1\}$ and $\mathcal{B}_{\|\cdot\|_\infty} := \{x \in \mathbb{R}^2 \mid \max\{|x_1|, |x_2|\} \leq 1\}$.
We prove some properties of tangent cones in the following and give an equivalent description for them. Afterwards we determine the tangent cone of several subsets of $\mathbb{R}^n$ that play an important role later on.

Lemma 3.21. *Let $K \subset \mathbb{R}^n$ be a convex set and $x \in K$. Then*

$$\frac{K - x}{\beta} \subset \frac{K - x}{\alpha}$$

whenever $0 < \alpha < \beta$.

Proof. [7] Let $0 < \alpha < \beta$ and $y \in K$ be arbitrary. Set $z = (1 - \alpha/\beta)x + (\alpha/\beta)y$. By convexity of K it holds that $z \in K$. Then

$$\frac{y - x}{\beta} = \frac{1}{\beta}\left(\frac{\beta}{\alpha}z - \frac{\beta}{\alpha}x + x - x\right) = \frac{z - x}{\alpha} \in \frac{K - x}{\alpha}$$

[7] [33, Lemma 1.4]

and the lemma is proved. □

Proposition 3.22. *Let $K \subset \mathbb{R}^n$ be a convex set and $x \in K$. then*

$$T_K(x) = \{w \in \mathbb{R}^n \mid \lim_{\lambda \searrow 0} \lambda^{-1} d(x + \lambda w, K) = 0\}. \tag{3.12}$$

Proof. [8] "$\subset$" First of all, consider the following equality.

$$\begin{aligned} \lambda^{-1} d(x + \lambda w, K) &= \inf_{y \in K} \frac{\|x + \lambda w - y\|}{\lambda} = \inf_{y \in K} \left\| w - \frac{y - x}{\lambda} \right\| \\ &= \inf \left\{ \|w - z\| \;\middle|\; z \in \frac{K - x}{\lambda} \right\} = d\left(w, \frac{K - x}{\lambda}\right), \quad \lambda > 0. \end{aligned}$$

Now, let $w \in T_K(x)$ and $\varepsilon > 0$. Then there exist $y \in K, \delta > 0$ such that $\|w - \delta^{-1}(y - x)\| < \varepsilon$ or equivalently, $d(w, (K - x)/\delta) < \varepsilon$. By the above equality and Lemma 3.21 we receive

$$\lambda^{-1} d(x + \lambda w, K) = d(w, (K - x)/\lambda) \leq d(w, (K - x)/\delta) < \varepsilon$$

for all $0 < \lambda < \delta$, and therefore

$$\lim_{\lambda \searrow 0} \lambda^{-1} d(x + \lambda w, K) = 0. \tag{3.13}$$

"$\supset$" If (3.13) holds for a vector $w \in \mathbb{R}^n$, then for each $\varepsilon > 0$ there exists $\lambda > 0$ such that $d(w, (K - x)/\lambda) < \varepsilon$. This yields

$$w \in \operatorname{cl} \bigcup_{\lambda > 0} \frac{K - x}{\lambda},$$

and the claim is true. □

Corollary 3.23. *Let $w \in T_K(x)$ and $\lambda_k \searrow 0$. Then there exists a sequence $(u_k) \subset \mathbb{R}^n, u_k \to 0$ such that*

$$x + \lambda_k w + \lambda_k u_k \in K \quad \textit{for all} \quad k \in \mathbb{N}.$$

Proof. Let y_k be in the set $\pi(x + \lambda_k w, K)$ for each $k \in \mathbb{N}$. Define the vector u_k by the equation $y_k = x + \lambda_k w + \lambda_k u_k$. Then

$$\begin{aligned} &\lim_{k \to \infty} \lambda_k^{-1} d(x + \lambda_k w, K) = \lim_{k \to \infty} \lambda_k^{-1} \|x + \lambda_k w - y_k\| \\ &= \lim_{k \to \infty} \lambda_k^{-1} \|\lambda_k u_k\| = \lim_{k \to \infty} \|u_k\| = 0 \end{aligned}$$

by Proposition 3.22 and therefore $u_k \to 0$, which was to show. □

Proposition 3.24.

(i) The tangent cone of $\mathbb{R}^n_+$ at a point $x \in \mathbb{R}^n_+$ is

$$T_{\mathbb{R}^n_+}(x) = \{z \in \mathbb{R}^n \mid z_i \geq 0 \textit{ if } x_i = 0\}.$$

[8][33, Proposition 1.3]

(ii) The tangent cone of $\Sigma := \{x \in \mathbb{R}^{2n}_+ \mid x_i + x_{n+i} \leq 1 \ \forall i = 1, \dots, n\}$ *at a point* $x \in \Sigma$ *is*

$$T_\Sigma(x) = \{z \in \mathbb{R}^{2n} \mid z_i \geq 0 \text{ if } x_i = 0 \text{ and } z_i + z_{n+i} \leq 0 \text{ if } x_i + x_{n+i} = 1\}.$$

(iii) The tangent cone of $\Pi := [0,1]^n$ *at a point* $x \in \Pi$ *is*

$$T_\Pi(x) = \{z \in \mathbb{R}^n \mid z_i \geq 0 \text{ if } x_i = 0 \text{ and } z_i \leq 0 \text{ if } x_i = 1\}.$$

Proof. (i) "$\subset$" Let $x \in \mathbb{R}^n_+$ and $z \in T_{\mathbb{R}^n_+}(x)$. Then, by definition, there exist sequences $(\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+ \setminus \{0\}, (y^k)_{k\in\mathbb{N}} \subset \mathbb{R}^n_+$ such that

$$z = \lim_{k\to\infty} \lambda_k^{-1}(y^k - x),$$

and thus,

$$z_i = \lim_{k\to\infty} \lambda_k^{-1}(y_i^k - x_i) \quad \forall i = 1, \dots, n.$$

If $x_i = 0$, then $\lambda_k^{-1} y_i^k \geq 0$ for all $k \in \mathbb{N}$ and consequently $z_i \geq 0$. Hence the inclusion holds.
"$\supset$" Let $x \in \mathbb{R}^n_+$ and $z \in \mathbb{R}^n$ with $z_i \geq 0$ if $x_i = 0$. We aim to find $y \in \mathbb{R}^n_+$ and $\lambda > 0$ such that $z = \lambda^{-1}(y - x)$. Equivalently, $y_i = \lambda z_i + x_i$ for all $i = 1, \dots, n$. Hence, the vector y is completely determined after having set λ. And by choosing $\lambda > 0$ appropriately, we assure that $y \in \mathbb{R}^n_+$. For this, we set $\lambda := \min\{-\frac{x_i}{z_i} \mid z_i < 0\}$ if this set is nonempty and $\lambda := 1$ else. In each case, λ is positive. Now, if $x_i = 0$ then by assumption $z_i \geq 0$ and $y_i = \lambda z_i \geq 0$. If $x_i > 0$ and $z_i \geq 0$, then as well $y_i = \lambda z_i + x_i \geq 0$. If at last $x_i > 0$ and $z_i < 0$, then due to our definition of λ we have $\lambda \leq -\frac{x_i}{z_i} \Leftrightarrow \lambda z_i + x_i \geq 0$. Hence $y \in \mathbb{R}^n_+$ and the claim is true.
(ii) "$\subset$" Let $x \in \Sigma$ and $z \in T_\Sigma(x)$. Then, by definition, there exist sequences $(\lambda_k)_{k\in\mathbb{N}} \subset \mathbb{R}_+ \setminus \{0\}, (y^k)_{k\in\mathbb{N}} \subset \Sigma$ such that

$$z = \lim_{k\to\infty} \lambda_k^{-1}(y^k - x).$$

Considering each component, we get

$$z_i = \lim_{k\to\infty} \lambda_k^{-1}(y_i^k - x_i) \quad \forall i = 1, \dots, 2n,$$

and therefore,

$$z_i + z_{n+i} = \lim_{k\to\infty} \lambda_k^{-1}\Big(y_i^k + y_{n+i}^k - (x_i + x_{n+i})\Big).$$

By assumption, $0 \leq y_i^k + y_{n+i}^k \leq 1$. If $x_i + x_{n+i} = 1$, then

$$\lambda_k^{-1}\Big(y_i^k + y_{n+i}^k - (x_i + x_{n+i})\Big) \leq 0$$

and consequently $z_i + z_{n+i} \leq 0$. For the other condition we refer to the proof of (i) "$\subset$": If $x_i = 0$, then $z_i \geq 0$. Hence the inclusion holds.
"$\supset$" Let $x \in \Sigma$ and $z \in \mathbb{R}^{2n}$ with $z_i \geq 0$ if $x_i = 0$ and $z_i + z_{n+i} \leq 0$ if $x_i + x_{n+i} = 1$. We aim to find $y \in \Sigma$ and $\lambda > 0$ such that $z = \lambda^{-1}(y - x)$ or

equivalently $y_i = \lambda z_i + x_i$ for all $i = 1, \dots, 2n$. The vector y has to satisfy the following requirements:

$$0 \le y_i = \lambda z_i + x_i \quad \forall i = 1, \dots, 2n \quad \text{and}$$
$$y_i + y_{n+i} = \lambda(z_i + z_{n+i}) + x_i + x_{n+i} \le 1 \quad \forall i = 1, \dots, n.$$

At first – as before in (i) "$\supset$" – we assure that y_i is nonnegative for all $i = 1, \dots, 2n$ by choosing $0 < \lambda \le \lambda_1$ where $\lambda_1 := \min\{-\frac{x_i}{z_i} \mid z_i < 0\}$ if this set is nonempty and $\lambda_1 := 1$ else. In a second step we have to specify λ such that $y_i + y_{n+i} \le 1$ for all $i = 1, \dots, n$. For this, let

$$\lambda_2 := \min \left\{ \frac{1 - x_i - x_{n+i}}{z_i + z_{n+i}} \mid z_i + z_{n+i} > 0 \right\},$$

provided that this set is not empty and $\lambda_2 := 1$ else. In each case, λ_2 is positive. Now set $\lambda := \min\{\lambda_1, \lambda_2\}$ and consider the sum $y_i + y_{n+i}$. If $x_i + x_{n+i} = 1$ then by assumption $z_i + z_{n+i} \le 0$, which yields $y_i + y_{n+1} \le 1$. The same holds if $x_i + x_{n+i} < 1$ and $z_i + z_{n+i} \le 0$. At last, consider the case $x_i + x_{n+i} < 1$ and $z_i + z_{n+i} > 0$. Then

$$\lambda \le \frac{1 - x_i - x_{n+i}}{z_i + z_{n+i}} \;\Leftrightarrow\; \lambda(z_i + z_{n+i}) + x_i + x_{n+i} = y_i + y_{n+i} \le 1$$

for all $i = 1, \dots, n$. Hence $y \in \Sigma$ and the claim is true.
(iii) The proof is entirely analogous to (ii). Drop all terms with indices $j \ge n+1$ and you receive the claim. $\square$

At this point we shall give an inequality that turns out to be crucial for the stability analysis of epidemiological models.

Lemma 3.25. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and $x \in \mathbb{R}^n \setminus \{0\}$ be arbitrary. $T_B(x)$ denotes the tangent cone to B at x, where $B := \|x\| \mathcal{B}_{\|\cdot\|}$. Consider a vector y in the dual of x. Then*

$$\langle z, y \rangle \le 0 \qquad \forall z \in T_B(x). \tag{3.14}$$

Proof. At first we show that $\langle w, y \rangle \le \langle x, y \rangle$ holds for all $w \in B$. For this, choose $w \in B$ arbitrary. By definition of B, $\|w\| \le \|x\|$. Recall (3.6) and (3.10), then

$$\langle w, y \rangle \le \|w\| \|y\|^* \le \|x\| \|y\|^* = \langle x, y \rangle.$$

Now consider $z = \lambda^{-1}(w - x)$ for $\lambda > 0$. Then

$$\left\langle \frac{w - x}{\lambda}, y \right\rangle = \frac{1}{\lambda} \Big(\langle w, y \rangle - \langle x, y \rangle \Big) \le 0.$$

From here we conclude that the above inequality holds for all $z \in T_B(x)$ which is (3.14). $\square$

Before we pass over to *convex functions* we would like to point out that it is possible to define tangent cones even for nonconvex sets. The difficulty herein is that there does not exist a unique definition for the tangent cone. Instead a variety of notions have been established for nonconvex sets, see for instance [32, Chapter 11]. The reader finds a very illuminating history of tangent cones in the book [4, Chapter 4] of Aubin and Frankowska. Now we turn to the next important notion in convex analysis.

Definition 3.26. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if for all $x, y \in \mathbb{R}^n$, $\lambda \in [0,1]$ we have*

$$f(\lambda x + (1 - \lambda y)) \leq \lambda f(x) + (1 - \lambda) f(y). \tag{3.15}$$

In the theory of convex functions it is usually allowed that f maps to $\mathbb{R} \cup \{\infty\}$. Since this is not relevant for us we will omit this detail. Convex functions are connected to convex sets in the following way. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its *epigraph*

$$\text{epi}(f) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq y\}$$

is a convex set. Before we discuss further properties of convex functions, we shall give an important example. Every norm $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Let $x, y \in \mathbb{R}^n$ be arbitrary and $\lambda \in [0,1]$. Then

$$\|\lambda x + (1 - \lambda) y\| \leq \lambda \|x\| + (1 - \lambda) \|y\|$$

which shows convexity.

Proposition 3.27. *Every convex function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous.*

Proof. The proof is given in [33, Theorem 1.9] or in [32, Corollary 1.4.2]. □

As we already know by Rademacher's Theorem 3.4, every locally Lipschitz continuous function is almost everywhere differentiable. So are convex functions. It holds even more: Convex functions possess directional derivatives at each point $x \in \mathbb{R}^n$ in every direction $z \in \mathbb{R}^n$, see [33, Theorem 1.12]. It is desired to generalize the concept of differentiability for convex functions. For this, consider a convex function $f : \mathbb{R}^n \to \mathbb{R}$ which is differentiable in $x \in \mathbb{R}^n$. Then for any $y \in \mathbb{R}^n$,

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle,$$

where $\nabla f(x)$ denotes the gradient of f in x, see [31, Theorem 25.1]. This remarkable property motivates the definition of the following set.

Definition 3.28. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. A vector x^* is said to be a* subgradient *of f at a point x if*

$$f(y) \geq f(x) + \langle x^*, y - x \rangle \quad \forall y \in \mathbb{R}^n. \tag{3.16}$$

The set of all subgradients is denoted by $\partial f(x)$ and is called subdifferential *of f at x. If $\partial f(x)$ is nonempty then f is called* subdifferentiable *at x.*

The reader can find comprehensive information about convex functions and subdifferentials in the book [31, §23] of Rockafellar. In fact, a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is subdifferentiable everywhere and $\partial f(x)$ is a nonempty compact convex set for each $x \in \mathbb{R}^n$. The notion of subdifferentiability is consistent with differentiability in the sense that if a subdifferentiable function is differentiable in x then $\partial f(x) = \{\nabla f(x)\}$. We are now interested in a characterization of the set $\partial f(x)$ if f is a norm. Here the concept of dual vectors we introduced in Section 3.2 is of help.

Proposition 3.29. *Let $v : \mathbb{R}^n \to \mathbb{R}$ be a norm and $x \in \mathbb{R}^n$. $\mathcal{D}(x)$ shall denote the set of all dual vectors to x. Then*

$$\partial v(x) = \mathcal{D}(x). \tag{3.17}$$

Proof. [9] Recall the facts deduced in Section 3.3.
"$\subset$" Let x^* be a subgradient of $x \in \mathbb{R}^n$. Then by definition the inequality (3.16) holds for all $y \in \mathbb{R}^n$. For $y = 0$ we receive $\langle -x, x^* \rangle \leq -v(x)$ or $\langle x, x^* \rangle \geq v(x)$, respectively. For $y = 2x$ we get $\langle 2x - x, x^* \rangle \leq v(2x) - v(x)$ or equivalently, $\langle x, x^* \rangle \leq v(x)$. The combination of the the two inequalities yields

$$\langle x, x^* \rangle = v^*(x).$$

To see that $v(x^*) \leq 1$, let $z \in \mathbb{R}^n$ be arbitrary. Then

$$\langle z, x^* \rangle = \langle z + x - x, x^* \rangle \leq v(z + x) - v(x) \leq v(z) + v(x) - v(x) = v(z), \tag{3.18}$$

again by (3.16) and triangle inequality. Let $z_0 \in \mathcal{B}_v$ be a vector for which $v^*(x^*) = \langle z_0, x^* \rangle$ holds. Thus, by (3.18), we have

$$v^*(x^*) = \langle z_0, x^* \rangle \leq v(z_0) \leq 1$$

This gives $x^* \in \mathcal{D}(x)$.
"$\supset$" Let $x^* \in \mathcal{D}(x)$ and $y \in \mathbb{R}^n$ be arbitrary. Then

$$\begin{aligned}\langle y - x, x^* \rangle &= \langle y, x^* \rangle - \langle x, x^* \rangle = \langle y, x^* \rangle - v(x) \\ &\leq v(y)v^*(x^*) - v(x) \leq v(y) - v(x),\end{aligned}$$

by considering the definition of a dual vector and inequality (3.6). □

There exist formulae to compute the subdifferential of certain convex functions. For example, if $\lambda > 0$ and $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex, then $\partial(\lambda f)(x) = \lambda \partial f(x)$ and $\partial(f + g)(x) = \partial f(x) + \partial g(x)$. By contrast, we cannot expect a product rule for subgradients, since a product of two convex functions is not necessarily convex, consider for instance $f(x)g(x) = x \cdot x^2$.
The desire to develop a useful differential calculus even for nonconvex functions led to a theory which is known nowadays under the keyword *nonsmooth analysis*. An approach to this field is given by Clarke et al. in the book [11]. The theory described in this book builds the core for the analysis of our switched epidemiological models.
Consider a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$. We already know due to RADEMACHER, Theorem 3.4, that f is differentiable almost everywhere. Moreover if f is Lipschitz continuous on a neighborhood U of x with a constant $L > 0$, then $\|\nabla f(y)\| \leq L$ for every $y \in U$ where f is differentiable. Consider the set

$$\partial_C f(x) := \operatorname{conv} \left\{ \lim_{k \to \infty} \nabla f(x_k) \;\middle|\; x_k \to x,\ \nabla f(x_k) \text{ exists} \right\}. \tag{3.19}$$

$\partial_C f(x)$ is called the *Clarke generalized gradient* of f at x. We will call elements of $\partial_C f(x)$ *subgradients*, analogously to subgradients of convex functions. It holds that $\partial_C f(x)$ is a nonempty compact convex set for each $x \in \mathbb{R}^n$, see [11,

[9] See also [32, Proposition 4.6.2].

Theorem 0.1.5], and that it is related to the subgradient or the usual gradient in the following way. If f is a convex function, then the Clarke generalized gradient and the subdifferential defined above coincide, that is $\partial_C f(x) = \partial f(x)$ for every x, see [11, Proposition 2.4.3]. From the definition of the set $\partial_C f(x)$ it follows directly that $\partial_C f(x) = \{\nabla f(x)\}$ if f is continuously differentiable in x. In Chapter 4 we will need to compute the Clarke generalized gradient of a product of a Lipschitz continuous function and a continuously differentiable function.

Proposition 3.30. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous and $g : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Then*

$$\partial_C(fg)(x) \subset \partial_C f(x)g(x) + f(x)\nabla g(x). \tag{3.20}$$

Proof. Let $x \in \mathbb{R}^n$ and consider a sequence $x_k \to x$ where $\nabla f(x_k)$ exists for every $k \in \mathbb{N}$. Then fg is differentiable in x_k and

$$\nabla(fg)(x_k) = \nabla f(x_k)g(x_k) + f(x_k)\nabla g(x_k), \qquad \forall k \in \mathbb{N}, \tag{3.21}$$

by classical differential calculus. In the following, suppose $g(x) \neq 0$. In this case, the limit $\lim_{k\to\infty} \nabla(fg)(x_k)$ exists if and only if $\lim_{k\to\infty} \nabla f(x_k)$ exists due to (3.21). If $\nabla f(x_k)$ converges we obtain

$$\begin{aligned}\lim_{k\to\infty} \nabla(fg)(x_k) &= \lim_{k\to\infty} \Big(\nabla f(x_k)g(x_k) + f(x_k)\nabla g(x_k)\Big) \\ &= \lim_{k\to\infty} \nabla f(x_k) \cdot g(x) + f(x)\nabla g(x).\end{aligned}$$

Therefore,

$$\begin{aligned}&\Big\{ \lim_{k\to\infty} \nabla(fg)(x_k) \;\Big|\; x_k \to x, \nabla(fg)(x_k) \text{ exists}\Big\} \\ &= \Big\{ \lim_{k\to\infty} \nabla f(x_k) \cdot g(x) + f(x)\nabla g(x) \;\Big|\; x_k \to x, \lim_{k\to\infty} \nabla f(x_k) \text{ exists}\Big\} \\ &= \Big\{ \lim_{k\to\infty} \nabla f(x_k) \;\Big|\; x_k \to x, \lim_{k\to\infty} \nabla f(x_k) \text{ exists}\Big\} \cdot g(x) + f(x)\nabla g(x) \qquad (3.22)\end{aligned}$$

and hence

$$\begin{aligned}\partial_C(fg)(x) &= \operatorname{conv} \Big\{ \lim_{k\to\infty} \nabla(fg)(x_k) \;\Big|\; x_k \to x, \nabla(fg)(x_k) \text{ exists}\Big\} \\ &\subset \partial_C f(x) \cdot g(x) + f(x)\nabla g(x). \qquad (3.23)\end{aligned}$$

The set inclusion holds due to (3.22) and the fact that the right-hand side in (3.23) is a convex set, recall Proposition 3.17.
Now suppose $g(x) = 0$. Since f is Lipschitz continuous, $\|\nabla f(x_k)\|$ is locally bounded around x, i. e. for each k big enough. Thus, $\lim_{k\to\infty} \nabla f(x_k)g(x_k) = 0$ and

$$\lim_{k\to\infty} \nabla(fg)(x_k) = \lim_{k\to\infty} f(x_k)\nabla g(x_k) = f(x)\nabla g(x).$$

Therefore,

$$\begin{aligned}\partial_C(fg)(x) &= \operatorname{conv} \Big\{ \lim_{k\to\infty} \nabla(fg)(x_k) \;\Big|\; x_k \to x, \nabla(fg)(x_k) \text{ exists}\Big\} \\ &= \{f(x)\nabla g(x)\} \subset \partial_C f(x)g(x) + f(x)\nabla g(x),\end{aligned}$$

which proves (3.20). □

We place another statement at this point which will be needed in Chapter 4.

Lemma 3.31. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and $v : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous. Then the map $\zeta : \mathbb{R}^n \to \mathbb{R}$ defined by*

$$\zeta(x) := \max\{\langle z, f(x)\rangle \mid z \in \partial_C v(x)\}$$

is upper semicontinuous*, that is for each $x \in \mathbb{R}^n$,*

$$\limsup_{x_k \to x} \zeta(x_k) \leq \zeta(x),$$

or equivalently, for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\zeta(y) < \zeta(x) + \varepsilon \quad \forall y \in \mathbb{R}^n \text{ with } \|y - x\| < \delta.$$

Proof. At first, note that ζ is well defined since for a fixed $x \in \mathbb{R}^n$, $\partial_C v(x)$ is a nonempty compact set and $\langle \cdot, f(x)\rangle : \mathbb{R}^n \to \mathbb{R}$ is continuous, such that the latter admits its maximum on the set $\partial_C v(x)$. The claim itself is a consequence of Propositions 1.5(b) and 1.1(b) in [11, Chapter 2]. □

It is well known that continuous functions $\mathbb{R}^n \to \mathbb{R}$ admit minimum and maximum on compact sets. Upper semicontinuous functions still admit their maximum on compact sets.

Lemma 3.32. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be upper semicontinuous and $K \subset \mathbb{R}^n$ a compact set. Then there exists $\xi \in K$ such that $f(x) \leq f(\xi)$ for all $x \in K$.*

Proof. Let $y := \sup_{x \in K} f(x)$ which could a priori be infinite. Consider a sequence (x_n) in K for which $\lim_{n\to\infty} f(x_n) = y$. Since K is compact, there exists a convergent subsequence (x_k) with $\lim_{k\to\infty} x_k = \xi \in K$. By upper semicontinuity of f we have $f(\xi) \geq \lim_{k\to\infty} f(x_k) = y$. Therefore $f(\xi) = y$. □

3.5 Ordinary Differential Equations

3.5.1 Carathéodory-type Differential Equations

In the theory of ordinary differential equations one considers usually an initial value problem

$$\dot{x} = f(t, x) \tag{3.24}$$

$$x(t_0) = x^0, \tag{3.25}$$

where $f : T \times X \to \mathbb{R}^n$ is defined on an open set $T \times X \subset \mathbb{R} \times \mathbb{R}^n$ and $(t_0, x^0) \in T \times X$. A *solution* to (3.24) is a continuously differentiable function $\varphi : I \to \mathbb{R}^n$ on an interval $I \subset T$ for which $(t, \varphi(t)) \in T \times X$, and it holds that $\dot{\varphi}(t) = f(t, \varphi(t))$ for all $t \in I$.
It is well known[10] that (3.24)–(3.25) has a unique solution if f is continuous and locally Lipschitz continuous in x, uniformly in t. Consider the integral equation

$$x(t) = x^0 + \int_{t_0}^{t} f(s, x(s))ds \tag{3.26}$$

[10] [5, Satz 2.4.1]

which is equivalent to (3.24)–(3.25) in the following sense. If f is continuous and $\varphi : I \to \mathbb{R}^n$ is continuously differentiable on an interval I containing t_0, then φ solves (3.24) and (3.25) if and only if it solves (3.26). However, continuity of f is not necessary for the integral to exist. One could suppose that there exists a function that solves (3.26) under milder conditions than those mentioned above. It was CARATHÉODORY who treated that problem and developed a more general existence theory. For this, we relax the notion of a solution to (3.24).

An absolutely continuous function $\varphi : I \to \mathbb{R}^n$ is called a *solution of* (3.24) *in the extended sense on* I if

(i) $(t, \varphi(t)) \in T \times X$ for all $t \in I$,

(ii) $\dot{\varphi}(t) = f(t, \varphi(t))$ almost everywhere on I.

We already stated in Theorem 3.10 that an absolutely continuous function is almost everywhere differentiable, such that (ii) is reasonable. Now we cite the following theorem given in [19, Theorem 2.1.14].

Theorem 3.33 (Carathéodory). *Assume $f : T \times X \to \mathbb{R}^n$, where $T \subset \mathbb{R}$ is an open interval and X an open subset of $\mathbb{R}^n$. Suppose that f satisfies the following "Carathéodory conditions":*

(i) $f(\cdot, x)$ is measurable for each fixed $x \in X$,

(ii) $f(t, \cdot)$ is continuous for each fixed $t \in T$,

(iii) $\|f(\cdot, \tilde{x})\|$ is locally integrable on T for some $\tilde{x} \in X$,

(iv) for each compact set $C = I \times K \subset T \times X$ there exists an integrable function $L_C(\cdot) : I \to \mathbb{R}_+$ such that

$$\|f(t, x) - f(t, y)\| \le L_C(t)\|x - y\| \quad \forall (t, x), (t, y) \in C.$$

Then for any $(t_0, x^0) \in T \times X$ there exists a unique solution $x(\cdot) = \varphi(\cdot; t_0, x^0)$ of (3.24) in the extended sense on some maximal open interval $T(t_0, x^0) \subset T$ containing t_0, such that $x(t_0) = x^0$. Moreover,

(i) if $t_+(t_0, x^0) := \sup T(t_0, x^0) < \sup T$ then $x(t)$ is unbounded as $t \nearrow t_+(t_0, x^0)$ or the boundary of X is not empty and $d(x(t), \operatorname{bd} X) \to 0$ as $t \nearrow t_+(t_0, x^0)$.

(ii) if $\mathbb{D}_\varphi$ is the domain of definition of the general solution φ,

$$\mathbb{D}_\varphi = \{(t, t_0, x^0) \mid t \in T(t_0, x^0), (t_0, x^0) \in T \times X\},$$

then $\mathbb{D}_\varphi$ is open in $T^2 \times \mathbb{R}^n$ and $\varphi : \mathbb{D}_\varphi \to \mathbb{R}^n$ is continuous.

Remark 3.34. From (i) in the above theorem we deduce that, if the boundary of X is empty and $x(t)$ is bounded as $t \nearrow t_+(t_0, x^0)$, then $t_+(t_0, x^0) = \sup T$.

For the rest of the text we will always investigate functions that are solutions of differential equations in the extended sense, such that we drop the addendum 'in the extended sense' and just speak of a solution. Continuing from the last theorem, we state the following two propositions.

Proposition 3.35. [11] *Given the assumptions of Theorem 3.33, let* $(t_0, x^0) \in T \times X$ *and* $t_1 \in T(t_0, x^0)$*. If* $x^1 = \varphi(t_1; t_0, x^0)$*, then* $T(t_1, x^1) = T(t_0, x^0)$ *and*

$$\varphi(t; t_1, x^1) = \varphi(t; t_0, x^0) \qquad \forall t \in T(t_0, x^0). \tag{3.27}$$

The identity (3.27) *is called* cocycle property *of the general solution* φ*.*

Proposition 3.36. [12] *Suppose* $T \subset \mathbb{R}$ *is an open interval,* $X \subset \mathbb{R}^n$ *is open and* $f : T \times X \to \mathbb{R}^n$ *is affinely bounded, that is*

$$\|f(t,x)\| \leq M(t)\|x\| + m(t), \quad (t,x) \in T \times X, \tag{3.28}$$

where $M(\cdot)$*,* $m(\cdot)$ *are locally integrable nonnegative functions on* T*. Then every solution of* (3.24) *is bounded on every finite interval* (t_1, t_2)*,* $t_1, t_2 \in T$*,* $t_1 < t_2$ *on which it is defined. If moreover* $X = \mathbb{R}^n$ *then every solution of* (3.24) *can be continued to all of* T*.*

We revisit briefly some properties of linear differential equations. The reader can find these results in [19, Chapter 2.2] or [5, Kapitel 6]. Consider

$$\dot{x}(t) = A(t)x(t), \tag{3.29}$$

$$x(t_0) = x^0, \tag{3.30}$$

where $A : I \to \mathbb{R}^{n \times n}$ is a measurable matrix function defined on an interval $I \subset \mathbb{R}$ and $(t_0, x^0) \in I \times \mathbb{R}^n$. Let $\varphi(\cdot; t_0, x^0)$ denote the solution of (3.29)–(3.30). Then for each $t \in I$ there exists a linear map $\Phi(t, t_0) : \mathbb{R}^n \to \mathbb{R}^n$ satisfying

$$\varphi(t; t_0, x^0) = \Phi(t, t_0)x^0. \tag{3.31}$$

$\Phi(t, t_0)$ is called the *fundamental matrix* associated with (3.29). It has the following properties.

(i) $\Phi(t_0, t_0) = I$ for all $t_0 \in I$,

(ii) $\Phi(t_2, t_0) = \Phi(t_2, t_1)\Phi(t_1, t_0)$ for all $t_2 \geq t_1 \geq t_0 \in I$,

(iii) $\Phi(t, t_0)$ is invertible for all $t_0, t \in I$.

For a linear system

$$\dot{x}(t) = A(t)x(t) + g(t), \qquad x(t_0) = x^0, \tag{3.32}$$

with measurable inhomogeneity $g : I \to \mathbb{R}^n$, the solution φ can be expressed by the *variation of constants* formula

$$\varphi(t; t_0, x^0) = \Phi(t, t_0)x^0 + \int_{t_0}^{t} \Phi(t, s)g(s)ds. \tag{3.33}$$

Assume that $A : \mathbb{R} \to \mathbb{R}^{n \times n}$ satisfies the conditions of Proposition 3.36. Then the solution $\varphi(t; t_0, x^0)$ of (3.29)–(3.30) exists for every $t \in \mathbb{R}$ and is bounded on every finite interval (t_0, τ) with $\tau > t_0$. Fix $t_0 \in \mathbb{R}$ and consider $\|\varphi(\cdot; t_0, \cdot)\| : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ as a function of time t and the initial value x^0. Then it is continuous

[11] See [5, Satz 2.6.5].
[12] See [19, Proposition 2.1.19].

by Theorem 3.33 and admits its maximum on the compact set $[t_0, \tau] \times \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. Thus the fundamental matrix $\Phi(\cdot, t_0) : [t_0, \tau] \to \mathbb{R}^{n\times n}$ is bounded as well since for the operator norm it holds

$$\|\Phi(t,t_0)\| = \max_{\|x^0\|=1} \|\Phi(t,t_0)x^0\| = \max_{\|x^0\|=1} \|\varphi(t;t_0,x^0)\|, \qquad \forall t \in [t_0,\tau],$$

such that there exists a constant $M > 0$ with

$$\|\Phi(t,t_0)\| \leq M \quad \forall t \in [t_0,\tau]. \tag{3.34}$$

3.5.2 Differential Inequalities

The following theorem will be substantial in Chapter 4. It uncovers an important relation between trajectories of SIS and SIR systems. Before we start we shall explain a particular notion. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *increasing in the variable* x_i, $i = 1, \ldots, n$, if for $x, y \in \mathbb{R}^n$ with $x_j = y_j$, $j \neq i$, and $x_i < y_i$ we have $f(x) \leq f(y)$.

Theorem 3.37. *Given the initial value problem* (3.24)–(3.25), *assume the right hand side f satisfies the Carathéodory conditions of Theorem 3.33. Suppose that the function $f_i(t,x)$ is increasing in the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ for every $i \in \{1, \ldots, n\}$. Let $\varphi : I \to \mathbb{R}^n$ be the unique solution of* (3.24)–(3.25) *on an interval $I \subset \mathbb{R}$ including t_0. Then the following proposition is true.*
Let $\psi : I \to \mathbb{R}^n$ be an absolutely continuous function satisfying the initial inequality

$$\psi(t_0) \leq \varphi(t_0)$$

and differential inequality

$$\dot{\psi}(t) \leq f(t, \psi(t)) \tag{3.35}$$

almost everywhere on the interval I. Then

$$\psi(t) \leq \varphi(t) \quad \forall t \in I.$$

Proof. See [35, Theorem 16.2]. □

We give a simple corollary of the last theorem which will be needed later.

Corollary 3.38. *Suppose an absolutely continuous function $\psi : [0,T] \to \mathbb{R}$ satisfies*

$$\dot{\psi}(t) \leq c \cdot \psi(t) \tag{3.36}$$

for some $c \in \mathbb{R}$ almost everywhere on $[0,T]$. Then

$$\psi(t) \leq e^{ct}\psi(0) \quad \forall t \in [0,T]. \tag{3.37}$$

Proof. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = cx$ and the corresponding initial value problem $\dot{x} = f(x)$, $x(0) = \psi(0)$. The unique solution to it is given by $\varphi(t) = e^{ct}\psi(0)$. We are now in the situation of Theorem 3.37. Inequalities (3.35) and (3.36) are equivalent and, by construction, $\psi(0) = \varphi(0)$. Application of Theorem 3.37 yields (3.37). □

3.5.3 Switched Systems and Differential Inclusions

In the theory of Dynamical Systems, one distinguishes usually between systems with continuous and those with discrete dynamics. However, there are a lot of applications – mainly in physics and engineering – where the two concepts appear in combination. Think of an electric circuit where current flows continuously whereas we can influence its behaviour by discrete events like turning on and off switches. You can also imagine a continuous data flow in a computer network or an industrial plant working continuously which is controlled by discrete-time inputs. These use cases led researchers from different fields, like computer scientists, mathematicians and engineers, to work on those problems and develop a new theory, which is now known under the keyword *hybrid systems* or *switched systems*. D. Liberzon gives an overview to the field in his book [27]. There exists a variety of possibilities for modeling switched systems. We concentrate in the following on those with *time-dependent switching*, which we want to specify now.
Suppose we have a family of functions $f_j : D \subset \mathbb{R}^n \to \mathbb{R}^n, j \in \mathcal{I}$, where $\mathcal{I}$ is an arbitrary index set. This gives rise to a family of systems

$$\dot{x} = f_j(x), \quad j \in \mathcal{I}, \tag{3.38}$$

evolving on $\mathbb{R}^n$. The above family becomes a switched system, if we add a *switching signal*. This is simply a map $\sigma : \mathbb{R}_+ \to \mathcal{I}$. In the following we restrict us to the case $\mathcal{I} = \{1, \ldots, m\}$ for some $m \in \mathbb{N}$. In general, we will require that $\sigma(\cdot)$ is measurable and define

$$\mathcal{S} := \{\sigma : \mathbb{R}_+ \to \{1, \ldots, m\} \mid \sigma \text{ is measurable}\}.$$

If this is too abstract for you, you can also think of a piecewise constant function $\sigma : \mathbb{R}_+ \to \{1, \ldots, m\}$ being continuous from the right. That is, there exists some $\tau > 0$ such that $|t - s| \geq \tau$ for each two points t, s of discontinuity of σ and it holds that $\lim_{t \searrow t_0} \sigma(t) = t_0$ for each $t_0 \in \mathbb{R}_+$.[13] The points of discontinuity are called *switching times* or *switching instants*.
A switched system is now described by the equation

$$\dot{x}(t) = f_{\sigma(t)}(x(t)). \tag{3.39}$$

An immediate question is whether there exist solutions to the above system. The answer is positive, if all the f_j's are locally Lipschitz continuous! We briefly demonstrate that in the following lemma.

Lemma 3.39. *Let $f_j : D \to \mathbb{R}^n, j \in \{1, \ldots, m\}$, be a finite family of locally Lipschitz continuous maps, defined on an open subset $D \subset \mathbb{R}^n$. Further, let $\sigma : \mathbb{R}_+ \to \{1, \ldots, m\}$ be measurable. Then the function $f(t, x) := f_{\sigma(t)}(x)$ satisfies the Carathéodory conditions of Theorem 3.33 and therefore the switched system* (3.39) *has a unique solution for each $(t_0, x^0) \in \mathbb{R}_+ \times D$.*

Proof. Fix $x \in D$. Then, $f(\cdot, x) : \mathbb{R}_+ \to \mathbb{R}^n$ is measurable since $\sigma(\cdot) : \mathbb{R}_+ \to \{1, \ldots, m\}$ is measurable.

[13]This point of view is justified by the fact that a measurable function $\sigma : \mathbb{R}_+ \to \{1, \ldots, m\}$ can be approximated, on any compact interval, arbitrarily well by piecewise constant, right-continuous maps. See [1, p. 4].

Fix $t \in \mathbb{R}_+$. Then $f(t,\cdot) = f_j(\cdot)$ for a certain $j \in \{1,\dots,m\}$ and is hence continuous, since each f_j is locally Lipschitz continuous.
Let $x \in D$ and consider an arbitrary bounded interval $I \subset \mathbb{R}_+$. Then the integral $\int_I \|f(t,x)\|dt < \infty$, since $f(\cdot,x)$ is a measurable function and takes finite values for each $t \in \mathbb{R}_+$. Hence, $\|f(\cdot,x)\|$ is locally integrable for each $x \in D$.
Finally, let $C = I \times K \subset \mathbb{R}_+ \times D$ be an arbitrary compact set. From Proposition 3.2 we know that each $f_j(\cdot)$ is Lipschitz continuous on K with a constant $L_j > 0$. Set $L := \max\{L_j \mid j = 1,\dots,m\}$. Then

$$\|f(t,x) - f(t,y)\| \leq L\|x-y\| \quad \text{for all } (t,x),(t,y) \in C.$$

The constant map $L_C(\cdot) : I \to \mathbb{R}_+, L_C(t) \equiv L$, is obviously integrable, such that all Carathéodory conditions are satisfied. Thus, for each initial condition $(t_0, x^0) \in \mathbb{R}_+ \times D$, there exists a unique solution of (3.39). □

With the above lemma we have verified that each of the epidemiological models introduced in Chapter 2 has a unique solution.

The field of Switched Systems is contained in the more general theory of *Differential Inclusions*. There, one no longer considers a differential equation $\dot{x} = f(t,x)$, but an inclusion of the form

$$\dot{x}(t) \in F(t, x(t)),$$

where $F : \mathbb{R}^{n+1} \to \mathcal{P}(\mathbb{R}^n)$ is a set-valued map, i.e. for every $x \in \mathbb{R}^n$, the image $F(x)$ is a subset of $\mathbb{R}^n$. An introduction to the field is given by Smirnov [33]. In the following we restrain to the case of time-invariant inclusions $\dot{x} \in F(x)$.

Definition 3.40. *Consider a differential inclusion*

$$\dot{x}(t) \in F(x(t)), \tag{3.40}$$

where $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is a set-valued map and let $x^0 \in \mathbb{R}^n$. If there exist $T > 0$ and an absolutely continuous function $x : [0,T] \to \mathbb{R}^n$ such that $x(0) = x^0$ and the inclusion (3.40) *holds almost everywhere on the interval $[0,T]$, then $x(\cdot)$ is called a* solution to differential inclusion (3.40) *with initial condition x^0.*

As for ordinary differential equations, there has been developed an existence theory of solutions to differential inclusions. The following theorem specifies conditions under which differential inclusions possess solutions and how the latter depend on initial conditions. For this, we would like to introduce the normed vector space $\mathcal{C}([0,T],\mathbb{R}^n)$ of continuous functions from the interval $[0,T]$ to $\mathbb{R}^n$. The space shall be endowed with the norm

$$\|x(\cdot)\|_{\mathcal{C}} := \max\{\|x(t)\| \mid t \in [0,T]\}.$$

The set of solutions to (3.40) with initial condition $x^0 \in \mathbb{R}^n$ is a subset of $\mathcal{C}([0,T],\mathbb{R}^n)$ and will be denoted by $\mathbb{S}_{[0,T]}(F,x^0)$. Finally, for a nonempty subset $C \subset \mathbb{R}^n$,

$$\mathbb{S}_{[0,T]}(F,C) := \bigcup_{x^0 \in C} \mathbb{S}_{[0,T]}(F,x^0).$$

Theorem 3.41. *Let $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a set-valued map with closed convex values for all $x \in \mathbb{R}^n$ and assume that F is Lipschitz continuous on every compact set $K \subset \mathbb{R}^n$. Then for every $x^0 \in \mathbb{R}^n$ there exist solutions to differential inclusion* (3.40) *with $x(0) = x^0$.*
If $C \subset \mathbb{R}^n$ is a compact set, then $\mathbb{S}_{[0,T]}(F, C)$ is a compact subset of the space $\mathcal{C}([0, T], \mathbb{R}^n)$. In particular there exists a constant $M > 0$ such that the inequality

$$\|x(\cdot)\|_{\mathcal{C}} \leq M \tag{3.41}$$

holds for all $x(\cdot) \in \mathbb{S}_{[0,T]}(F, C)$.

Proof. The proof can be found in [33]. The first part is Corollary 4.1 there, the second part follows from Corollary 4.5. See also the proof of Theorem 4.11. □

The connection between switched systems and differential inclusions is the following. Assume we have given a switched system $\dot{x} = f_{\sigma(t)}(x)$, built by a family of locally Lipschitz continuous functions $f_j : \mathbb{R}^n \to \mathbb{R}^n$, $j = 1, \dots, m$, and a switching signal $\sigma \in \mathcal{S}$. If $x(\cdot; x^0, \sigma)$ is a solution to the switched system with initial condition $x^0 \in \mathbb{R}^n$, then it is a solution to the differential inclusion $\dot{x} \in F(x)$ where $F(x) := \{f_j(x) \mid j = 1, \dots, m\}$. We aim to apply Theorem 3.41 to switched systems. Therefore a Lipschitz continuous set-valued function with closed convex values is needed. We already know that F as defined above is Lipschitz continuous on every compact set due to Proposition 3.7. But $F(x)$ is not a convex set. To get rid of this difficulty, we define $\hat{F}(x) := \operatorname{conv} F(x)$ for each $x \in \mathbb{R}^n$. This function has obviously convex values and they are closed due to Proposition 3.17(iii). It is also Lipschitz continuous on compact sets as we will see now. Let $K \subset \mathbb{R}^n$ be compact and $L > 0$ the constant such that

$$F(x) \subset F(y) + L\|x - y\|\mathcal{B}$$

for all $x, y \in K$. Then

$$F(x) \subset \operatorname{conv} F(y) + L\|x - y\|\mathcal{B},$$

which is equivalent to

$$\operatorname{conv} F(x) \subset \operatorname{conv} F(y) + L\|x - y\|\mathcal{B}. \tag{3.42}$$

This follows from Proposition 3.17 and Example 3.18: $L\|x - y\|\mathcal{B}$ is a convex set and therefore the sum on the right hand side of (3.42) is convex, too. Since $\operatorname{conv} F(x)$ is the smallest convex set containing $F(x)$, it has to be a subset of $\operatorname{conv} F(y) + L\|x - y\|\mathcal{B}$. Therefore, $\hat{F}(x) \subset \hat{F}(y) + L\|x - y\|\mathcal{B}$. This shows Lipschitz continuity of $\hat{F}$.
Consider a family of matrices $\mathcal{M} = \{A_j \in \mathbb{R}^{n \times n} \mid j = 1, \dots, m\}$ and a related linear switched system

$$\dot{x} = A_{\sigma(t)}x, \quad \sigma \in \mathcal{S}.$$

From Proposition 3.36 we know that a solution $\varphi(\cdot; x^0, \sigma)$ with an initial condition $x^0 \in \mathbb{R}^n$ exists for every $t \geq 0$. If we set

$$\mathcal{A}(x) := \operatorname{conv}\{A_j x \mid j = 1, \dots, m\},$$

then $\varphi(\cdot; x^0, \sigma)$ is also a solution to the differential inclusion

$$\dot{x} \in \mathcal{A}(x).$$

Let $T > 0$ be arbitrary and $C = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. Then Theorem 3.41 tells us that there exists a constant $M > 0$ such that

$$\forall t \in [0,T],\ x^0 \in C,\ \sigma \in \mathcal{S}: \quad \|\varphi(t; x^0, \sigma)\| \leq M,$$

and therefore,

$$\forall t \in [0,T],\ \sigma \in \mathcal{S}: \quad \|\Phi_\sigma(t,0)\| \leq M. \tag{3.43}$$

That is, the fundamental matrix $\Phi_\sigma(\cdot, 0)$ is uniformly bounded over all $\sigma \in \mathcal{S}$ on bounded intervals $[0,T]$.
If we are interested in the behavior of solutions to switched systems or differential inclusions, then it would be useful to know that there exist subsets in $\mathbb{R}^n$, such that solutions which enter one of those sets never leave it again. This property is reflected by the term *invariance.*

Definition 3.42. *A set $C \subset \mathbb{R}^n$ is called* invariant *by differential inclusion* (3.40)*, if any solution of* (3.40) *with initial value $x_0 \in C$ remains in C for every time t, i.e. $x(0) = x_0 \in C \Rightarrow x(t) \in C$ for all $t \geq 0$.*

Remark 3.43. We have shown above how a differential inclusion $\dot{x} \in F(x)$ is related to a switched system $\dot{x} = f_{\sigma(t)}(x)$. We will therefore call a set $C \in \mathbb{R}^n$ invariant by a given switched system, if it is invariant by the related differential inclusion.

Theorem 3.44. *Let $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a set-valued map with nonempty compact values for each $x \in \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ a closed convex set. Assume that F is Lipschitz continuous on every compact subset $K \subset \mathbb{R}^n$. Then the following conditions are equivalent.*

(i) The set C is invariant by differential inclusion (3.40)*.*

(ii) For any point $x \in C$ the following inclusion holds:

$$F(x) \subset T_C(x). \tag{3.44}$$

Proof. "(i)⇒(ii)" Let $x \in C$ be arbitrary. By assumption, every solution $x(\cdot)$ of (3.40) remains in C for every time $t \geq 0$, i. e. $d(x(t), C) \equiv 0$. Pick any $w \in F(x)$. Then there exists a solution to (3.40) which is differentiable in $t = 0$ with $x(0) = x$ and $\dot{x}(0) = w$, see the proof of Theorem (4.4) in [10] and references therein. Remember from basic analysis that if $x(t)$ is differentiable in t, then $x(t + \lambda) = x(t) + \lambda\dot{x}(t) + r(\lambda)$ where $r(\cdot)$ is a function which satisfies $r(\lambda)/\lambda \to 0$ if $\lambda \to 0$. Thus

$$\begin{aligned}
\lim_{\lambda \searrow 0} \frac{1}{\lambda} d(x + \lambda w, C) &= \lim_{\lambda \searrow 0} \frac{1}{\lambda} d(x(0) + \lambda\dot{x}(0), C) \\
&= \lim_{\lambda \searrow 0} \frac{1}{\lambda} d(x(0) + \lambda\dot{x}(0) + r(\lambda), C) \\
&= \lim_{\lambda \searrow 0} \frac{1}{\lambda} d(x(\lambda), C) = 0.
\end{aligned}$$

Hence $w \in T_C(x)$ according to Proposition 3.22.
"(ii)⇒(i)"[14] Let $x \in C$. Consider a solution of (3.40) with $x(0) = x$. By definition, $x(\cdot)$ is absolutely continuous. The function $g(t) := d(x(t), C)$ is

[14] [33, Theorem 5.6]

absolutely continuous as well by Propositions 3.5 and 3.9. Our goal is to show that $g(t) = 0$ for all $t \geq 0$. This implies that $x(t) \in C$ for all $t \geq 0$.
The derivatives $\dot{x}(t)$ and $\dot{g}(x(t), C)$ exist for almost every $t > 0$. Choose one of these t. Let $y \in \pi(x(t), C)$ and $w \in T_C(y)$. For an arbitrary sequence $\alpha_k \searrow 0$ there exists a sequence $u_k \to 0$ such that $y + \alpha_k w + \alpha_k u_k \in C$ for all $k \in \mathbb{N}$ by Corollary 3.23. If $x(t)$ is differentiable in t, then $x(t+\alpha) = x(t) + \alpha\dot{x}(t) + r(\alpha)$ where $r(\cdot)$ satisfies $r(\alpha)/\alpha \to 0$ if $\alpha \to 0$. Now,

$$\begin{aligned}\dot{g}(t) &= \lim_{\alpha\to 0} \frac{1}{\alpha}\Big[d(x(t+\alpha), C) - d(x(t), C)\Big] \\ &= \lim_{\alpha\to 0} \frac{1}{\alpha}\Big[d(x(t) + \alpha\dot{x}(t) + r(\alpha), C) - d(x(t), C)\Big]\end{aligned}$$

Consider a sequence $\alpha_k \searrow 0$. For each $k \in \mathbb{N}$ we have

$$\begin{aligned}&\frac{1}{\alpha_k}\Big[d(x(t) + \alpha_k\dot{x}(t) + r(\alpha_k), C) - d(x(t), C)\Big] \\ \leq\ &\frac{1}{\alpha_k}\Big[\big\|x(t) + \alpha_k\dot{x}(t) + r(\alpha_k) - (y + \alpha_k w + \alpha_k u_k)\big\| - \big\|x(t) - y\big\|\Big]\end{aligned}$$

Taking the limit, we conclude

$$\begin{aligned}&\lim_{k\to\infty} \frac{1}{\alpha_k}\Big[\big\|x(t) + \alpha_k\dot{x}(t) + r(\alpha_k) - (y + \alpha_k w + \alpha_k u_k)\big\| - \big\|x(t) - y\big\|\Big] \\ =&\lim_{k\to\infty} \frac{1}{\alpha_k}\Big[\big\|x(t) + \alpha_k\dot{x}(t) - (y + \alpha_k w)\big\| - \big\|x(t) - y\big\|\Big] \\ \leq&\lim_{k\to\infty} \frac{1}{\alpha_k}\Big[\big\|x(t) - y\big\| + \alpha_k\big\|\dot{x}(t) - w\big\| - \big\|x(t) - y\big\|\Big] \\ =&\big\|\dot{x}(t) - w\big\|.\end{aligned}$$

Since $w \in T_C(y)$ was arbitrary and with the assumption (3.44), we obtain

$$\dot{g}(t) \leq d(\dot{x}(t), T_C(y)) \leq d(\dot{x}(t), F(y)).$$

Choose a compact set $K \subset \mathbb{R}^n$ which contains both $x(t)$ and y. By assumption, $F(\cdot)$ is Lipschitz continuous on K with a constant $L > 0$ and it holds that $F(x(t)) \subset F(y) + L\|x(t) - y\|\mathcal{B}$. Since $\dot{x}(t) \in F(x(t))$, we can continue our inequality by

$$d(\dot{x}(t), F(y)) \leq L\|x(t) - y\| = Ld(x(t), C) = Lg(t).$$

We started with $x(0) \in C$, so $g(0) = 0$. By applying Corollary 3.38, we arrive at $g(t) \equiv 0$ and hence $x(t) \in C$ for all $t \geq 0$. □

We will extensively use the above theorem in Chapter 4 to establish invariance of sets that are of interest when investigating epidemiological models. At this point we would like to consider linear switched systems. Assume we have given a finite family of matrices $\mathcal{M} = \{A_j \in \mathbb{R}^{n\times n} \mid j = 1, \ldots, m\}$, whose off-diagonal entries are nonnegative. That is, for $A_j = (a^j_{ik}) \in \mathcal{M}$ we have $a^j_{ik} \geq 0$ for every $i \neq k$. Such matrices are called *Metzler*. Now the related switched system

$$\dot{x} = A_{\sigma(t)}x, \quad \sigma \in \mathcal{S}, \tag{3.45}$$

gives rise to a differential inclusion

$$\dot{x} \in \mathcal{A}(x), \tag{3.46}$$

where $\mathcal{A}(x) := \{A_j x \mid j = 1, \dots, m\}$. We aim to show that (3.45) is a positive system. The strategy to do this is to prove that the positive orthant $\mathbb{R}^n_+$ is an invariant set by differential inclusion (3.46). Remember that the tangent cone of the positive orthant is given by

$$T_{\mathbb{R}^n_+}(x) = \{z \in \mathbb{R}^n \mid z_i \geq 0 \text{ if } x_i = 0\},$$

see Proposition 3.24. The set-valued map $\mathcal{A} : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ surely satisfies the conditions of Theorem 3.44. Hence all we have to do is to verify that for all $x \in \mathbb{R}^n_+$ the inclusion $\mathcal{A}(x) \subset T_{\mathbb{R}^n_+}(x)$ holds. This is evident since for every $x \geq 0$ with $x_i = 0$ and every $j = 1, \dots, m$ we have

$$(A_j x)_i = \sum_{k=1}^{n} a^j_{ik} x_k = \sum_{k \neq i} a^j_{ik} x_k \geq 0$$

by the definition of A_j. We conclude by the above theorem that the system (3.45) is positive.

Remark 3.45. Positive switched linear systems have a remarkable property. Since

$$\varphi(t; x^0, \sigma) = \Phi_\sigma(t, 0) x^0 \geq 0 \quad \forall t > 0, x \in \mathbb{R}^n_+, \sigma \in \mathcal{S},$$

we can deduce from Lemma 3.11(iv) that $\Phi_\sigma(t, 0) \geq 0$ for all $t > 0$. It holds even more. Let $t_1 > 0$ and $\sigma \in \mathcal{S}$ be arbitrary. Define the switching signal ω by $\omega(t) := \sigma(t + t_1)$. If $\varphi(t; t_1, x^1)$ denotes the solution to

$$\dot{x} = A_{\sigma(t)} x, \quad x(t_1) = x^1,$$

and $\psi(t; 0, x^1)$ denotes the solution to

$$\dot{x} = A_{\omega(t)} x, \quad x(0) = x^1,$$

then $\psi(t - t_1; 0, x^1) = \varphi(t; t_1, x^1)$ for all $t \geq t_1$ and all $x^1 \in \mathbb{R}^n$. Hence we receive

$$\Phi_\omega(t - t_1, 0) x^1 = \psi(t - t_1; 0, x^1) = \varphi(t; t_1, x^1) = \Phi_\sigma(t, t_1) x^1$$

and therefore $\Phi_\omega(t - t_1, 0) = \Phi_\sigma(t, t_1)$. From above we know that $\Phi_\omega(t - t_1, 0)$ is a nonnegative matrix and so

$$\Phi_\sigma(t, t_1) \geq 0 \qquad \forall t \geq t_1 \geq 0. \tag{3.47}$$

3.5.4 Stability of Switched Systems

Our goal in the analysis of epidemiological models is the investigation of stability properties. For this, we recapitulate the standard notions of Lyapunov stability and introduce the specialties concerned with switched systems.

Let

$$\dot{x} = f(x) \tag{3.48}$$

be a time-invariant differential equation with locally Lipschitz continuous right-hand side $f : D \to \mathbb{R}^n$, where $D \subset \mathbb{R}^n$ is open. With $x(t; x^0)$ we denote the solution to (3.48) with initial condition $x(0; x^0) = x^0$. Assume $\bar{x} \in D$ is an equilibrium point, i.e. $f(\bar{x}) = 0$.

Definition 3.46. *We say that $\bar{x}$ is*

(i) stable, *if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that*

$$\|x^0 - \bar{x}\| < \delta \;\Rightarrow\; \|x(t;x^0) - \bar{x}\| < \varepsilon \quad \forall t \geq 0.$$

(ii) attractive, if there exists an $R > 0$ such that

$$\|x^0 - \bar{x}\| < R \;\Rightarrow\; \lim_{t\to\infty} x(t;x^0) = \bar{x}.$$

(iii) asymptotically stable, if it is stable and attractive.

(iv) globally asymptotically stable, if it is asymptotically stable and the region of attraction

$$A(\bar{x}) := \{x^0 \in D \mid \lim_{t\to\infty} x(t;x^0) = \bar{x}\}$$

is the whole domain D.

Note that the above definitions could also be applied to a nonempty compact set $M \subset D$ instead of a single point $\bar{x}$. For this, simply replace the norm $\|\cdot\|$ by the distance function $d(\cdot, M)$ and "$=$" by "$\in$" in (ii) and (iv).[15]
Now we attend to stability notions for switched systems. Recall the setting of Section 3.5.3. For a given $\sigma \in \mathcal{S}$ and $x^0 \in D$, $x(\cdot;x^0,\sigma)$ denotes the unique solution of $\dot{x}(t) = f_{\sigma(t)}(x(t))$ satisfying $x(0;x^0,\sigma) = x^0$.

Definition 3.47. *A nonempty compact set $M \subset D$ is called*

(i) uniformly stable, if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that for every $\sigma \in \mathcal{S}$,

$$d(x^0, M) < \delta \;\Rightarrow\; d(x(t;x^0,\sigma), M) < \varepsilon \quad \forall t \geq 0.$$

(ii) uniformly attractive, if for each $\eta > 0$ there exist an $R > 0$ and a time $T > 0$ such that for each $\sigma \in \mathcal{S}$,

$$d(x^0, M) < R \;\Rightarrow\; d(x(t;x^0,\sigma), M) < \eta \quad \forall t \geq T.$$

(iii) globally uniformly attractive, if it is uniformly attractive and the region of attraction

$$A(M) := \{x^0 \in D \mid \lim_{t\to\infty} x(t;x^0,\sigma) \in M \quad \forall \sigma \in \mathcal{S}\}$$

is the whole domain D.

(iv) (globally) uniformly asymptotically stable, if it is uniformly stable and (globally) uniformly attractive.

The main extension of Definition 3.47 compared to Definition 3.46 is that all constants in the latter definition are independent of the choice of $\sigma \in \mathcal{S}$. This property is reflected by the notion *uniform*.
The immediate question now is, how can we verify whether a given set or an equilibrium point satisfies one of the stability notions above? We will study this

[15] Stability notions of compact sets are properly defined in the book [9] of Bhatia and Szegő, Chapter V.

problem for linear systems first and attend to nonlinear systems afterwards. From stability theory of linear time-invariant systems we know, that $\bar{x} = 0$ is an asymptotically stable equilibrium for the system $\dot{x} = Ax$, $A \in \mathbb{R}^{n\times n}$, if and only if all eigenvalues of A have negative real part.[16] Such matrices are called *Hurwitz*. Note that in the following we use to speak about stability of a linear system instead of a particular solution to it. This is justified by the fact that for linear systems, an arbitrary solution has one of the stability properties mentioned above if and only if the trivial solution $\bar{x} \equiv 0$ has it.[17]
Assume we have given a set of matrices $\mathcal{M} = \{A_j \in \mathbb{R}^{n\times n} \mid j = 1, \dots, m\}$, and an arbitrary $\sigma \in \mathcal{S}$. One could suppose now, that the switched linear system

$$\dot{x} = A_{\sigma(t)}x \tag{3.49}$$

is uniformly asymptotically stable if all matrices A_j are Hurwitz. Actually this is not the case. There exist switched systems, where all subsystems are asymptotically stable, but where appropriate switching leads to an unstable switched system. This is illustrated in [27, Chapter 2]. However, there are several criteria which assure that arbitrary switching does not change stability characteristics of a system. We cite one of these, that will be of use later.

Theorem 3.48. *Let $\{A_j \in \mathbb{R}^{n\times n} \mid j = 1, \dots, m\}$ be a finite family of Hurwitz matrices. Assume that all matrices commute pairwise, i.e. $A_jA_k = A_kA_j$ for all $j, k \in \{1\dots, m\}$. Then the switched linear system $\dot{x} = A_{\sigma(t)}x$ is globally uniformly exponentially stable.[18] That is, there exist constants $c, \lambda > 0$ such that*

$$\|x(t; x^0, \sigma)\| \le ce^{-\lambda t}\|x^0\| \tag{3.50}$$

for all $t \ge 0$, $x^0 \in \mathbb{R}^n$, $\sigma \in \mathcal{S}$.

Remark 3.49. Uniform exponential stability can also be expressed using the fundamental matrix. If (3.49) is globally uniformly exponentially stable, then there exist $c, \lambda > 0$ such that

$$\|\Phi_\sigma(t, s)\| \le ce^{-\lambda(t-s)} \quad \forall t \ge s \ge 0,\ \sigma \in \mathcal{S}. \tag{3.51}$$

To see this, choose the operator norm on $\mathbb{R}^{n\times n}$

$$\|\Phi_\sigma(t, s)\| = \max_{x\neq 0} \frac{\|\Phi_\sigma(t, s)x\|}{\|x\|}$$

and let $x^0 \neq 0$ be a vector for which the right hand side achieves its maximum. Then

$$\|\Phi_\sigma(t, 0)\| = \frac{\|\Phi_\sigma(t, 0)x^0\|}{\|x^0\|} = \frac{\|x(t; x^0, \sigma)\|}{\|x^0\|} \le ce^{-\lambda t} \tag{3.52}$$

by (3.50).
Now define $\omega(t) := \sigma(t + s)$ for an arbitrary $s > 0$. Obviously, $\omega \in \mathcal{S}$. Let $\psi(\cdot) := \psi(\cdot; x(s; x^0, \sigma), \omega)$ be the solution of $\dot{x}(t) = A_{\omega(t)}x(t)$. Then, on one hand, $\|\psi(t)\| \le ce^{-\lambda t}\|\psi(0)\|$ for all $t \ge 0$ by (3.50). On the other hand, $\psi(t) =$

[16]See for instance [5, Satz 7.5.5].
[17]See [5, Satz 7.5.1].
[18]See [27, Theorem 2.5]. Note that exponential stability is a stronger property than asymptotic stability and implies the latter one, cf. Chapter 2.1.1 in the book.

$x(t+s;x^0,\sigma)$ for all $t \geq 0$. Thus, $\|x(t+s;x^0,\sigma)\| \leq ce^{-\lambda t}\|x(s;x^0,\sigma)\|$ and by substituting $\tau := t+s$ we get

$$\|x(\tau;x^0,\sigma)\| \leq ce^{-\lambda(\tau-s)}\|x(s;x^0,\sigma)\| \tag{3.53}$$

for all $\tau \geq s$. A combination of (3.52) and (3.53) yields (3.51).

There exists another quantity to characterize uniform exponential stability of a switched linear system, the *joint Lyapunov exponent*. For this, consider the set of time t fundamental matrices associated to (3.49) which is given by

$$\mathcal{H}_t := \{\Phi_\sigma(t,0) \mid \sigma \in \mathcal{S}\} \quad \forall t > 0, \qquad \mathcal{H}_0 := \{I\},$$

and the set $\mathcal{H} := \bigcup_{t\in\mathbb{R}_+} \mathcal{H}_t$. $\mathcal{H}$ is a *semigroup*, i. e. for two matrices $S, T \in \mathcal{H}$ it holds that $TS \in \mathcal{H}$. To see this, take $t_1, t_2 > 0$ and $\sigma_1, \sigma_2 \in \mathcal{S}$. Then $\Phi_{\sigma_1}(t_1,0) \in \mathcal{H}_{t_1}$ and $\Phi_{\sigma_2}(t_2,0) \in \mathcal{H}_{t_2}$. Define $\sigma : \mathbb{R}_+ \to \mathcal{M}$ by

$$\sigma(t) = \begin{cases} \sigma_1(t) & 0 \leq t \leq t_1, \\ \sigma_2(t-t_1) & t_1 < t. \end{cases}$$

Then $\sigma \in \mathcal{S}$ and

$$\Phi_\sigma(t_1+t_2,0) = \Phi_{\sigma_2}(t_2,0)\Phi_{\sigma_1}(t_1,0) \in \mathcal{H}_{t_1+t_2} \subset \mathcal{H}. \tag{3.54}$$

We define the growth at time t by

$$\rho_t(\mathcal{M}) := \sup_{\sigma\in\mathcal{S}} \frac{1}{t} \log \|\Phi_\sigma(t,0)\|.$$

Then the *joint Lyapunov exponent* is given by

$$\rho(\mathcal{M}) := \lim_{t\to\infty} \rho_t(\mathcal{M}).$$

Note that due to equivalence of norms on $\mathbb{R}^{n\times n}$ the joint Lyapunov exponent is independent of the chosen operator norm. Now the following proposition holds.

Proposition 3.50. *If the switched linear system* (3.49) *is uniformly stable, then* $\rho(\mathcal{M}) \leq 0$*. The system is uniformly exponentially stable if and only if* $\rho(\mathcal{M}) < 0$*.*

Proof. If the system (3.49) is uniformly stable, then for a fixed $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\sigma \in \mathcal{S}$ the following implication holds:

$$\|x^0\| < \delta \;\Rightarrow\; \|x(t;x^0,\sigma)\| = \|\Phi_\sigma(t,0)x^0\| < \varepsilon \quad \forall t \geq 0.$$

Take the operator norm on $\mathbb{R}^{n\times n}$. For each $\sigma \in \mathcal{S}$ and each $t \geq 0$ there exists a vector $x \in \mathbb{R}^n$ with $\|x\| = 1$ and

$$\|\Phi_\sigma(t,0)\| = \frac{\|\Phi_\sigma(t,0)x\|}{\|x\|}. \tag{3.55}$$

This equality still holds for $\tilde{x} := (\delta/2)x$. Together with the stability assumption we obtain

$$\|x(t;\tilde{x},\sigma)\| = \|\Phi_\sigma(t,0)\tilde{x}\| = \|\Phi_\sigma(t,0)\|\|\tilde{x}\| = \|\Phi_\sigma(t,0)\|\frac{\delta}{2} < \varepsilon,$$

or equivalently,

$$\|\Phi_\sigma(t,0)\| < \frac{2\varepsilon}{\delta} =: M \quad \forall \sigma \in \mathcal{S}, t \geq 0.$$

This yields $\rho_t(\mathcal{M}) \leq \frac{1}{t} \log M$ and therefore $\rho(\mathcal{M}) \leq 0$.
Let us turn to the second statement. If the system (3.49) is uniformly exponentially stable, then there exist constants $M > 0$, $\beta < 0$ such that

$$\|\Phi_\sigma(t,0)\| \leq Me^{\beta t} \quad \forall t \geq 0, \sigma \in \mathcal{S}.$$

Therefore the following inequality holds for all $t \geq 0$.

$$\begin{aligned}\rho_t(\mathcal{M}) &= \sup_{\sigma \in \mathcal{S}} \frac{1}{t} \log \|\Phi_\sigma(t)\| \\ &\leq \frac{1}{t} \log(Me^{\beta t}) = \frac{1}{t} \log(M) + \frac{1}{t} \log(e^{\beta t}) = \frac{1}{t} \log(M) + \beta\end{aligned}$$

Taking the limit $t \to \infty$ we receive $\rho(\mathcal{M}) \leq \beta$ which was supposed to be negative.
Conversely, assume that $\rho(\mathcal{M}) < 0$. Then for every $\rho(\mathcal{M}) < \beta < 0$ there exists a time $T > 0$ such that

$$\forall t \geq T: \ \rho_t(\mathcal{M}) = \sup_{\sigma \in \mathcal{S}} \frac{1}{t} \log \|\Phi_\sigma(t,0)\| < \beta.$$

Equivalenty,

$$\forall t \geq T, \Phi(t,0) \in \mathcal{H}_t: \ \frac{1}{t} \log \|\Phi_\sigma(t,0)\| < \beta,$$

and thus,

$$\forall t \geq T, \Phi(t,0) \in \mathcal{H}_t: \ \|\Phi_\sigma(t,0)\| < e^{\beta t}.$$

By (3.43) we know that there exists a constant $\tilde{M} > 0$ such that $\|\Phi_\sigma(t,0)\| \leq \tilde{M}$ for all $0 \leq t \leq T$ and all $\sigma \in \mathcal{S}$. Set $M = \tilde{M}e^{-\beta T}$. Then

$$\forall t \geq 0, \sigma \in \mathcal{S}: \quad \|\Phi_\sigma(t,0)\| \leq Me^{\beta t},$$

which shows uniform exponential stability. □

Note that $\rho(\mathcal{M}) \leq 0$ is not sufficient to verify uniform stability of a linear switched system. Consider the example

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x. \tag{3.56}$$

The fundamental matrix as well as the solution to the above differential equation are given by

$$\Phi(t,0) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad \varphi(t;x^0) = \Phi(t,0)x^0 = \begin{pmatrix} x_1^0 + tx_2^0 \\ x_2^0 \end{pmatrix}.$$

The column sum norm of the fundamental matrix is $\|\Phi(t,0)\|_1 = 1 + t$ for all $t > 0$. This gives $\rho(\mathcal{M}) = 0$. But the solution $\varphi(\cdot;x^0)$ is obviously unbounded for $t \to \infty$ if $x_2^0 \neq 0$ and therefore the system (3.56) is not stable.
In the analysis of switched linear systems a special type of norm plays a fundamental role as we will see in Chapter 4.

Definition 3.51. *Let $\mathcal{M} \subset \mathbb{R}^{n\times n}$ be a finite set of matrices. A norm v on $\mathbb{R}^n$ is called* extremal *for the associated semigroup $\mathcal{H}$, if for all $x \in \mathbb{R}^n$ and all $t \geq 0$ we have*

$$v(Sx) \leq e^{\rho(\mathcal{M})t} v(x) \quad \forall S \in \mathcal{H}_t. \tag{3.57}$$

Extremal norms have the following notable property. Assume the linear switched system (3.49) is stable and therefore $\rho(\mathcal{M}) \leq 0$ by Proposition 3.50. Thus any solution $\varphi(\cdot; x^0, \sigma)$ of (3.49) is decreasing in the sense that

$$v(\varphi(t; x^0, \sigma)) \leq v(x^0) \quad \forall t \geq 0. \tag{3.58}$$

Since extremal norms are useful for the stability analysis of linear systems (and even nonlinear systems, as we will see later), it is essential to know conditions under which they exist. For later use it will be useful to claim that such a norm is additionally absolute.

Proposition 3.52. *Let $\mathcal{M}$ be a finite set of Metzler matrices. If the associated switched linear system* (3.49) *is uniformly stable and $\rho(\mathcal{M}) = 0$, then there exists an absolute extremal norm for this system.*

Proof. Let $\|\cdot\|$ be an absolute norm on $\mathbb{R}^n$. Then define

$$v(x) := \sup_{S\in\mathcal{H}} \|Sx\| \qquad \forall x \in \mathbb{R}^n_+, \tag{3.59}$$

and set $v(x) := v(|x|)$ for $x \in \mathbb{R}^n$. It follows directly from the definition that if v is a norm, then it is absolute. At first we have to show that $v : \mathbb{R}^n \to \mathbb{R}$, i.e. $v(x) < \infty$ for all $x \in \mathbb{R}^n$. Since the system is uniformly stable, it holds that

$$\forall \varepsilon > 0\ \exists \delta > 0 : \|x\| < \delta \Rightarrow \|Sx\| < \varepsilon \quad \forall S \in \mathcal{H}.$$

Let $x \in \mathbb{R}^n$ be arbitrary and choose $\eta > 0$ such that $\|\eta x\| < \delta$. Then $\|S(\eta x)\| < \varepsilon \Leftrightarrow \|Sx\| < \varepsilon/\eta$ for all $S \in \mathcal{H}$. This shows that $v(x) < \infty$ for every $x \in \mathbb{R}^n$. We continue to verify the norm axioms. (i) Since $I \in \mathcal{H}$ we have $v(x) = \sup_{S\in\mathcal{H}} \|S|x|\| \geq \|x\|$ and $v(0) = 0$. Hence v is positive definite. (ii) For $\alpha \in \mathbb{R}$ it holds that

$$v(\alpha x) = \sup_{S\in\mathcal{H}} \|S|\alpha x|\| = |\alpha| \sup_{S\in\mathcal{H}} \|S|x|\| = |\alpha| v(x),$$

thus v is absolutely homogeneous. (iii) The triangle inequality holds due to

$$\begin{aligned} v(x+y) &= \sup_{S\in\mathcal{H}} \|S|x+y|\| \leq \sup_{S\in\mathcal{H}} \|S|x| + S|y|\| \\ &\leq \sup_{S\in\mathcal{H}} \|S|x|\| + \sup_{S\in\mathcal{H}} \|S|y|\| = v(x) + v(y), \end{aligned}$$

where the first inequality follows from Lemma 3.11 and the fact that an absolute norm is monotone. Note that every $S \in \mathcal{H}$ is nonnegative, see Remark 3.45. Finally, v is extremal since

$$v(Sx) = v(|Sx|) \leq v(S|x|) = \sup_{T\in\mathcal{H}} \|TS|x|\| \leq \sup_{R\in\mathcal{H}} \|R|x|\| = v(x).$$

Here, the first inequality follows due to Lemma 3.11(ii) and monotonicity of $\|\cdot\|$. The second inequality is true since $\mathcal{H}$ is a semigroup and $TS \in \mathcal{H}$ for any two matrices $S, T \in \mathcal{H}$. ☐

The condition $\rho(\mathcal{M}) = 0$ is necessary in the above proposition. If a switched linear system is uniformly exponentially stable and therefore $\rho(\mathcal{M}) < 0$, an extremal norm need not exist. Consider the example

$$\dot{x} = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} x.$$

The fundamental matrix to this system is given by

$$\Phi(t,0) = e^{-t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

The column sum norm of it is $\|\Phi(t,0)\|_1 = e^{-t}(1+t)$ and $\rho_t = \frac{1}{t}\log(e^{-t}(1+t)) = -1 + \frac{1}{t}\log(1+t)$ such that $\rho = -1$. Let $t > 0$ be arbitrary and set

$$P = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \text{ then } P^n = \begin{pmatrix} 1 & nt \\ 0 & 1 \end{pmatrix}.$$

If there existed an extremal norm v for the above system, then for all $x \in \mathbb{R}^n$,

$$v(\Phi(t,0)x) \leq e^{-t}v(x) \quad \Leftrightarrow \quad v(\Phi(t,0)) \leq e^{-t}$$

would hold for the induced operator norm. Consequently, $v(P) \leq 1$. But this inequality is wrong for every operator norm on $\mathbb{R}^{n\times n}$. To see this, take the column sum norm: $\|P^n\|_1 = 1 + nt$ such that $\lim_{n\to\infty} \|P^n\|_1 = \infty$. Since norms on $\mathbb{R}^{n\times n}$ are equivalent, the limit holds for an arbitrary operator norm. Operator norms are submultiplicative such that $v(P^n) \leq v(P)^n$ holds for all $n \in \mathbb{N}$ and hence $\infty = \lim_{n\to\infty} v(P^n) \leq \lim_{n\to\infty} v(P)^n$ which implies $v(P) > 1$. Therefore an extremal norm cannot exist for the above system.
If $\rho(\mathcal{M}) < 0$ for a switched linear system, then we can construct an absolute norm which is “almost extremal”.

Proposition 3.53. *Let $\mathcal{M}$ be a finite set of Metzler matrices. If $\rho(\mathcal{M}) < 0$, then for all $\beta \in \mathbb{R}$ satisfying $\rho(\mathcal{M}) < \beta < 0$ there exists an absolute norm $v : \mathbb{R}^n \to \mathbb{R}$ such that for all $x \in \mathbb{R}^n$ and all $t \geq 0$ we have*

$$v(Sx) \leq e^{\beta t}v(x) \quad \forall S \in \mathcal{H}_t. \tag{3.60}$$

Proof. Let $\|\cdot\|$ be an absolute norm on $\mathbb{R}^n$. Then define

$$v(x) := \sup_{t\geq 0, S\in\mathcal{H}_t} \|e^{-\beta t}Sx\| \qquad \forall x \in \mathbb{R}^n_+,$$

and set $v(x) := v(|x|)$ for $x \in \mathbb{R}^n$. At first we have to show that $v(x) < \infty$ for all $x \in \mathbb{R}^n$. From Proposition 3.50 we know that the switched linear system associated to $\mathcal{M}$ is uniformly exponentially stable and for any $\rho(\mathcal{M}) < \beta < 0$ there exists an $M > 0$ such that

$$\|Sx\| \leq Me^{\beta t}\|x\| \quad \forall x \in \mathbb{R}^n, t \geq 0, S \in \mathcal{H}_t.$$

With that we receive

$$\begin{aligned} v(x) &= \sup_{t\geq 0, S\in\mathcal{H}_t} \|e^{-\beta t}S|x|\| = \sup_{t\geq 0, S\in\mathcal{H}_t} e^{-\beta t}\|S|x|\| \\ &\leq \sup_{t\geq 0} e^{-\beta t}Me^{\beta t}\||x|\| = M\||x|\| < \infty \quad \forall x \in \mathbb{R}^n. \end{aligned}$$

Verification of norm axioms goes analogously to that of the norm given in the proof of Proposition 3.52. By construction, v is absolute. Finally let $\tau \geq 0$ be arbitrary and $S \in \mathcal{H}_\tau$. Then for all $x \in \mathbb{R}^n$,

$$\begin{aligned} v(Sx) &= \sup_{t\geq 0, T\in\mathcal{H}_t} \|e^{-\beta t}T|Sx|\| \leq \sup_{t\geq 0, T\in\mathcal{H}_t} \|e^{-\beta t}TS|x|\| \\ &\leq \sup_{t\geq 0, R\in\mathcal{H}_{t+\tau}} \|e^{-\beta t}R|x|\| = e^{\beta\tau} \sup_{t\geq 0, R\in\mathcal{H}_{t+\tau}} \|e^{-\beta(t+\tau)}R|x|\| \\ &\leq e^{\beta\tau} \sup_{t\geq 0, R\in\mathcal{H}_t} \|e^{-\beta t}R|x|\| = e^{\beta\tau}v(x). \end{aligned}$$

Here the first inequality follows from Lemma 3.11 and monotonicity of $\|\cdot\|$. The second inequality holds since $\mathcal{H}$ is a semigroup. If $S \in \mathcal{H}_\tau$ and $T \in \mathcal{H}_t$ then $TS \in \mathcal{H}_{t+\tau}$, see (3.54). □

We shall now study stability properties of nonlinear systems. The main tool for stability analysis of nonlinear systems are *Lyapunov functions.* Assume $f : D \to \mathbb{R}^n$, defined on an open subset $D \subset \mathbb{R}^n$, is locally Lipschitz continuous and consider the nonlinear time-invariant differential equation

$$\dot{x} = f(x). \tag{3.61}$$

Definition 3.54. *Let $x(\cdot; x^0)$ be the solution to* (3.61) *with initial condition $x(0; x^0) = x^0 \in D$ and $M \subset D$ be a nonempty compact set. A continuous map $V : D \to \mathbb{R}$ is called a* Lyapunov function *for the differential equation* (3.61) *at M on D if*

(i) $x \in M \Rightarrow V(x) = 0$ *and* $x \notin M \Rightarrow V(x) > 0$, *(positive definiteness)*

(ii) $V(x(t; x^0)) < V(x^0)$ *for* $x^0 \notin M$, $t > 0$. *(decrease condition)*

If the inequality in (ii) is not strict, then V is called a weak Lyapunov function*. A (weak) Lyapunov function is called* uniformly unbounded *on D, if for any $\alpha > 0$ there exists a compact set $K \subset D$, $K \neq D$, such that $V(x) \geq \alpha$ for each $x \notin K$.*

A function V that satisfies the first property is called *positive definite away from* M. If $M = \{0\}$ then V is simply called *positive definite.* The next theorem shows the relation between Lyapunov functions and stability.

Theorem 3.55. *Given the assumptions of Definition 3.54, the following propositions hold.*

(i) If there exists a weak Lyapunov function for (3.61) *at M on D, then M is stable.*

(ii) If there exists a Lyapunov function for (3.61) *at M on D, then M is asymptotically stable.*

(iii) If there exists a uniformly unbounded Lyapunov function for (3.61) *at M on D, then M is globally asymptotically stable.*

Proof. All assertions are proved in [9, Chapter V]. (i),(ii),(ii) are Theorems 4.5, 2.2, 2.13 there, respectively. □

The great challenge in stability theory is to find a function that fits as a Lyapunov function for a given system. If we have constructed a candidate $V : D \subset \mathbb{R}^n \to \mathbb{R}$, we have to verify whether the function satisfies properties (i) and (ii) of Definition 3.54. But how could we verify the second property if we didn't know the solutions to our system? The good news is, if V is continuously differentiable, then verification of the decrease condition becomes rather simple. Assume we have given system (3.61) and a continuously differentiable solution $x(\cdot) = x(\cdot; x^0)$ with $x(0; x^0) = x^0$, further a continuously differentiable function $V : D \to \mathbb{R}$. Consider $V(x(\cdot)) : \mathbb{R} \to \mathbb{R}$. Its derivative is given by

$$\frac{d}{dt}V(x(t)) = \nabla V(x(t)) \cdot \dot{x}(t) = \nabla V(x(t)) \cdot f(x(t)). \tag{3.62}$$

On the other side,

$$\int_0^t \frac{d}{d\tau}V(x(\tau))d\tau = V(x(t)) - V(x^0). \tag{3.63}$$

Hence if we can assure

$$\nabla V(x)f(x) < 0 \quad \forall x \notin M, \tag{3.64}$$

then $V(x(t)) - V(x^0) < 0$ for all $x^0 \notin M, t > 0$ and therefore V satisfies property (ii) of Definition 3.54. The same argument holds if "$<$" is replaced by "$\leq$" in both the definition and (3.64). Thus for differentiable functions we have a practical method at hand to verify the decrease condition. But it means a limitation only to take differentiable functions into account when looking for a Lyapunov function. Assume the set M we want to test for stability is the point $\bar{x} = 0$. Now look at the first requirement in Definition 3.54. Any norm on $\mathbb{R}^n$ is positive definite and would therefore perfectly fit this condition! So we would have a large class of functions at our disposal to build Lyapunov functions. But norms are not differentiable, so the decrease condition is hard to verify. Fortunately there exists a theory to solve this problem which uses the concept of generalized gradients we introduced at the end of Section 3.4.
Assume V is locally Lipschitz continuous. Then the Clarke generalized gradient $\partial_C V(x)$ exists for every $x \in D$. Assume further an absolutely continuous function $\varphi : \mathbb{R} \to D$ with $\varphi(0) = x^0 \in D$. Then the composition $V \circ \varphi : \mathbb{R} \to \mathbb{R}$ is absolutely continuous by Proposition 3.9 and thus (3.63) holds for every $t \in \mathbb{R}$ by Theorem 3.10. An interesting question is whether there exists a formula to compute $\frac{d}{dt}V(\varphi(t))$ similar to (3.62). The following proposition, which is stated in the paper [6], answers this question under a certain assumption to which we attend subsequently.

Proposition 3.56. *Let $V : D \subset \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous and nonpathological, and $\varphi : \mathbb{R} \to D$ be absolutely continuous. Then for almost every $t \in \mathbb{R}$ we have*

$$\{\langle p, \dot{\varphi}(t)\rangle \mid p \in \partial_C V(x)\} = \left\{\tfrac{d}{dt}V(\varphi(t))\right\}. \tag{3.65}$$

With this "chain rule" we are able to formulate an analogous criterion to (3.64) to establish (asymptotic) stability of a nonlinear system. Assume φ is a solution to (3.61) with $\varphi(0) = x^0$, then for almost every $t \in \mathbb{R}$ we have $\dot{\varphi}(t) = f(\varphi(t))$

and $\frac{d}{dt}V(\varphi(t)) = \langle p, f(\varphi(t))\rangle$ for all $p \in \partial_C V(\varphi(t))$ due to (3.65). Hence if we assure

$$\big\langle p, f(x)\big\rangle < 0 \quad \forall p \in \partial_C V(x),\ x \neq 0, \tag{3.66}$$

then the left-hand side in (3.63) is negative and thus V satisfies the decrease condition (ii) in Definition 3.54. The same holds if "$<$" is replaced by "$\leq$" in both the definition and (3.66). Lyapunov functions where the approach (3.66) is used to verify the decrease condition are sometimes referred to as *nonsmooth Lyapunov functions*. Now what are nonpathological functions? Baciotti and Ceragioli give a definition of this property in their paper [6], but we shall not impose this notion on the reader. It suffices to know that convex functions and differentiable functions are nonpathological as well as products of the two.
The concept of Lyapunov functions is also suitable to prove stability properties of switched systems. Consider a family of locally Lipschitz continuous functions $f_j : D \to \mathbb{R}^n$, $j \in \{1, \dots, m\}$ on a common open domain $D \subset \mathbb{R}^n$ and the associated systems

$$\dot{x} = f_j(x), \quad j \in \{1, \dots, m\}. \tag{3.67}$$

Given a switching signal $\sigma \in \mathcal{S}$, the related switched system would be

$$\dot{x}(t) = f_{\sigma(t)}(x(t)). \tag{3.68}$$

Theorem 3.57. *Assume that M is a nonempty compact set in the common domain D of all maps f_j, $j = 1, \dots, m$.*

(i) If there exists a function $V : D \to \mathbb{R}$ which is a common weak Lyapunov function for all systems of family (3.67) at M on D, then M is uniformly stable.

(ii) If there exists a function $V : D \to \mathbb{R}$ which is a common Lyapunov function for all systems of family (3.67) at M on D, then M is uniformly asymptotically stable. If moreover V is uniformly unbounded, then M is globally uniformly asymptotically stable.

Proof. The second part can be found in the book [27] of Liberzon. Clarke et al. treat the subject in [11, Chapter 4] in a very general setting, considering differential inclusions and nonsmooth Lyapunov functions. In addition the authors consider stability from a control theoretic point of view. In this context, stability is known under the keyword *strong invariance*. □

At the end of this section we would like to point to a small detail concerning Lyapunov functions. In many books, the requirement (3.64) is replaced by the following. There shall exist a positive definite function $W : D \to \mathbb{R}$ away from M such that for all $x \notin M$,

$$\nabla V(x) f(x) \leq -W(x).$$

For certain switched systems, this slightly stronger assumption compared to (3.64) is essential. Consider the following example which is taken from the book [27, Example 2.1].

Example 3.58. Let $j \in \mathcal{I} := (0, 1]$ and define $f_j(x) = -jx$. This gives a family of systems $\dot{x} = f_j(x)$. The equilibrium $\bar{x} = 0$ is asymptotically stable and

$V(x) = x^2/2$ is a suitable Lyapunov function for each of these systems. The related switched system

$$\dot{x} = -\sigma(t)x$$

has the solutions

$$x(t) = e^{-\int_0^t \sigma(\tau)d\tau} x(0).$$

Hence every every switching signal $\sigma \in L^1([0, \infty), \mathcal{I})$, that is a function for which $\int_0^\infty |\sigma(t)|dt < \infty$, produces a trajectory that does not converge to zero. This happens because the rate of decay of V along the jth constituent system

$$\nabla V(x) f_j(x) = -jx^2$$

gets smaller for small values of j. If $\sigma(\cdot)$ goes to zero too fast, we do not have asymptotic stability. Accordingly there does not exist a positive definite function $W : \mathbb{R} \to \mathbb{R}$, only depending on x and not on j, such that $\nabla V(x) f_j(x) < -W(x)$ for all $j \in \mathcal{I}$. Condition (3.64) is not sufficient in this situation to establish asymptotic stability. However, if we are concerned with only finitely many constituent systems, then (3.64) is adequate. In this case we could define W by $W(x) := -\max\{\nabla V(x) f_j(x) \mid j = 1, \dots, m\}$.

4 Analysis of Switched Epidemiological Models

Now all mathematical instruments are available to investigate the switched epidemiological models we introduced in Section 2.5. The question we want to study is, can we indicate conditions under which an epidemic dies out, independent of the initial number of infectives in a population? That is, the number of infectives converges to zero as t goes to infinity. In reality we are faced with several difficulties when modeling infectious diseases. For example the infectivity of a disease will change during time. This is why we introduced time-variant switched models. Although we are aware of this fact, it is in general not possible to determine exact infectivity rates and the instants at which they change. Instead, we might have several estimates for infectivity rates but don't know at which time they apply. Despite these uncertainties we would like to know if an epidemic vanishes. The mathematical concept to describe such a behavior is that of *uniform asymptotic stability.* Assume we have given different scenarios of spread of a disease, as considered in Section 2.5. If the state $\bar{x} = 0$, where no infectives occur, is globally uniformly asymptotically stable, then the number of infectives will decrease to zero, no matter how many infectives were initially present and no matter when each of the different scenarios exactly applies. Before we get in the analysis of the different epidemiological models, we shall recapitulate the general framework.
Let $n, m \in \mathbb{N}$. For $j = 1, \ldots, m$, the matrices $B_j \in \mathbb{R}^{n\times n}$ are nonnegative. $\Gamma_j, M_j, \Omega_j \in \mathbb{R}^{n\times n}$ and $D_j := \Gamma_j + M_j$ are diagonal matrices with positive entries on the diagonal. For a fixed index j and $i \in \{1, \ldots, n\}$, the diagonal entry of $\Gamma_j, M_j, \Omega_j, D_j$ with index ii is described by $\gamma_i, \mu_i, \omega_i, \alpha_i$, respectively, and $B_j = (b_{ik})$, $i, k = 1, \ldots, n$. The set of all Lebesgue measurable functions mapping from the interval $[0, \infty)$ to the index set $\{1, \ldots, m\}$ is defined by $\mathcal{S} := \{\sigma : \mathbb{R}_+ \to \{1, \ldots, m\} \mid \sigma \text{ is measurable}\}$. We begin our investigation with the SIS model, since it builds the basis for the analysis of SIR and SIRS models.

4.1 The SIS Model

In this section, the state vector $x = (x_1, \ldots, x_n)^T$ is n-dimensional. Given multiple triples of matrices (B_j, Γ_j, M_j), $j = 1, \ldots, m$, as defined above, each of these triples constitutes an SIS epidemiological model

$$\dot{x} = (-D_j + B_j)x - \operatorname{diag}(x)B_j x =: g_j(x). \tag{4.1}$$

If we add a switching signal $\sigma \in \mathcal{S}$, the system

$$\dot{x} = (-D_{\sigma(t)} + B_{\sigma(t)})x - \operatorname{diag}(x)B_{\sigma(t)}x =: g_{\sigma(t)}(x) \tag{4.2}$$

gives rise to a switched SIS epidemiological model. Its linearization is then given by

$$\dot{x} = (-D_{\sigma(t)} + B_{\sigma(t)})x. \tag{4.3}$$

First of all we would like to show that the switched multigroup SIS model (4.2) is well defined in the following way. The state variables $x_i(t)$ in this model represent the fractions $I_i(t)/N_i$ of infectives within a population and should therefore take values in the interval $[0, 1]$ for all $t > 0$ if we assume $x(0) \in [0, 1]^n$. With the vocabulary we developed in Chapter 3 we have to show that the set

$\Pi = [0,1]^n$ is invariant by the system (4.2). It holds even more. The switched SIS system is a positive system, that is the solution $x(t)$ to (4.2) is nonnegative for all times $t > 0$ if we start in the positive orthant, i. e. $x(0) \geq 0$.

Proposition 4.1.

(i) The positive orthant $\mathbb{R}^n_+$ is an invariant set by the system (4.2) for any $\sigma \in \mathcal{S}$.

(ii) The set $\Pi = [0,1]^n$ is invariant by the system (4.2) for any $\sigma \in \mathcal{S}$.

Proof. We profit from the intensive preliminary work done in Chapter 3. Thus we will exploit Proposition 3.24 and apply Theorem 3.44. Given a family of functions $g_j : \mathbb{R}^n \to \mathbb{R}^n$ describing an SIS model, we define the set-valued map $G(x) := \{g_j(x) \mid j = 1, \ldots, m\}$. G is Lipschitz continuous on every compact set $K \subset \mathbb{R}^n$ by Propositions 3.2 and 3.7. Both $\mathbb{R}^n_+$ and Π are closed convex sets. Hence the assumptions of Theorem 3.44 are satisfied.
(i) Let $x \in \mathbb{R}^n_+$. If $x_i = 0$ for an index $1 \leq i \leq n$, then the ith component of g_j is given by

$$g_{j,i}(x) = -\alpha_i x_i + (1 - x_i) \sum_{k=1}^{n} b_{ik} x_k = \sum_{k=1}^{n} b_{ik} x_k \geq 0.$$

Thus, considering Proposition 3.24, $G(x) \subset T_{\mathbb{R}^n_+}(x)$ and Theorem 3.44 yields invariance of $\mathbb{R}^n_+$.
(ii) Let $x \in \Pi$. If $x_i = 1$ for $1 \leq i \leq n$, then

$$g_{j,i}(x) = -\alpha_i x_i + (1 - x_i) \sum_{k=1}^{n} b_{ik} x_k = -\alpha_i \leq 0.$$

Considering part (i) and Proposition 3.24, $G(x) \subset T_{\Pi}(x)$, and Theorem 3.44 yields invariance of Π. □

At the end of Section 3.5.3 we explained that linear switched systems like (4.3) are positive systems. The next lemma tells us, that the solution of the nonlinear system (4.2) is being majorized by the solution of the linearization (4.3).

Lemma 4.2. *Consider system (4.2) and its linearization (4.3). Let $x^0 \in \mathbb{R}^n_+$ and $\varphi(t; x^0, \sigma)$, $\Phi_\sigma(t,0)x^0$ be the solutions of (4.2) and (4.3), respectively. Then*

$$\varphi(t; x^0, \sigma) \leq \Phi_\sigma(t,0)x^0 \quad \forall t \geq 0. \tag{4.4}$$

Proof. The claim follows from the variation-of-constants formula. Define $\varphi(t) := \varphi(t; x^0, \sigma)$, then the solution to the system

$$\dot{x}(t) = (-D_{\sigma(t)} + B_{\sigma(t)})x(t) - \operatorname{diag}(\varphi(t))B_{\sigma(t)}\varphi(t), \quad x(0) = x^0,$$

is given by

$$\varphi(t; x^0, \sigma) = \Phi_\sigma(t,0)x^0 - \int_0^t \Phi_\sigma(t,s) \operatorname{diag}(\varphi(s))B_{\sigma(s)}\varphi(s) ds.$$

Consider the integral. Since the linear system (4.3) is positive, we deduce from Remark 3.45 that $\Phi_\sigma(t,s) \geq 0$ for all $s \in [0,t]$. By definition, $B_{\sigma(s)} \geq 0$. From Proposition 4.1 we know that $\varphi(s) \geq 0$ for all $s \geq 0$. Hence the integral is nonnegative and therefore inequality (4.4) holds. □

The following theorem about stability of SIS models was proved in [1]. We will exploit it to establish similar results on stability for switched SIR and SIRS models.

Theorem 4.3. *Assume that the linear switched system* (4.3) *is uniformly stable. Then the disease free equilibrium* $\bar{x} = 0 \in \mathbb{R}^n$ *of the nonlinear system* (4.2) *is globally*[19] *uniformly asymptotically stable.*

Proof. Let $\mathcal{M} := \{-D_j + B_j \mid j = 1, \ldots, m\}$, then uniform stability of the linear switched system (4.3) implies that the joint Lyapunov exponent $\rho(\mathcal{M}) \leq 0$ by Proposition 3.50. We will treat the cases $\rho(\mathcal{M}) < 0$ and $\rho(\mathcal{M}) = 0$ separately.
(i) Assume $\rho(\mathcal{M}) < 0$, then the linear switched system (4.3) is uniformly exponentially stable and there exists $\beta < 0$ and an absolute norm $v : \mathbb{R}^n \to \mathbb{R}$ such that for all $x^0 \in \mathbb{R}^n_+$ and all $\sigma \in \mathcal{S}$ we have

$$v(\Phi_\sigma(t,0)x^0) \leq e^{\beta t} v(x^0) \quad \forall t \geq 0.$$

From Lemma 4.2 we know that for the solution $\varphi(\cdot; x^0, \sigma)$ of the nonlinear system (4.2) the inequality $0 \leq \varphi(t; x^0, \sigma) \leq \Phi_\sigma(t, 0)x^0$ holds for all $t \geq 0$. If we combine the two inequalities we receive

$$v(\varphi(t; x^0, \sigma)) \leq v(\Phi_\sigma(t,0)x^0) \leq e^{\beta t} v(x^0) \qquad \forall x^0 \in \mathbb{R}^n_+, \sigma \in \mathcal{S}, t \geq 0,$$

since v is monotone. This proves that the equilibrium $\bar{x} = 0$ is globally uniformly asymptotically stable for the switched SIS system (4.2).
(ii) Assume $\rho(\mathcal{M}) = 0$. Then there exists an absolute extremal norm v for the linear system (4.3), that is, for all $x^0 \in \mathbb{R}^n_+$ and all $\sigma \in \mathcal{S}$ we have

$$v(\Phi_\sigma(t,0)x^0) \leq v(x^0) \quad \forall t \geq 0. \tag{4.5}$$

At first we show that v is a weak Lyapunov function for each of the systems (4.1). Consider an initial value $x \in \mathbb{R}^n_+$, $x \neq 0$. Inequality (4.5) tells us that the closed ball $v(x)\mathcal{B}$ with respect to the norm v is an invariant set by the linear system (4.3). If we consider the related differential inclusion $\dot{x} \in \mathcal{A}(x)$ with $\mathcal{A}(x) := \{(-D_j + B_j)x \mid j = 1, \ldots, m\}$, then we deduce from Theorem 3.44 that $\mathcal{A}(x) \subset T_{v(x)\mathcal{B}}(x)$. Consider a subgradient vector $y \in \partial v(x)$. We have seen in Proposition 3.29 that y is a dual vector to x. Hence we use Lemma 3.25 to conclude that $\langle y, z \rangle \leq 0$ for all $z \in T_{v(x)\mathcal{B}}(x)$. This implies

$$\langle y, (-D_j + B_j)x \rangle \leq 0 \tag{4.6}$$

for all $j \in \{1, \ldots, m\}$. Further, as $x_i > 0$ implies $y_i \geq 0$ by Lemma 3.14 it follows that

$$\langle y, -\operatorname{diag}(x)B_j x \rangle \leq 0, \tag{4.7}$$

so that we get

$$\langle y, g_j(x) \rangle = \langle y, (-D_j + B_j)x \rangle + \langle y, -\operatorname{diag}(x)B_j x \rangle \leq 0. \tag{4.8}$$

This shows that v is a weak Lyapunov function for each of the systems (4.1) and therefore $\bar{x} = 0$ is uniformly stable for (4.2) by Theorem 3.57.

[19] Since we are dealing with a positive system, ***globally*** means that the region of attraction of the equilibrium $\bar{x} = 0$ is $A(\bar{x}) = \mathbb{R}^n_+$.

The idea is now to construct a Lyapunov function out of the extremal norm v such that for all subgradients of this new function, inequality (4.8) is always strict. For this we have to investigate under which condition (4.8) may fail to be strict. Due to (4.6) and (4.7) this can only be the case if in both of these equality holds. Assuming equality in (4.7) we get

$$0 = -\langle y, \operatorname{diag}(x) B_j x\rangle = -\sum_{i=1}^{n} y_i x_i (B_j x)_i = -\sum_{x_i, y_i \neq 0} y_i x_i (B_j x)_i$$

Since $x_i > 0$ implies $y_i \geq 0$ we deduce that all summands in the latter sum have to be nonnegative and if the equality holds, then all summands have to vanish. Thus, for each $i \in \{1, \ldots, n\}$, either $y_i = 0$ or $(B_j x)_i = 0$. Plugging this into (4.6) we obtain

$$0 = \langle y, (-D_j + B_j)x\rangle = -\sum_{x_i > 0} y_i x_i \alpha_i + \sum_{x_i = 0} y_i (B_j x)_i.$$

Consider the first sum. We know that $y_i \geq 0$ if $x_i > 0$, and the α_i are positive throughout. Thus all summands are nonnegative. y is a dual vector to x, i. e. $\langle x, y\rangle = \|x\| \neq 0$ by assumption. Consequently there has to be an index i such that $x_i > 0$ and $y_i > 0$. Therefore the first sum together with the minus sign in front is negative. The second sum has to be positive to compensate this. So there exists an index $k \in \{1, \ldots, n\}$ with $x_k = 0$ and $(B_j x)_k > 0$.
Now we want to exploit this property and therefore modify the weak Lyapunov function v. To this end let $\psi : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^\infty$ function with support contained in $(-\infty, 1]$ and so that

$$\psi(z) > 0, \;\; \psi'(z) < 0, \;\; \psi''(z) \geq 0, \quad z \in [0, 1).$$

One might ask if such a function exists. For example $\psi(z) = \exp(\frac{2}{z-1})$, if $z < 1$, and $\psi(z) = 0$ else, satisfies the above requirements.
For $\varepsilon \in (0, 1)$ we define $\psi_\varepsilon(z) := \psi(z + (1-\varepsilon))$ and note that the support of ψ_ε and all of its derivatives is contained in $(-\infty, \varepsilon]$. Moreover, $\eta(\varepsilon) := |\psi'_\varepsilon(0)| = \max_{z \in [0,\varepsilon]} |\psi'_\varepsilon(z)| > 0$ since ψ''_ε is nonnegative on $(0, \varepsilon)$, and $\eta(\varepsilon) \to 0$ if $\varepsilon \to 0$.
Fix $0 < l < L$. We aim to show that for $0 < \varepsilon = \varepsilon(l, L) < 1$ small enough the function

$$V_\varepsilon(x) := v(x)\left(1 + \sum_{i=1}^{n} \psi_\varepsilon(x_i)\right) = v(x)\left(1 + \sum_{x_i < \varepsilon} \psi_\varepsilon(x_i)\right) \tag{4.9}$$

is a Lyapunov function for each of the systems (4.1) on the set

$$\mathbb{L} := \{x \in \mathbb{R}^n_+ \mid l \leq v(x) \leq L\}.$$

As $l > 0$ may be chosen arbitrarily small and $L > 0$ arbitrarily large, this shows global asymptotic stability of $\bar{x} = 0$ for each of the systems (4.1).
We proceed to verify the properties of a Lyapunov function. Obviously, V_ε is locally Lipschitz continuous, positive definite and uniformly unbounded on $\mathbb{R}^n_+$. It remains to show that there exists an $\varepsilon > 0$ such that

$$\forall j \in \{1, \ldots, m\}, \; x \in \mathbb{L}, \; p \in \partial_C V_\varepsilon(x) : \quad \langle p, g_j(x)\rangle < 0. \tag{4.10}$$

We will do this in several steps. As we know from Proposition 3.30, each $p \in \partial_C V_\varepsilon(x)$ is of the form

$$p = y\left(1 + \sum_{x_i<\varepsilon} \psi_\varepsilon(x_i)\right) + v(x)\sum_{x_i<\varepsilon} \psi'_\varepsilon(x_i)e_i, \tag{4.11}$$

where e_i denotes the ith unit vector and $y \in \partial v(x)$, i. e. y is dual to x. For the sake of estimation, set $d_{\max} := \max\{(D_j)_{ii} \mid i = 1,\ldots,n, j = 1,\ldots,m\}$. In the following we assume a fixed $j \in \{1,\ldots,m\}$. Define the function $\zeta : \mathbb{R}^n \to \mathbb{R}$ by

$$\zeta(x) := \max\{\langle y, g_j(x)\rangle \mid y \in \partial v(x)\}.$$

Since $\partial v(x)$ is a nonempty compact set for each x, the function is well defined. From Lemma 3.31 we deduce that ζ is upper semicontinuous and from (4.8) that $\zeta(x) \le 0$ for $x \in \mathbb{R}^n_+$. Fix $x \in \mathbb{L}$. We now distinguish two cases. First, assume $\zeta(x) < 0$. Then it is possible to choose $0 < \varepsilon < 1$ so that

$$Lnd_{\max}\eta(\varepsilon)\varepsilon < |\zeta(x)| \tag{4.12}$$

and this inequality is still true if we replace ε by any $0 < \tilde{\varepsilon} < \varepsilon$. In fact, since $\zeta(\cdot)$ is upper semicontinuous there exists a neighborhood U around x and $\tau > 0$ such that $\zeta(\tilde{x}) < \zeta(x) + \tau < 0$ for every $\tilde{x} \in U$. Thus we can choose ε such that (4.12) holds for all $\tilde{x}$ on this neighborhood. We will need this observation later. Now we obtain for $p \in \partial_C V_\varepsilon(x)$, given by (4.11), that

$$\begin{aligned}
\langle p, g_j(x)\rangle &= \left(1 + \sum_{x_i<\varepsilon} \psi_\varepsilon(x_i)\right)\langle y, g_j(x)\rangle + v(x)\sum_{x_i<\varepsilon} \psi'_\varepsilon(x_i)\langle e_i, g_j(x)\rangle && (4.13)\\
&\le \langle y, g_j(x)\rangle + v(x)\sum_{x_i<\varepsilon} \psi'_\varepsilon(x_i)(-\alpha_i x_i + (1-x_i)(B_j x)_i)\\
&\le \zeta(x) + L\sum_{x_i<\varepsilon} \psi'_\varepsilon(x_i)\big(-\alpha_i x_i + (1-x_i)(B_j x)_i\big)\\
&= \zeta(x) + L\sum_{x_i<\varepsilon} |\psi'_\varepsilon(x_i)|\big(\alpha_i x_i - (1-x_i)(B_j x)_i\big), && (4.14)
\end{aligned}$$

where the last equality holds since $\psi'_\varepsilon(z) \le 0$ for all $z \ge 0$. Further, $|\psi'_\varepsilon(z)| \le \eta(\varepsilon)$ for all $z \in [0,\varepsilon]$ by definition of η. The matrix B_j is nonnegative such that $\psi'_\varepsilon(x_i)(1-x_i)(B_j x)_i \le 0$ for all i and we can drop these terms in the estimation. We continue with

$$\begin{aligned}
&\le \zeta(x) + L\sum_{x_i<\varepsilon} |\psi'_\varepsilon(x_i)|\alpha_i x_i \;\le\; \zeta(x) + L\sum_{x_i<\varepsilon} \eta(\varepsilon)d_{\max}x_i\\
&\le \zeta(x) + Lnd_{\max}\eta(\varepsilon)\varepsilon < 0
\end{aligned}$$

by (4.12). Secondly, consider the case $\zeta(x) = 0$. Then we have seen that there exists an index k such that $x_k = 0$ and $(B_j x)_k > 0$. Hence we can choose $0 < \varepsilon < 1$ such that

$$nd_{\max}\varepsilon < (B_j x)_k, \tag{4.15}$$

and this inequality is still true if we replace ε by any $0 < \tilde{\varepsilon} < \varepsilon$. At this point we remark the same thing as we did in the first case. Since $x \mapsto B_j x$ is a

continuous map, there exists a neighborhood U around x and $\tau > 0$ such that $(B_j\tilde{x})_k > \tau > 0$ for every $\tilde{x} \in U$. Thus we can choose ε such that (4.15) holds for all $\tilde{x}$ on this neighborhood.
Continuing from (4.13) we obtain

$$\begin{aligned}\langle p, g_j(x)\rangle &\leq v(x) \sum_{x_i<\varepsilon} |\psi_\varepsilon'(x_i)|\big(\alpha_i x_i - (1-x_i)(B_j x)_i\big) \\ &= v(x)\Big(\psi_\varepsilon'(0)(B_j x)_k + \sum_{x_i<\varepsilon, i\neq k} |\psi_\varepsilon'(x_i)|\big(\alpha_i x_i - (1-x_i)(B_j x)_i\big)\Big) \\ &\leq v(x)\Big(\psi_\varepsilon'(0)(B_j x)_k + \sum_{x_i<\varepsilon, i\neq k} \eta(\varepsilon) d_{\max}\varepsilon\Big)\end{aligned}$$

and using that $\psi_\varepsilon'(0) = -\eta(\varepsilon)$

$$\begin{aligned}&\leq -v(x)\eta(\varepsilon)\big((B_j x)_k - n d_{\max}\varepsilon\big) \\ &\leq -l\eta(\varepsilon)\big((B_j x)_k - n d_{\max}\varepsilon\big) < 0\end{aligned}$$

by (4.15). What we have shown up to now is that for fixed $j \in \{1,\ldots,m\}$ and $x \in \mathbb{L}$ there exists $0 < \varepsilon < 1$ such that for all $p \in \partial_C V_\varepsilon(x)$ the decrease condition $\langle p, g_j(x)\rangle < 0$ is satisfied. In both cases we argued that there exists a neighborhood U_x around each $x \in \mathbb{L}$ and an appropriate $\varepsilon(U_x) > 0$ such that the strict inequality still holds on this neighborhood. All these neighborhoods give an open cover of $\mathbb{L}$ and since $\mathbb{L}$ is compact we can select a finite cover of it. If we now define ε_j as the minimum of all $\varepsilon(U_x)$ belonging to this finite cover, then $\langle p, g_j(x)\rangle < 0$ holds for all $x \in \mathbb{L}$ and all $p \in \partial_C V_{\varepsilon_j}(x)$. In other words, V_{ε_j} is a Lyapunov function for the jth constituent system on the set $\mathbb{L}$. Since we have finitely many constituent systems, we set $\varepsilon := \min\{\varepsilon_j \mid j = 1,\ldots,m\} > 0$ and V_ε finally serves as a common Lyapunov function for all m systems on the set $\mathbb{L}$ and (4.10) is satisfied for this choice of ε.
Analogously to $\zeta(\cdot)$ we define $Z_j(x) := \max\{\langle p, g_j(x)\rangle \mid p \in \partial_C V_\varepsilon(x)\}$ and observe that this function is upper semicontinuous as well. We argued that $Z_j(x) < 0$ for all $x \in \mathbb{L}$ and since upper semicontinuous functions admit their maximum on compact sets by Lemma 3.32, there exists a constant $c_j < 0$ such that $\max_{x\in\mathbb{L}} Z_j(x) = c_j$. This holds for every j such that if we set $c := \max\{c_j \mid j = 1,\ldots,m\} < 0$, then

$$\forall j \in \{1,\ldots,m\},\ x \in \mathbb{L},\ p \in \partial_C V_\varepsilon(x): \quad \langle p, g_j(x)\rangle \leq c < 0. \tag{4.16}$$

Since we do not have a Lyapunov function for the switched system on the whole positive orthant, we cannot directly apply Theorem 3.57 to establish global uniform asymptotic stability. Anyhow, uniform attractivity may be seen as follows. Let $\delta < 0$ be arbitrary. For any $x^0 \in \mathbb{R}^n_+$ we can choose $L > 0$ such that $v(x^0) \leq L$ and $\varepsilon = \varepsilon(\delta, L)$ such that V_ε is a Lyapunov function on $\mathbb{L} = \{x \in \mathbb{R}^n_+ \mid \delta \leq v(x) \leq L\}$. Let $c < 0$ be the constant for which (4.16) is satisfied. Let $\sigma \in \mathcal{S}$ and $\varphi(t) := \varphi(t; x^0, \sigma)$ be the solution of (4.2) with initial condition x^0. From Proposition 3.9(ii) we deduce that $V_\varepsilon \circ \varphi : \mathbb{R} \to \mathbb{R}$ is absolutely continuous and therefore

$$V_\varepsilon(\varphi(t)) - V_\varepsilon(x^0) = \int_0^t \frac{d}{d\tau} V_\varepsilon(\varphi(\tau)) d\tau \tag{4.17}$$

for every $t > 0$ by Theorem 3.10. Proposition 3.56 says that the equality

$$\frac{d}{dt} V_\varepsilon(\varphi(t)) = \langle p, \dot{\varphi}(t) \rangle, \quad p \in \partial_C V_\varepsilon(\varphi(t)),$$

holds almost everywhere on $\mathbb{R}_+$. From above we know

$$\langle p, g_{\sigma(t)}(\varphi(t)) \rangle \leq c < 0,$$

and thus we obtain

$$V_\varepsilon(\varphi(t)) - V_\varepsilon(x^0) \leq \int_0^t c \, d\tau \;\Leftrightarrow\; V_\varepsilon(\varphi(t)) \leq V_\varepsilon(x^0) - ct$$

as long as $V_\varepsilon(\varphi(t)) \geq \delta$. From the definition of V_ε follows that $v(\varphi(t)) \leq V_\varepsilon(\varphi(t))$ such that we have

$$v(\varphi(t; x^0, \sigma) \leq \delta \qquad \forall t \geq (V_\varepsilon(x^0) - \delta)/c, \; \sigma \in \mathcal{S}.$$

As $\delta > 0$ and $x^0 \in \mathbb{R}^n_+$ are arbitrary, this shows global uniform attractivity of the equilibrium $\bar{x} = 0$ for the switched SIS system (4.2). □

4.2 The SIR and SIRS Model

From now on, the state vector x is again $2n$-dimensional. Given multiple matrices $B_j, \Gamma_j, M_j, \Omega_j$, $j = 1, \ldots, m$, as defined at the beginning of the chapter, and setting $D_j = \Gamma_j + M_j$, then for each j an SIR epidemiological model is given by the differential equation

$$\dot{x} = \begin{pmatrix} -D_j + B_j & 0 \\ \Gamma_j & -M_j \end{pmatrix} x \;-\; \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_j x^1 \\ 0 \end{pmatrix} =: f_j(x). \tag{4.18}$$

If we add a switching signal $\sigma \in \mathcal{S}$, the system

$$\dot{x}(t) = f_{\sigma(t)}(x(t)) \tag{4.19}$$

with right-hand side

$$f_{\sigma(t)}(x) := \begin{pmatrix} -D_{\sigma(t)} + B_{\sigma(t)} & 0 \\ \Gamma_{\sigma(t)} & -M_{\sigma(t)} \end{pmatrix} x \;-\; \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_{\sigma(t)} x^1 \\ 0 \end{pmatrix}$$

gives rise to a switched SIR epidemiological model. The according equations for an SIRS epidemiological model are

$$\dot{x} = \begin{pmatrix} -D_j + B_j & 0 \\ \Gamma_j & -M_j - \Omega_j \end{pmatrix} x \;-\; \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_j x^1 \\ 0 \end{pmatrix} =: h_j(x) \tag{4.20}$$

and

$$\dot{x}(t) = h_{\sigma(t)}(x(t)) \tag{4.21}$$

with right-hand side

$$h_{\sigma(t)}(x) := \begin{pmatrix} -D_{\sigma(t)} + B_{\sigma(t)} & 0 \\ \Gamma_{\sigma(t)} & -M_{\sigma(t)} - \Omega_{\sigma(t)} \end{pmatrix} x \;-\; \begin{pmatrix} \operatorname{diag}\left(x^1 + x^2\right) B_{\sigma(t)} x^1 \\ 0 \end{pmatrix}.$$

As we did earlier for the switched SIS system, we would like to verify that our SIR and SIRS systems are well defined. The first n components of the state variable $x(t) \in \mathbb{R}^{2n}$ represent the infective fractions $I_i(t)/N_i(t)$, whereas the second n components decribe the recovered fractions $R_i(t)/N_i(t)$. So $x(t)$ should be a nonnegative vector for $t \geq 0$. Moreover, since for every time we have $\big(S_i(t) + I_i(t) + R_i(t)\big)/N_i = 1$, we would claim that for each $1 \leq i \leq n$ the inequality $x_i(t) + x_{n+i}(t) \leq 1$ holds for every $t \geq 0$.

Proposition 4.4. *The set*

$$\Sigma := \{x \in \mathbb{R}^{2n}_+ \mid x_i + x_{n+i} \leq 1 \ \forall i = 1, \dots, n\}$$

is invariant by systems (4.19) *and* (4.21) *for any* $\sigma \in \mathcal{S}$.

Proof. The proof is analogous to the one of Theorem 4.1. Given functions $f_j : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ which describe an SIR system, we define the set-valued map $F(x) := \{f_j(x) \mid j = 1, \dots, m\}$. F is Lipschitz continuous on every compact set $K \in \mathbb{R}^{2n}$ by Propositions 3.2. This holds equally for functions h_j representing an SIRS system and setting $H(x) := \{h_j(x) \mid j = 1, \dots, m\}$. Moreover Σ is a closed convex set. Hence the assumptions of Theorem 3.44 are satisfied.
Let $x \in \Sigma$ and fix $j \in \{1, \dots, m\}$. If $x_i = 0$ for an index $1 \leq i \leq n$, then the ith component of f_j is given by

$$f_{j,i}(x) = -\alpha_i x_i + (1 - x_i - x_{n+i}) \sum_{k=1}^{n} b_{ik} x_k = (1 - x_{n+i}) \sum_{k=1}^{n} b_{ik} x_k \geq 0.$$

If $x_{n+i} = 0$ for $1 \leq i \leq n$, then

$$f_{j,n+i}(x) = \gamma_i x_i - \mu_i x_{n+i} = \gamma_i x_i \geq 0.$$

If $x_i + x_{n+i} = 1$ for $1 \leq i \leq n$, then

$$\begin{aligned} f_{j,i}(x) + f_{j,n+i}(x) &= (1 - x_i - x_{n+i}) \sum_{k=1}^{n} b_{ik} x_k \ - \alpha_i x_i + \gamma_i x_i - \mu_i x_{n+i} \\ &= -\gamma_i x_i - \mu_i x_i + \gamma_i x_i - \mu_i x_{n+i} = -\mu_i \leq 0. \end{aligned}$$

Considering Proposition 3.24 we conclude that $F(x) \subset T_\Sigma(x)$, and Theorem 3.44 yields invariance of Σ by the SIR system (4.19). Now consider the right-hand side $h_j(x)$ of an SIRS system. For $1 \leq i \leq n$ the equality $h_{j,i}(x) = f_{j,i}(x)$ holds and therefore $h_{j,i}(x) \geq 0$ if $x_i = 0$. Further, $h_{j,n+i}(x) = f_{j,n+i}(x) - \omega_i x_{n+i}$ such that $h_{j,n+i}(x) = \gamma_i x_i \geq 0$ if $x_{n+i} = 0$. Finally, $h_{j,i}(x) + h_{j,n+i}(x) = -\mu_i - \omega_i x_{n+i} \leq 0$ if $x_i + x_{n+i} = 1$. We conclude that $H(x) \subset T_\Sigma(x)$ which shows invariance of Σ by the system (4.21). □

We will now come to the end of the thesis and present the main result, which continues the work done in the article [1]. There a sufficient criterion has been given to establish global uniform asymptotic stability of the disease free equilibrium of a switched multigroup SIS epidemiological model. We carried out the full proof in Theorem 4.3. It is remarkable that the same criterion suffices to establish a similar result for switched multigroup SIR and SIRS models. The main idea is that the infective components of the SIR(S) state vector are majorized at any time by the state variables of the according SIS system.

Theorem 4.5. *Assume that the linear switched system*

$$\dot{x}^1 = (-D_{\sigma(t)} + B_{\sigma(t)})x^1 \tag{4.22}$$

is uniformly stable in $\bar{x}^1 = 0 \in \mathbb{R}^n$. *Then the disease free equilibrium* $\bar{x} = 0 \in \mathbb{R}^{2n}$ *of the nonlinear SIR system* (4.19), *SIRS system* (4.21) *respectively, is uniformly asymptotically stable and the region of attraction includes* Σ.

Proof. We give the proof for the SIR system. The proof for the SIRS system goes analogous. Let $\xi \in \Sigma$, $\sigma \in \mathcal{S}$ be arbitrary and $\varphi(\cdot) = \varphi(\cdot; \xi, \sigma)$ be the solution of (4.2) with initial condition ξ. We already know by Proposition 4.4 that Σ is invariant and hence $\varphi(t)$ exists for all $t > 0$ and never leaves Σ. (Consider Remark 3.34.) We aim to show that $\varphi^1(t)$ will be majorized by the solution $\psi(\cdot) = \psi(\cdot; \xi^1, \sigma), \psi(0) = \xi^1$, of the according switched SIS system

$$\dot{x}^1 = (-D_{\sigma(t)} + B_{\sigma(t)})x^1 - \operatorname{diag}(x^1)B_{\sigma(t)}x^1 =: g_{\sigma(t)}(x^1) \in \mathbb{R}^n, \tag{4.23}$$

i.e. for all $t > 0$ it holds that $\varphi^1(t) \leq \psi(t)$. We want to make use of Theorem 3.37 about differential inequalities and show in the following that the assumptions of the theorem are satisfied. Fix $j \in \{1, \dots, m\}$ and consider the ith component function of $g_j(x^1)$, which will be denoted by $g_{j,i}(x^1)$. Let $l \in \{1, \dots, n\}, l \neq i$, and choose two vectors $x, y \in [0,1]^n$ with $x_k = y_k$ for $k \neq l$ and $y_l > x_l$. Then

$$\begin{aligned} g_{j,i}(y) - g_{j,i}(x) &= (1 - y_i) \sum_{k=1}^n b_{ik} y_k - \alpha_i y_i - (1 - x_i) \sum_{k=1}^n b_{ik} x_k + \alpha_i x_i \\ &= (1 - x_i) b_{il} (y_l - x_l) \geq 0. \end{aligned}$$

Hence, the function $g_{j,i}(x^1)$ is increasing as claimed in Theorem 3.37 for every j and every i on the set $[0,1]^n$. Switching does not change anything at this property, such that $g_{\sigma(t)}(x^1)$ is increasing as well. Further, the following differential inequality holds almost everywhere on $[0, +\infty)$.

$$\begin{aligned} \dot{\varphi}^1(t) &= (-D_{\sigma(t)} + B_{\sigma(t)})\varphi^1(t) - \operatorname{diag}(\varphi^1(t) + \varphi^2(t))B_{\sigma(t)}\varphi^1(t) \\ &\leq (-D_{\sigma(t)} + B_{\sigma(t)})\varphi^1(t) - \operatorname{diag}(\varphi^1(t))B_{\sigma(t)}\varphi^1(t) \\ &= g_{\sigma(t)}(\varphi^1(t)). \end{aligned}$$

We conclude by Theorem 3.37 that

$$\varphi^1(t) \leq \psi(t) \quad \forall t \in [0, +\infty). \tag{4.24}$$

Now, as already mentioned, $\bar{x}^1 = 0$ is globally asymptotically stable for system (4.23), uniformly across all switching signals $\sigma \in \mathcal{S}$. This yields that the compact set $\hat{X} := \Sigma \cap \{x \in \mathbb{R}^{2n} \mid x^1 = 0\}$ is uniformly asymptotically stable to system (4.19) for all trajectories starting in Σ. It remains to show that $\bar{x}^2 = 0$ is uniformly asymptotically stable for the system

$$\dot{x}^2(t) = -M_{\sigma(t)}x^2(t) + \Gamma_{\sigma(t)}x^1(t). \tag{4.25}$$

Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary norm and use the induced operator norm on $\mathbb{R}^{n \times n}$. For every $j \in \{1, \dots, m\}$, $-M_j$ is Hurwitz and because of diagonal structure, all the M_j's are commuting. Then, by Theorem 3.48 the linear switched system

$$\dot{x}^2 = -M_{\sigma(t)}x^2 \tag{4.26}$$

is globally uniformly exponentially stable. That is, there exist $c, \lambda > 0$ such that for the fundamental matrix of (4.26) it holds that

$$\|\Phi_\sigma(t,s)\| \le ce^{-\lambda(t-s)} \quad \forall t,s \ge 0,\ \sigma \in \mathcal{S}. \tag{4.27}$$

First, we show uniform stability of $\bar{x}^2 = 0$. Let $\varepsilon > 0$. Define $\gamma := \max\{\|\Gamma_j\| \mid j = 1,\dots,m\}$. It is possible to choose ξ such that

$$\|\xi^2\| < \frac{\varepsilon}{2c} \quad \text{and} \quad \|\varphi^1(t;\xi,\sigma)\| < \frac{\lambda\varepsilon}{2\gamma c} \quad \text{for all } t \ge 0,\ \sigma \in \mathcal{S}$$

by the first part of the proof. Now the variation of constants formula (3.33) yields

$$\begin{aligned}
\|\varphi^2(t;\xi,\sigma)\| &\le \|\Phi_\sigma(t,0)\|\|\xi^2\| + \int_0^t \|\Phi_\sigma(t,s)\|\|\Gamma_{\sigma(s)}\|\|\varphi^1(s)\|ds \\
&\le ce^{-\lambda t}\|\xi^2\| + \int_0^t ce^{-\lambda(t-s)}\gamma\|\varphi^1(s)\|ds \\
&< ce^{-\lambda t}\frac{\varepsilon}{2c} + c\gamma\frac{\lambda\varepsilon}{2\gamma c}\int_0^t e^{-\lambda(t-s)}ds \\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\lambda\frac{1}{\lambda}(1-e^{-\lambda t}) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

for an arbitrary $\sigma \in \mathcal{S}$ and all $t \ge 0$. Hence $\bar{x}^2 = 0$ is uniformly stable. For uniform attractivity, let $\eta > 0$ and $\xi \in \Sigma$ be arbitrary. Define $\alpha := \max_{t\ge 0}\|\varphi^1(t)\|$. Again,

$$\begin{aligned}
\|\varphi^2(t;\xi,\sigma)\| &\le \|\Phi_\sigma(t,0)\|\|\xi^2\| + \int_0^t \|\Phi_\sigma(t,s)\|\|\Gamma_{\sigma(s)}\|\|\varphi^1(s)\|ds \\
&\le ce^{-\lambda t}\|\xi^2\| + \int_0^T ce^{-\lambda(t-s)}\gamma\alpha ds + \int_T^t ce^{-\lambda(t-s)}\gamma\|\varphi^1(s)\|ds
\end{aligned}$$

for any $t > T > 0$. Since $\varphi^1(\cdot)$ converges uniformly to 0, we can choose T such that

$$\|\varphi^1(t)\| < \frac{\lambda\varepsilon}{2c\gamma} \quad \forall t \ge T,\ \sigma \in \mathcal{S}.$$

With that, we continue

$$\begin{aligned}
\|\varphi^2(t;\xi,\sigma)\| &< ce^{-\lambda t}\|\xi^2\| + c\gamma\alpha\frac{1}{\lambda}(e^{\lambda T}-1)e^{-\lambda t} + \frac{\lambda\varepsilon}{2c\gamma}c\gamma\frac{1}{\lambda}(1-e^{-\lambda(t-T)}) \\
&\le \underbrace{\left(c\|\xi^2\| + c\gamma\alpha\frac{1}{\lambda}(e^{\lambda T}-1)\right)}_{=:\,C} e^{-\lambda t} + \frac{\varepsilon}{2}.
\end{aligned}$$

Finally, choose $T_0 > T$ such that $Ce^{-\lambda T_0} < \varepsilon/2$. Then

$$\|\varphi^2(t;\xi,\sigma)\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $t \ge T_0$ and all $\sigma \in \mathcal{S}$. This shows uniform attractivity of $\bar{x}^2 = 0$ and we are done.
To see that $\bar{x} = 0$ is also uniformly asymptotically stable for the SIRS system (4.21), note that $h_{j,i}(x) = f_{j,i}(x)$ for $1 \le i \le n$. So the first part of the proof remains unchanged. For the second part, replace $M_{\sigma(t)}$ by $M_{\sigma(t)} + \Omega_{\sigma(t)}$ in (4.25) and (4.26). The ensuing argumentation stays the same. □

5 Conclusion

In the preceding work we saw how modern mathematics may be used to model the spread of epidemics as well as to analyze these models in order to gain deeper insight to the dynamics of infectious diseases.
We began with two standard models in mathematical epidemiology, the SIR and SIS model, and also considered a hybrid of both, the SIRS model. These basic models are already well understood and Theorem 2.1 gives a comprehensive answer to the question how an SIR system behaves depending on its characteristic parameters. There exist two equilibria, the *disease-free equilibrium* where no infectives occur, and the *endemic equilibrium* where a constant positive fraction of infectives appears. Each solution either approaches the disease-free or the endemic equilibrium, only depending on infection, recovery and mortality rates and not on the specific initial numbers of infectives and susceptibles.
These simplistic models ignore important aspects of an epidemic. In reality someone is faced with individual infection rates for different groups within a population and in general these rates will not be constant but change over time. This is the reason why epidemiological models with multiple subgroups were introduced and time-variant parameters were considered by using switched systems.
The more realistic, the more complex a model gets, and it is a challenging task to analyze the behavior of solutions of a switched multigroup epidemiological model. We established many mathematical tools to approach this problem in Chapter 3. The seminal notions and theorems arise from convex and nonsmooth analysis as well as from the theory of differential inclusions. They made it possible to proof Theorem 4.3 where a sufficient condition – uniform stability of the linearized system – was stated to establish global uniform asymptotic stability of the disease-free equilibrium of a switched multigroup SIS model. Based on this result which originates from the article [1] we continued to proof analogous statements for switched multigroup SIR and SIRS models.
Several questions emerge from this point. We only considered stability properties of the disease-free equilibrium. In another scenario persistence could occur, where the numbers of infectives do not converge to zero. It is interesting to know whether there exists an endemic equilibrium and, if this is the case, what stability properties it possesses. Another aspect arises from a control theoretic point of view. Even if the disease-free equilibrium is not uniformly asymptotically stable, is it possible to indicate a switching law that forces trajectories to converge to zero? In real life this would be helpful since it is desired to have measures available that erase an infectious disease from the population. Some of these problems for switched SIS models are treated in [1], but they are still open for switched SIR(S) models.

References

[1] Mustapha Ait Rami, Vahid Bokharaie, Oliver Mason, and Fabian R. Wirth. Stability Criteria for SIS Epidemiological Models under Switching Policies. 2013. http://arxiv.org/abs/1306.0135.

[2] Jürgen Appell. *Analysis in examples and counterexamples. An introduction to the theory of real functions. (Analysis in Beispielen und Gegenbeispielen. Eine Einführung in die Theorie reeller Funktionen.).* Berlin: Springer, 2009.

[3] Jean-Pierre Aubin. *Viability theory.* Boston, MA etc.: Birkhäuser, 1991.

[4] Jean-Pierre Aubin and Hélène Frankowska. *Set-valued analysis.* Boston etc.: Birkhäuser, 1990.

[5] Bernd Aulbach. *Ordinary differential equations. (Gewöhnliche Differenzialgleichungen.) 2nd ed.* Heidelberg: Elsevier/Spektrum Akademischer Verlag, 2004.

[6] Andrea Bacciotti and Francesca Ceragioli. Nonpathological Lyapunov functions and discontinuous Carathéodory systems. *Automatica*, 42(3):453–458, 2006.

[7] Norman T.J. Bailey. *The mathematical theory of infectious diseases and its applications. 2nd ed.* Hafner, New York, 1975.

[8] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences.* Philadelphia, PA: SIAM, 1994.

[9] N.P. Bhatia and G.P. Szegő. *Stability theory of dynamical systems. Reprint of the 1970 edition.* Berlin: Springer, 2002.

[10] F. H. Clarke. Generalized gradients and applications. *Trans. Am. Math. Soc.*, 205:247–262, 1975.

[11] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth analysis and control theory.* New York, NY: Springer, 1998.

[12] Earl A. Coddington and Norman Levinson. *Theory of ordinary differential equations.* McGraw-Hill, 1955.

[13] D.J. Daley and J. Gani. *Epidemic modelling: an introduction.* Cambridge: Cambridge University Press, 1999.

[14] O. Diekmann and J.A.P. Heesterbeek. *Mathematical epidemiology of infectious diseases. Model building, analysis and interpretation.* Chichester: Wiley, 1999.

[15] Jürgen Elstrodt. *Measure and integration theory. (Maß- und Integrationstheorie.) 6th corrected ed.* Berlin: Springer, 2009.

[16] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions.* Boca Raton: CRC Press, 1992.

[17] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Rev.*, 42(4):599–653, 2000.

[18] Adrian T. Hill and Achim Ilchmann. Exponential stability of time-varying linear systems. *IMA J. Numer. Anal.*, 31(3):865–885, 2011.

[19] Diederich Hinrichsen and Anthony J. Pritchard. *Mathematical systems theory. I. Modelling, state space analysis, stability and robustness. 1st ed., corrected printing.* Berlin: Springer, 2010.

[20] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, 1990.

[21] Matt J. Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals.* Princeton, NJ: Princeton University Press, 2008.

[22] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings Royal Soc. London (A)*, 115:700–721, 1927.

[23] Konrad Königsberger. *Analysis 1. 6., durchgesehene Aufl.* Berlin: Springer, 2004.

[24] Konrad Königsberger. *Analysis 2. 5., korrigierte Aufl.* Berlin: Springer, 2004.

[25] A. Krämer and R. Reintjes. *Infektionsepidemiologie.* Springer, Berlin, 2003.

[26] V. Lakshmikantham and S. Leela. *Differential and integral inequalities. Theory and applications. Vol. I: Ordinary differential equations (Mathematics in Science and Engineering. Vol. 55).* Academic Press, Inc., New York, 1969.

[27] Daniel Liberzon. *Switching in systems and control.* Boston, MA: Birkhäuser, 2003.

[28] Daniel Liberzon. Switched systems. In D. Hristu-Varsakelis and W.S. Levine, editors, *Handbook of Networked and Embedded Control Systems*, pages 559–574. Birkhäuser, Boston, 2005.

[29] James D. Murray. *Mathematical biology. Vol. 1: An introduction. 3rd ed.* New York, NY: Springer, 2002.

[30] Martin L. Puterman. *Markov decision processes: discrete stochastic dynamic programming. Reprint of the 1994 hardback ed.* Hoboken, NJ: John Wiley & Sons, reprint of the 1994 hardback ed. edition, 2005.

[31] R.Tyrrell Rockafellar. *Convex analysis.* Princeton, NJ: Princeton University Press, 1997.

[32] Winfried Schirotzek. *Nonsmooth analysis.* Berlin: Springer, 2007.

[33] Georgi V. Smirnov. *Introduction to the theory of differential inclusions.* Providence, RI: AMS, American Mathematical Society, 2002.

[34] Eduardo D. Sontag. *Mathematical control theory. Deterministic finite dimensional systems. 2nd ed.* New York, NY: Springer, 1998.

[35] J. Szarski. *Differential inequalities. 2nd ed., revised.* PWN, Warszawa, 1967.

[36] Dirk Werner. *Functional analysis. (Funktionalanalysis.) 6th corrected ed.* Berlin: Springer, 2007.

[37] Fabian Wirth. The generalized spectral radius and extremal norms. *Linear Algebra Appl.*, 342(1-3):17–40, 2002.

Ich versichere hiermit, dass ich die Masterarbeit selbstständig verfasst und nur die in Text und Literaturverzeichnis angeführten Quellen und Hilfsmittel benutzt habe. Ich habe die Arbeit bis zu diesem Zeitpunkt keiner anderen Prüfungsbehörde vorgelegt.

..................................., den	....	....
Ort	Datum	Unterschrift des Verfassers