Dynamical systems

January 10, 2016

1 Introduction

In this text the term 'dynamical system' means nothing else than a system of ordinary differential equations. Why is this term used here? The aim is to emphasize that the ordinary differential equations and their solutions are considered here in a particular way. A mathematical problem which ordinary differential equations pose is that of solving them explicitly. This means looking for formulae for the solutions in terms of combinations of certain 'elementary functons' such as powers, the exponential function, sine etc. It is known that this is not possible for all equations and one way of going further is to introduce new functions (for instance elliptic functions) which themselves are nothing other than solutions of special ordinary differential equations which have already been studied in detail. It is, however, the case that most ordinary differential equations cannot be solved explicitly in any useful sense.

What other options are there? Many people think (in particular many non-mathematicians) that there are only two possibilities. The first is to leave rigorous mathematics behind and to go over to heuristic approximation procedures. It is supposed, for instance, that in a particular application a certain quantity is small and the original equation is replaced by one in which this quantity is set to zero. The second method is to discretize the equations and solve the resulting discrete equations on a computer. In both cases an approximation has been introduced. From a mathematical point of view one system of equations has been replaced by a second and it must be established, what the solutions of the two systems have to do with each other. These methods can often give good results. What is to emphsized here is that there is also another option. (Often the best strategy is to combine all three methods.)

What is the third option? It is to prove mathematically rigorous statements about the qualitative behaviour of solutions. In this context it is often better to consider the relations between solutions rather than studying each solution on its own. It is also to consider the problem geometrically. A system of ordinary differential equations consists of equations of the form

$$\frac{dx_i}{dt} = f_i(t, x_j) \tag{1}$$

for real-valued functions $x_i(t), 1 \leq i \leq m$ and real-valued functions f_i of $m+1$ variables. Together the f_i define a mapping with values in $\mathbb{R}^m$ which is denoted by f. If f does not depend on t then the system is called autonomous. In this course we will mainly concentrate on the case of autonomous systems since the point of view of dynamical systems is particularly helpful in that case. If desired, the non-autonomous case can be reduced to the autonomous one by considering the extended system

$$\frac{dx_i}{dt} = f(y, x_j), \quad \frac{dy}{dt} = 1 \tag{2}$$

The existence of a solution $x(t)$ of the original system with $x(t_0) = x_0$ is equivalent to the existence of a solution $(x(t), y(t))$ of the extended system with $x(t_0) = x_0$ and $y(t_0) = t_0$. From a geometrical standpoint the x_i are considered as the components of a point in $\mathbb{R}^m$ and f_i are considered as the components of a vector field. The solutions of the equations are the integral curves of the vector field. It is then natural in the autonomous case to write the equations (1) as the vector-valued equation

$$\frac{dx}{dt} = f(x) \tag{3}$$

The function f is defined on an open subset G of $\mathbb{R}^m$. Initial conditions $x_i(t_0) = a_i$ of the form for a solution $x_i(t)$ mean that the solution is at the point with coordinates a_i at the time t_0. This way of looking at things is typical for the point of view of dynamical systems. The mathematical objects are the same. It is just that we talk and think about them in a different way in order to try to mobilize another geometric intuition. Here we only consider first order systems because it is easy to reduce a system of order k to a first order system by introducing the derivatives up to order $k-1$ as new variables.

To be able to work with solutions which are not explicit it is necessary to be able to fix which solution is being considered. This means that it is necessary to know how many solutions there are and how they can be parametrized. The usual way of doing this is the initial value problem and for this reason the next section treats this subject.

The concept 'dynamical system' is often used in a wider sense where more general types of evolution equations are allowed. These could, for instance, be partial differential equations or delay equations. Here an analogy is used between ordinary differential equations and these other equations where the Eucldean space is replaced by an infinite dimensional function space. There are importent differences between these classes of equations and this can make the analogies dangerous. On the hand there are a lot of similarities which can make these analogies very useful. The following text is restricted to the case of ordinary differential equations.

2 The initial value problem

The initial value problem for ordinary differential equations belongs to the subject matter of basic analysis courses at university. This theme will nevertheless be treated here in order to recall the methods and to provide a point of departure for generalizations. The fundamental result on existence and uniqueness is

Theorem 1 Let f be a continuous function defined on the set

$$\{(t,x) \in \mathbb{R} \times \mathbb{R}^m : t_0 \leq t \leq t_0 + a, |x - x_0| \leq b\} \tag{4}$$

with values in $\mathbb{R}^m$ which satisfies a Lipschitz condition with respect to x. Let M be a bound for $|f|$ and $\alpha = \min\{a, b/M\}$. The the equation $\frac{dx}{dt} = f(t,x)$ has a unique solution on the interval $[t_0, t_0 + \alpha]$ with $x(t_0) = x_0$.

Proof A sequence of functions will be defined which in the end converges to the desired solution. Let $x_0(t) = x_0$. If a continuous function x_n is defined on the interval $[t_0, t_0 + \alpha]$ and satisfies $|x_n(t) - x_0| \leq b$ let

$$x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, x_n(s))ds. \tag{5}$$

These conditions define a sequence $\{x_n\}$ of continuous functions on $[t_0, t_0 + \alpha]$. They satisfy the inequalities

$$|x_{n+1}(t) - x_0| \leq \int_{t_0}^{t} |f(s, x_n(s))|ds \leq M\alpha \leq b. \tag{6}$$

It can be shown by induction that

$$|x_{n+1}(t) - x_n(t)| \leq \frac{MK^n(t - t_0)^{n+1}}{(n+1)!} \tag{7}$$

for all n, where K is a Lipschitz constant for f. The start of the induction is clear. For the inductive step we use the fact that for $n \geq 1$

$$\begin{aligned} |x_{n+1}(t) - x_n(t)| &\leq \int_{t_0}^{t} |f(s, x_n(s)) - f(s, x_{n-1}(s)|ds \\ &\leq K \int_{t_0}^{t} |x_n(s) - x_{n-1}(s)|ds \leq \frac{MK^n(t - t_0)^{n+1}}{(n+1)!}. \end{aligned} \tag{8}$$

It follows that the series $x_0 + \sum_{n=0}^{\infty}(x_{n+1}(t) - x_n(t))$ converges uniformly and we define $x(t)$ to be this sum. It is then possible to pass to the limit $n \to \infty$ in the integral equation with the result that

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds. \tag{9}$$

Hence $x(t)$ is a solution of the differential equation with the desired initial value. To show uniqueness let $y(t)$ be any solution of the equation with the given initial value. Then it can be shown by induction that

$$|x_n(t) - y(t)| \leq \frac{MK^n(t-t_0)^{n+1}}{(n+1)!}. \tag{10}$$

The right hand side of this inequality converges to zero for $n \to \infty$ and it follows that $x(t) = y(t)$.

It has now been shown that the solution is unique on the interval on which it was constructed in the theorem. In fact the solution of (3) is unique with a given initial value is unique on any interval where it is defined, provided f is locally Lipschitz. Consider two solutions x and y on an interval $[t_0, t_1)$ with $x(t_0) = y(t_0)$. Here t_1 is allowed to be infinity. Let Sei t_* be the supremum of all numbers t with the property that $x = y$ on the interval $[t_0, t)$. Because of local uniqueness $t_* > 0$. The solutions x and y are equal on the interval $[t_0, t_*)$ and therefore, by continuity, assuming $t_* < \infty$, on the interval $[t_0, t_*]$. If $t_* < \infty$ we can consider the initial value problem with initial time t_* and initial condition $x(t_*) = y(t_*)$. It can be concluded from local uniqueness that $x = y$ on an interval of the form $[t_*, t_* + \alpha]$, a contradiciton to the definition of t_*. Here the case $t \geq t_0$ was considered but a similar argument holds for $t \leq t_0$.

Without the assumption of a local Lipschitz condition there is still a local existence result but this will not be proved here. This is Peano's theorem (cf. [4], Abschnitt II.2). Uniqueness no longer holds as can be seen from simple examples like $\frac{dx}{dt} = x^\alpha$, $0 < \alpha < 1$. Even when a local Lipschitz condition holds the local solution cannot in general be extended to a global solution, as can be seen in the simple example $\dot{x} = x^2$. The solution with $x(0) = 1$ is $\frac{1}{1-t}$ and it only exists up to $t = 1$. Because of uniqueness it is possible to define the maximal interval of existence of a solution with a given initial value. This interval can be characterized by a continuation criterion.

Before this statement is proved some metric properties of subsets of $\mathbb{R}^m$ will be discussed. Let A be a closed subset of $\mathbb{R}^m$. For a point $x \in \mathbb{R}^m$ let $d(x, A) = \inf_{y \in A} d(x, y)$. For a fixed subset A the function $d(x, A)$ is continuous, as will now be proved. Let x_1 and x_2 be points of $\mathbb{R}^m$ and $\epsilon > 0$. The exists a point $y \in A$ with $d(x_2, y) \leq d(x_2, A) + \frac{\epsilon}{2}$. It follows from the triangle inequality that

$$d(x_1, y) \leq d(x_1, x_2) + d(x_2, A) + \frac{\epsilon}{2} \leq d(x_2, A) + \epsilon \tag{11}$$

provided $d(x_1, x_2) \leq \frac{\epsilon}{2}$. Thus in this case $d(x_1, A) - d(x_2, A) \leq \epsilon$. The same argument shows that $d(x_2, A) - d(x_1, A) \leq \epsilon$ and the continuity of $d(x, A)$ has been proved. If $x \notin A$ the inequality $d(x, A) > 0$ holds. Because otherwise there would exist a sequence $x_n \in A$ with $d(x_n, x) \to 0$ as $n \to \infty$ and, since A is closed, this would imply that $x \in A$, a contradiction. If A_1 and A_2 are subsets of $\mathbb{R}^m$ let $d(A_1, A_2) = \inf_{y_1 \in A_1, y_2 \in A_2} d(y_1.y_2)$. This expression is symmetric in its arguments but can also be written in the form $d(A_1, A_2) = \inf_{y \in A_2} d(y, A_1)$. Suppose that A is closed, that K is compact and that $A \cap K = \emptyset$. Then $d(K, A)$

is the infimum of a continuous function on the compact set K and must be the minimum. It follows that $d(K, A) > 0$.

Consider the equation (3) where the function f is defined on an open subset G of $\mathbb{R}^m$. Let (t_-, t_+) be the maximal interval of existence of a solution with initial condition $x(t_0) = x_0 \in G$. If t_+ is finite then the solution leaves each compact subset K of G as $t \to t_+$. To prove this statement, suppose that it was false. Then there would exist a sequence $\{t_n\}$ with $t_n \to t_+$ and $x(t_n) \in K$ for all n. Let $\epsilon = d(K, \mathbb{R}^m \setminus G)$. Then $\epsilon > 0$. Let G_1 be the set of all points satisfying $d(x, K) < \frac{\epsilon}{2}$. G_1 is an open subset of G and its closure $\bar{G}_1$ is compact and contained in G. The distance between K and the complement of G_1 is no smaller than $\frac{\epsilon}{2}$. Hence the closed ball of radius $\frac{\epsilon}{2}$ about $x(t_n)$ is contained in G_1 for all n. At the same time f is bounded on G_1 by a positive number $M \geq 1$. By the local existence theorem there exists a solution on the interval $[t_n, t_n + \epsilon/2M]$ which has the same initial value as the original solution. It suffices to choose n large enough that $\epsilon/2M > t_n - t_*$ to obtain a contradiction to the defintion of t_*. This proof has been carried out for the case of an autonomous system. A corresponding result holds for a system which is not necessarily autonomous and the proof is similar.

3 An example: the fundamental system of virus dynamics

In this course the general theory is accompanied by examples coming from scientific applications. In this section one such example is introduced. The dynamical system considered, the fundamental system of virus dynamics, is used to model the spread of virus infections in the body. The results obtained with this model have contributed to important adavnces in medicine. The biological background will now be sketched briefly. More background information can be found in [8]. The disease AIDS was discovered in the 1980s and after a couple of years the virus causing the disease, HIV, had been isolated. The original optimism that it would soon be possible to cure the disease turned out to be unjustified. After an infection with HIV and a short initial phase with flu-like symptoms the disease normally causes no symptoms for a long time (about 10 years). Only then does AIDS become manifest. It was believed for a long time that in this symptom-free period the virus had become dormant for some reason but in the meantime it has become clear that this idea was false. Around 1995 it was realized that in these ten years a dynamical process takes place suring which huge numbers of virions are produced. Mathematical models played an important role in coming to this insight. It was in this context that the modern combination therapies for AIDS (HAART) were developed. Today the disease cannot be cured but its dangerous effects can be suppressed for a unlimited time with suitable drugs.

The model which will now be introduced, although it was used in AIDS research, has no special relation to HIV and can be used to model many diseases

which are caused by viruses. There are three variables. The population of cells which are not infected with the virus is denoted by x, the population of infected cells by y and the number of virions by v. The equations are

$$\dot{x} = \lambda - dx - \beta xv, \tag{12}$$

$$\dot{y} = \beta xv - ay, \tag{13}$$

$$\dot{v} = ky - uv. \tag{14}$$

Here the dot stands for the derivative with respect to t and the quanities λ, d, β, k and u are positive constants. Because of their interpretation the quantities x, y und v should be positive although the system of equations is well-defined and smooth on the whole of $\mathbb{R}^3$. The phenomena which correspond to the different parameters are the following. New cells are produced by cell devision (λ), uninfected cells die (d), virions infect cells (β), infected cells die (a), virions are produced by infected cells (k), virions are eliminated (u). An intervention of the immune system is not taken into account. In the case of HIV the cells involved are themselves immune cells (white blood cells) but this fact plays no role in the model.

To be precise, the quantity $v$is the number of virions outside the cells. The model neglects the fact that when a cell is infected the number of virions outside the cells is reduced by one. For this reason the equation for $\dot{v}$ should contain an extra term $-\beta xv$. It is argued, however, that this term is small in comparison to other terms in the same equation, so that it is justified to omit it. From a mathematical point of view including this effect leads to a new system which we call the 'modified fundamental model of virus dynamics'. The additional Term can be written as $-\delta\beta xv$, where the Parameter δ takes the value zero or one.

The functions on the right hand side of (12)-(14) are evidently locally Lipschitz and therefore the theorem on local existence and uniqueness can be applied to this system. Because of the intepretation of the unknowns initial values are considered which are positive (i.e. x, y und v are positive) and it is expected that the solutions stay positive. The proof of this fact is not obvious und will now be presented.

Lemma 1 Let $(x(t), y(t), v(t))$ be a solution of the system (12)-(14) on an interval $[t_0, t_1)$ with $x(t_0) = x_0$, $y(t_0) = y_0$ und $v(t_0) = v_0$. If x_0, y_0 and v_0 are positive then $x(t)$, $y(t)$ and $v(t)$ are positive for all $t \in [t_0, t_1)$.

Proof Call the variables x_i. If there is an index i and a time t for which $x_i(t) = 0$ let t_* be the infimum of all such t for any i. Then the restriction of the solution to the interval $[t_0, t_*)$ is positive and $x_i(t_*) = 0$ for a certain value of i. The equation for x_i can be written in the form $\dot{x}_i = -x_i f(x) + g(x)$, where $g(x)$ is non-negative. As a consequence $\dot{x}_i \geq -x_i f(x)$ and $\frac{d}{dt}(\log x_i) \geq -f(x) \geq -C$ for a positive constant C. To show this the fact is used that the solution remains in a compact set. It follows that $x_i(t_*) \geq x_i(t_0)e^{-(t_*-t_0)} > 0$, a contradiction.

It will now be shown, with the help of the continuation criterion, that all solutions of (12)-(14) with postive initial data exist globally in time in the future. To prove this it is enough to show that all variables a bounded above

on an arbitrary finite interval $[t_0, t)$. Taking the sum of the first two equations shows that $\frac{d}{dt}(x+y) \leq \lambda$ and hence that $x(t)+y(t) \leq x(t_0)+y(t_0)+\lambda(t-t_0)$. Thus x and y are bounded on any finite interval. The third equation then shows that $v(t)$ cannot grow faster than linearly and is also bounded on any finite interval. It is possible to show that

$$x(t)+y(t) \leq C_1 = \max\left\{x(t_0)+y(t_0), \frac{\lambda}{\min\{a,d\}}\right\}, \tag{15}$$

$$v(t) \leq C_2 = \max\left\{v(t_0), \frac{kC_1}{u}\right\}. \tag{16}$$

This means in particular that the solution is globally bounded. These claims will now be proved. Let us suppose that there exists a time t with $x(t)+y(t) = \alpha > C_1$. Let t_* be the infimum of those times where this inequality holds for a fixed choice of α. Then $\dot{x}(t_*)+\dot{y}(t_*) \geq 0$. On the other hand the equations (12) and (13) imply that this quantity is negative, a contradiction. This proves (15). The inequality (16) can be proved by the same method, using the bound for $x+y$ already obtained. With the same method it can be shown that the solutions of the modified system with ($\delta = 1$) are bounded and that the unknowns satisfy the same bounds as in the case $\delta = 0$. If the initial data satisfy the inequalities $x_0+y_0 \leq \frac{\lambda}{\min\{a,d\}}$ and $v_0 \leq \frac{k\lambda}{u\min\{a,d\}}$ then the whole solution must satisfy this inequality. This means that we have identified an invariant subset.

4 Dependence on initial data and parameters

We consider a solution $x(t) = \phi(t_0, t, x_0)$ of the equation $\dot{x} = f(t,x)$ with initial condition $x(t_0) = x_0$. The mapping ϕ is called the flow of the dynamical system. This solution is defined on a maximal interval of existence (t_-, t_+), where t_- and t_+ may depend on t_0 and x_0. In this section we consider the question of the continuity and differentiability of the function ϕ. In other words we are concerned with the question whether the solution of the initial value problem depends continuously or smoothly on initial data. We also consider the more general case of an equation $\dot{x} = f(t,x,z)$, where the coordinates of the point $z \in \mathbb{R}^k$ are considered as parameters.

Questions of dependence on initial data and parameters can often be related to each other by means of transformations of variables. Consider first the initial value problem without parameters. By introducing $\tilde{t} = t - t_0$ and $\tilde{x} = x - x_0$ we get the new problem $\frac{d\tilde{x}}{dt} = f(\tilde{t}+t_0, \tilde{x}+x_0)$ with $x(0) = 0$. Here t_0 and x_0 are treated as parameters and the initial condition is fixed. Conversely the problem with parameters can be replaced by the problem

$$\dot{x} = f(t,x,z), \dot{z} = 0; \quad x(t_0) = x_0, z(t_0) = z \tag{17}$$

without parameters in the equations. For this reason statements about problems containing parameters can often be reduced to statements without parameters.

Theorem 2 Let f be a continuous function on the set

$$\{(t,x,z) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^k : t_0 \leq t \leq t_0 + a, |x - x_0| \leq b, |z - z_0| \leq b'\} \qquad (18)$$

with values in $\mathbb{R}^m$ which satisfies a Lipschitz condition with respect to x. Let M be a bound for $|f|$ and $\alpha = \min\{a, b/M\}$. Then the equation (3) has a unique solution $\phi(t_0, t, x_0, z)$ on the interval $[t_0, t_0 + \alpha]$ with $\phi(t_0, t_0, x_0, z)) = x_0$. The mapping ϕ is continuous.

Proof The existence is already known. The new claim is the continuous dependence of the mapping ϕ on t_0, x_0 and z. As has already been explained it suffices to prove the case with $t_0 = 0$ und $x_0 = 0$. The proof is very similar to that in the case without parameters. We define a sequence on the product of the interval $[0, \alpha]$ with the ball of radius b' about z_0 by $x_0(t, z) = 0$ and

$$x_{n+1}(t,z) = \int_0^t f(s, x_n(s), z)ds. \qquad (19)$$

The functions x_n are continuous. They satisfy the same estimates as in the case without parameters and the sequence converges uniformly to the desired solution. This function is a uniform limit of continuous functions and therefore continuous.

If we want to consider the corresponding global problem we just need to pay attention to the dependence of the maximal interval of existence on x_0 and z. It can happen that $t_- = -\infty$ or $t_+ = +\infty$. For this reason it is convenient when describing these quantities to use the extended real numbers $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The union of the maximal intervals of existence for different values of t_0, x_0 and z is an open set. It follows in particular that the upper endpoint of the maximal interval of existence, considered as a function with values in $\bar{\mathbb{R}}$, is a lower semicontinuous function of x_0 und z. We recall the definition.

Definition A function F with values in $\bar{\mathbb{R}}$ on a topological space X is called lower semicontinuous if the condition $F(x_1) > a$ for a point $x_1 \in X$ and a constant a implies that there exists a neighbourhood U of x_1 with the property that $F(x) \geq a$ for all $x \in U$. A function F is called upper semicontinuous if $-F$ is lower semicontinuous.

The lower endpoint of the maximal interval of existence is upper semicontinuous. Intuitively this means that when initial conditions and parameters are varied the time of existence cannot suddenly shrink. The time of existence can, however, suddenly grow (t_+ need not be upper semicontinuous.) An example is $\dot{x} = x^2 - zx^3$ with $x(0) = 1$. For $z = 0$ we have $t_+ = 1$ but for $z > 0$ we have $t_+ = \infty$. In this case $\dot{x} > 0$ when $x = 0$. Thus a solution which is initially positive cannot become negative. On the other hand $\dot{x} < 0$ when $x > z$ so that the solution is bounded from above. It follows that global existence in the future holds for $z > 0$ as a consequence of the continuation criterion.

We have now proved continuous dependence for continuous parameters z. Corresponding results hold for discrete parameters, with a very similar proof.

Let $f_n(t,x)$ be a sequence of continuous functions which are defined for $t_0 \leq t \leq t_0 + a$ und $|x - x_0| \leq b$ and satisfy a Lipschitz condition with respect to x, the Lipschitz constant K being independent of n, which converges uniformly to $f(t,x)$. The function f is continuous and satisfies a Lipschitz condition with respect to x with the same Lipschitz constant K. Let Z be the subset of $\mathbb{R}$ which consists of the points 0 and $\frac{1}{n}$ for natural numbers n. Z is compact. We denote a point of Z by z. Let $\tilde{f}(t,x,0) = f(t,x)$ and $\tilde{f}(t,x,1/n) = f_n(t,x)$. This defines a function $\tilde{f}(t,x,z)$ on $[t_0, t_0+a] \times \bar{B}_b(x_0) \times Z$. It is continuous and satisfies a Lipschitz condition with respect to x with constant K. We consider an iteration as in the proof of Theorem 2.

$$\tilde{x}_{l+1}(t,z) = \int_0^t \tilde{f}(s, \tilde{x}_l(s), z)ds. \tag{20}$$

We obtain a sequence $\tilde{x}_l(t,z)$ on $[t_0, t_0+a] \times Z$ which converges uniformly to a continuous limit $\tilde{x}(t,z)$. Since its domain of definition is compact the function $\tilde{x}$ is uniformly continuous. If we define $x_n(t) = \tilde{x}(t,1/n)$ and $x(t) = \tilde{x}(t,0)$ then these functions satisfy the equations $\dot{x}_n = f_n(t,x)$ and $\dot{x} = f(t,x)$ with $x_n(t_0) = x(t_0) = x_0$. x_n converges uniformly to x because of the uniform continuity of $\tilde{x}$. This result implies a corresponding one for sequences of initial data by means of a suitable transformation of variables. Let $x_n(t)$ be the solutions of $\dot{x} = f(x)$ with $x(t_0) = x_{0,n}$ and $x_{0,n} \to x_0$ for $n \to \infty$. Then $x_n(t)$ converges to the solution of $\dot{x} = f(x)$ with $x(t_0) = x_0$.

With the statements about continuous dependence we can introduce some general concepts which are useful for the study of the global properties of solutions of a dynamical system. Consider a dynamical system which is defined on a subset G of $\mathbb{R}^m$ and a solution $x(t)$ of this system which is defined on the interval $[t_0, \infty)$. If there is a sequence $\{t_n\}$ in $[t_0, \infty)$ with $t_n \to \infty$ and $x(t_n) \to y \in \mathbb{R}^m$ für $n \to \infty$, then y is called an ω limit point of the solution x. The set of all ω limit points of the solution x is called the ω limit set of x. If a solution $x(t)$ is defined on the interval $(-\infty, t_0]$ and there is a sequence $\{t_n\}$ with $t_n \to -\infty$ and $x(t_n) \to y \in \mathbb{R}^m$ then y is called an α limit point of the solution x. The set of all α limit points of the solution x is called the α limit set of x. For every statement about ω limit points there is a corresponding statement about α limit points. To obtain this it suffices to consider the solution $\tilde{x}(t) = x(-t)$. For this reason we will only prove statements about ω limit points and leave it to the reader to derive the corresponding statements for α limit points.

The ω limit set is closed. Let $\{y_n\}$ be a sequence of ω limit points of a solution x which converges to $y \in \mathbb{R}^m$. Then there are sequences $\{t_{n,k}\}$ with $t_{n,k} \to \infty$ for $k \to \infty$ and $x(t_{n,k}) \to y_n$. When $\epsilon > 0$ there exists an N with the property $|y_n - y| < \frac{\epsilon}{2}$ for all $n \geq N$. In addition there exists $k(n) \geq n$ with the property that $|x(t_{n,k(n)}) - y_n| < \frac{\epsilon}{2}$. Hence $|x(t_{n,k(n)}) - y| < \epsilon$ for all $n \geq N$ and $t_{n,k(n)} \to \infty$ for $n \to \infty$. It follows that y belongs to the ω limit set of x and that this set is closed. If the solution is bounded then the ω limit set is compact. For when x is bounded its ω limit set is also bounded. Since it is closed it is compact. If the solution x is bounded then the ω limit set is

non-empty and connected. Let $\{t_n\}$ be an arbitrary sequence in $[t_0, \infty)$ with $t_n \to \infty$ for $n \to \infty$. Because this sequence is bounded it has a convergent subsequence which converges to some $y \in \mathbb{R}^m$. The point y is in the ω limit set of x and hence this set is non-empty. A toplogical space is called connected if the conditions $X = U \cup V$ and $U \cap V = \emptyset$ for open subsets U and V of X imply that either U or V is empty. Let X be the ω limit set of x and suppose that U and V are subsets with the properties just assumed. We suppose that neither of them is empty and obtain a contradiction. Let $y_1 \in U$ and $y_2 \in V$. There exist t_{2n} with $|x(t_{2n}) - y_1| < \frac{1}{n}$ and t_{2n+1} with $|x(t_{2n}) - y_2| < \frac{1}{n}$ and $t_{2n} \leq t_{2n+1} \leq t_{2n+2}$ for all n. For n large enough $t_{2n} \in U$ and $t_{2n+1} \in V$. Let $t'_n \leq t_{2n+1}$ be the first time after t_{2n} for which $x(t'_n)$ is not in U. A time of this kind exists. $x(t'_n)$ cannot lie in V because if it did $x(t)$ would lie in V for all t in an open neighbourhood of t'_n. In this way we obtain a sequence t'_n with the property that $x(t'_n)$ is in the complement of $U \cup V$. This sequence has a subsequence which converges to some $y \in \mathbb{R}^m$. The point y is an ω limit point of x and is in X, a contradiction.

If a dynamical system is defined on G a set A is called a forward-invariant subset if $x(t_0) \in A$ implies that $x(t) \in A$ for all $t \geq t_0$. If $x_0 \in G$ the forward integral curve through x_0 is the image of a solution on $[t_0, t_+)$ with $x(t_0) = x_0$. A subset A is forward-invariant exactly when the forward integral curve through any point of A is contained in A. Often the word 'forwards' is left out of these names, when it is implied by the context. If x is a solution and the ω limit set of x is contained in G then this ω limit set is invariant. To prove this statement let y_0 be a point in the ω limit set of a solution x. Let $\{t_n\}$ be a sequence with $x(t_n) \to y$. Let $\tilde{x}(t) = x(t - t_n)$. Then $\tilde{x}_n(0)$ converges to y_0. Let y be the solution with $y(0) = y_0$. By the continuous dependence of solutions on initial data it follows that $\tilde{x}_n(t)$ converges to $y(t)$. Since $x(t + t_n) = \tilde{x}_n(t)$ it can be concluded that $y(t)$ is in the ω limit set of x for all t for which this quantity is defined.

Next it will be shown that solutions also depend diffentiably on initial data and parameters.

Theorem 3 Suppose that the function f of Theorem 2 is continuously differentiable. Then the mapping $\phi(t_0, t, x_0, z)$ is also continuously differentiable. The derivatives satisfy the linear differential equations

$$\frac{d}{dt}\left(\frac{\partial x_i}{\partial x_{0,j}}\right) = \frac{\partial f_i}{\partial x_k}\frac{\partial x_k}{\partial x_{0,j}}, \tag{21}$$

$$\frac{d}{dt}\left(\frac{\partial x_i}{\partial z_j}\right) = \frac{\partial f_i}{\partial x_k}\frac{\partial x_k}{\partial z_j} + \frac{\partial f_i}{\partial z_j}. \tag{22}$$

with initial conditions δ^i_j and 0, respectively. In addition

$$\frac{\partial x_i}{\partial t_0} = -\frac{\partial x_i}{\partial x_{0,j}} f_j. \tag{23}$$

To prove this theorem we use the following strategy. Suppose for a moment that the differentiability holds and that derivatives with respect to time com-

mute with those with respect to the initial data and parameters. Then using the chain rule we can derive equations satisfied by the derivatives. We call these the variational equations. The solutions of the variational equations can be used to prove differentiability. We say that the equations are first differentiated formally. The results are then used to show that differentiation is really allowed. The last formula in the Theorem can be derived from the identity

$$\phi(t_0, t, \phi(t, t_0, x_0, z), z) = x_0 \tag{24}$$

by differentiating it formally with respect to t_0. In proving the theorem we use the following form of the mean value theorem. A subset G of $\mathbb{R}^m$ is called convex if $x \in G$, $y \in G$ and $0 \leq s \leq 1$ imply that $sx + (1-s)y \in G$.

Lemma 2 Let f be a continuous function on $(a, b) \times G$ with G a convex subset of $\mathbb{R}^m$ which has continuous partial derivatives with respect to the components of the second argument. Then there exist continuous functions $f_k(t, x_1, x_2)$ on $(a, b) \times G \times G$ with the properties that

$$f_k(t, x, x) = \frac{\partial f(t, x)}{\partial x_k} \tag{25}$$

und

$$f(t, x_2) - f(t, x_1) = f_k(t, x_1, x_2)(x_{2,k} - x_{1,k}). \tag{26}$$

The functions f_k can be defined by the following formula

$$f_k(t, x_1, x_2) = \int_0^1 \frac{\partial f(t, sx_2 + (1-s)x_1)}{\partial x_k} ds. \tag{27}$$

Proof Let $F(s) = f(t, sx_2 + (1-s)x_1)$ für $0 \leq s \leq 1$. Since G is convex the function F is well-defined. Is satisfies

$$\frac{dF}{ds} = \frac{\partial f}{\partial x_k}(t, sx_2 + (1-s)x_1)(x_{2,k} - x_{1,k}). \tag{28}$$

If f_k is defined as in the statement of the lemma then $F(1) - F(0)$ is equal to the right hand side of (26). Since $F(1) = f(t, x_2)$ and $F(0) = f(t, x_1)$ this completes the proof.

Proof of Theorem 3 The derivatives with respect to x_0 are considered first. For this purpose we can eliminate the parameter dependence by a change of variables. In this case we have a solution $\phi(t_0, t, x_0)$ which is given by Theorem 1. The statement is a completely local one and so we only need to prove it for a smal neighbourhood of a point. Let h be a real number and e_k the kth coordinate basis vector in $\mathbb{R}^m$. Let $x_h(t) = \phi(t_0, t, x_0 + he_k)$. It follows from Theorem 1 that x_h converges uniformly to x_0 for $h \to 0$. We have

$$\frac{d}{dt}[x_h(t) - x_0(t)] = f(t, x_h(t)) - f(t, x_0(t)). \tag{29}$$

The lemma gives

$$\frac{d}{dt}[x_h(t) - x_0(t)] = f_k(t, x_0(t), x_h(t))(x_{h,k}(t) - x_{0,k}(t)). \tag{30}$$

Let $y_h = h^{-1}[x_h(t) - x_0(t)]$ for $h \neq 0$. The existence of the derivative is equivalent to the existence of the limit of y_h as $h \to 0$. The initial condition implies that $x_h(t_0) = x_0 + he_k$ and $y_h(t_0) = e_k$. The function y_h satisfies the equation

$$\dot{y}_h = f_k(t, x_0(t), x_h(t))y_{h,k}. \tag{31}$$

The quantity $f_k(t, x_0(t), x_h(t))$ converges to $\frac{\partial f}{\partial y_k}(t, x_0(t))$ for $h \to 0$. We have a family of equations for the functions y_h which depends continuously on the parameter h. These equations have solutions with $y_h(t_0) = e_k$ which depend continuously on h, even for $h = 0$. The limit exists and is a solution of the equation (21). The partial derivative is the solution of an equation which depends continuously on parameters and is therefore itself continuous.

Consider now the derivative with respect to t_0. This time let

$$y_h(t) = \frac{\phi(t_0 + h, t, x_0) - \phi(t_0, t, x_0)}{h}, \quad h \neq 0. \tag{32}$$

We use the identity

$$\phi(t_0 + h, t, x_0) = \phi(t_0, t, \phi(t_0 + h, t_0, x_0)), \tag{33}$$

It follows that

$$hy_h(t) = \phi(t_0, t, \phi(t_0 + h, t_0, x_0)) - \phi(t_0, t, x_0). \tag{34}$$

Für $h \to 0$ we have $\phi(t_0 + h, t_0, x_0) \to x_0$ and the lemma gives

$$hy_h(t) = \left[\frac{\partial x}{\partial x_{0,k}} + o(1)\right](\phi_k(t_0 + h, t_0, x_0) - x_{0,k}). \tag{35}$$

By using the relation $\phi(t_0 + h, t_0 + h, x_0) = x_0$ and the mean value theorem we get

$$\phi_k(t_0 + h, t_0, x_0) - x_{0,k} = -\frac{\partial \phi_k}{\partial t}(t_0 + \theta h, t_0 + h, y_0) \tag{36}$$

for some θ in $(0, 1)$. The time derivative can be replaced using the differential equation. It follows that

$$y_h(t) = -\left[\frac{\partial x}{\partial x_{0,k}} + o(1)\right][f_k(t_0, x_0) + o(1)]. \tag{37}$$

We see that $\frac{\partial \phi}{\partial t_0}$ exists and satisfies the relation given in the theorem.

By using the results about the existence and continuity of the derivatives with respect to the initial data and the transformation which replaces parameters by initial data it is possible to get the statement about the differentiable dependence of the solutions on parameters. This also gives the evolution equation for the derivatives with respect to the variables z_i.

Consider now an equations of the form $\dot{x} = f(t, x, z, z^*)$ where the partial derivatives of first order with respect to x und z of the continuous function f exist and are continuous. This equation has a unique solution $x = \phi(t_0, t, x_0, z, z^*)$

with $x(t_0) = x_0$. This solution has first order partial derivatives with respect to t, t_0, x_0 and z and these derivatives are continuous as functions of (t_0, t, x_0, z, z^*). These statements can be proved just as in the proof of Theorem 3 since the variables z^* play no essential role.

With these results it is easy to prove the existence and continuity of higher derivatives of x when the existence and continuity of the corresponding derivatives of f is known. This can be proved by induction. To be concrete, consider a partial derivative of $\phi(t_0, t, x_0, z, z^*)$ with respect to the variables x_0 and z of order n. The derivative of ϕ with respect to x_0 satisfies the equation (21) and the coefficients of this equation have continuous derivatives with respect to x_0 and z of order up to $n-1$. Thus the solution also has derivatives of this kind. We can proceed similarly with the derivative of ϕ with respect to z. It can be concluded that that all derivatives of ϕ with respect to x and z up to order n exist and are continuous. Statements about derivatives with respect to t_0 can be obtained by using the results already obtained in the equation (23).

The statements about continuity and differentiability which have been proved can be used to prove something about the qualitative behaviour of solutions in the easiest case. This concerns the nature of a flow near a point where the vector field which generates it does not vanish.

Theorem 4 (Flow-box theorem) Let $\dot{x} = f(x)$ be an autonomous dynamical system where f is a continuously differentiable function on a subset G of $\mathbb{R}^m$. Let $x_0 \in G$ be a point with $f(x_0) \neq 0$. Then there exists an open neighbourhood of V of 0 in $\mathbb{R}^{m-1}$, a positive number ϵ and a diffeomorphism F from $[-\epsilon, \epsilon] \times V$ onto an open neighbourhood U of x_0 with $F(0) = x_0$ with the property that the flow ϕ of the system satisfies the relation

$$\phi(t, F(y))) = F(T_t(y)). \tag{38}$$

Here T_t is the translation by t in the direction of x_1, i.e. $(x_1, x_2, \ldots, x_m) \mapsto (x_1 + t, x_2, \ldots, x_m)$.

Proof It can be assumed w.l.o.g. that $x_0 = 0$ and $f(x_0) = e_1$. We introduce $\bar{x}$ as an abbreviation for $(x_2, \ldots, x_n)$. Let V be an open neighbourhood of 0 in $\mathbb{R}^{m-1}$ with the property that $(0, \bar{y}) \in G$ for all $\bar{y} \in V$. Let ϵ be so small that $\phi(t, (0, \bar{x})) \in G$ for all $\bar{x} \in V$ and $|t| \leq \epsilon$. Let $F(y) = \phi(y_1, (0, \bar{y}))$. The derivative of F at the origin is the identity. It follows from the inverse function theorem that there exists a neighbourhood W of the origin with the property that the restriction of F to W is a diffeomorphism onto its image. The size of V and ϵ can be reduced if necessary so that $[-\epsilon, \epsilon] \times V \subset W$. Die mapping F satisfies (38) because both sides of this equation are equal to $\phi(t + y_1, (0, \bar{y}))$.

This result says that an arbitrary vector field can be transformed by a diffeomorphism near any point where it does not vanish to the simple vector field with components $(1, 0, \ldots, 0)$. Intuitivrly, near any point where it does not vanish a vector field has no structure. If two vector fields with flows ϕ und ψ satisfy a relation of the form $F(\phi(t, x) = \psi(t, F(x))$ for a C^1 diffeomorphism F then they are called C^1 conjugate. It follows from the Flow-box Theorem that when $f(x_0)$

and $g(y_0)$ are non-vanishing the restrictions of f and g to appropriate neighbourhoods of x_0 and y_0 are C^1 conjugate. If the mapping F is only continuous f and g are called topologically conjugate. In this case the integral curves of f are mapped onto those of g and the direction of time is preserved. If these conditions are satisfied but the time coordinates on the integral curves which are related by F are not necessarily equal then f and g are called topologically equivalent.

5 Stationary solutions and their stability

We have seen that near points where a vector field does not vanish its flow has a very simple qualitative behaviour. Where the vector field has a zero things can become much more complicated. If $f(x_0) = 0$ then $x(t) = x_0$ is a time-independent solution, a stationary solution. The equation $f(x) = 0$ is difficult to solve in general. It is already difficult to say how many solutions it has. Next a result will be presented which guarantees the existence of a stationary solution under weak hypotheses. First we need a result from topology.

Theorem (Brouwer fixed point theorem) Let A be a topological space which is homeomorphic to a closed ball in $\mathbb{R}^m$ and let $\psi : A \to A$ be a continuous mapping. Then there exists a point $x \in A$ with $\psi(x) = x$.

In the next proof periodic solutions play a role. A solution $x(t)$ is called periodic if there is a number $T > 0$ with $x(T) = x(0)$. It then follows by uniqueness that $x(t + T) = x(t)$ for all t. The number T is called the period.

Theorem 5 Let a dynamical system be given on an open subset of $\mathbb{R}^m$ and let A be an invariant subset which is homeomorphic to a closed ball in $\mathbb{R}^m$. Then there is at least one stationary solution in A.

Proof Suppose that there were no stationary solutions in A. For a positiv number T and $x \in A$ let $\psi_T(x) = \phi(T, x)$. Since A is compact, ψ_T is well defined. This mapping is continuous and maps A into itself. As a consequence of the Brouwer fixed point theorem there exists a point z_T with $\psi_T(z_T) = z_T$. The solution with initial value z_T is periodic with period T. By choosing different values of T we can get periodic solutions with periods $1/n$ passing through points $z_{1/n}$ for all natural numbers n. Because A is compact this sequence has a convergent subsequence. Call it y_n and its limit y. Thus there are periodic solutions which start arbitrarily close to y and take an arbitrarily short time to return to their starting points. By assumption they are not stationary. Let K be a flow box for y. The function f is bounded on A by a constant M. A periodic solution with period T can never reach a point further from its starting point than MT. For ϵ small enough the open ball of radius 2ϵ about y lies in K. If a solution starts in the open ball of radius ϵ about y and its peiod is not greater than ϵ/M then it can never leave the flow box. But there are no periodic solutions in the flow box, a contradiction.

It is possible to apply this theorem to the fundamental system of virus dynamics with A being the invariant region which gor this system which we already found. It follows that the system has at least one non-negative stationary solution for any choice of the parameters. For this system it is the positive solutions which are most interesting. Other non-negative solutions are nevertheless of some interest. They can be used in some cases to get some information about the asymptotic behaviour of positive solutions. They can also be interesting limiting cases of the original system. For instance, a solution of the fundamental system of virus dynamics with $y = 0$ and $v = 0$ corresponds to the state of a healthy person (no free virus particles, no infected cells).

The fundamental system of virus dynamics is simple enough that it is possible to compute the stationary solutions explicitly. The third equation gives $y = \frac{u}{k}v$. When we substitute this relation into the second equation both sides contain a factor v. For a positive solution we can cancel this factor with the result that $x = \frac{au}{\beta k}$. Putting this relation into the first equation gives $v = \frac{d}{\beta}\left(\frac{\beta k\lambda}{adu} - 1\right)$. Let $R_0 = \frac{\beta k\lambda}{adu}$. This object is known to the biologists as the fundamental reproductive ratio. We see that a positive stationary solution can only exist when $R_0 > 1$. Substituting the equation for x into the second equation for stationary solutions gives $y = \frac{u}{k}v$ and $y = \frac{du}{\beta k}(R_0 - 1)$. Now two things have been proved. If $R_0 > 1$ there is exactly one positive solution which we have calculated explicitly. When $R_0 \leq 1$ no positive stationary solution exists. If solutions are allowed which are merely non-negative then $v = 0$ is also a possibility. Then $y = 0$ also holds. In the case the remaining equation says that $x = \frac{\lambda}{d}$.

The significance of stationary solutions depends on their stability. A stationary solution x^* is called stable if for any open neighbourhood U of x^* there exists a neighbourhood V of x^* such that any solution satisfying $x(t_0) \in V$ the condition $x(t) \in U$ for all $t \geq t_0$ follows. In words, each solution which starts in V stays in U as long as it exists. The stationary solution x^* is called asymptotically stable if it is stable and there exists a neighbourhood U of x^* with the property that $x(t_0) \in U$ implies $x(t) \to x^*$ for $t \to \infty$. The first condition in this definition does not follow from the second as is shown by the following example.

$$\dot{x} = x - rx - ry + xy, \tag{39}$$

$$\dot{y} = y - ry + rx - x^2. \tag{40}$$

This system is C^1. The qualitative behaviour is easier to see in polar coordinates where the system takes the form

$$\dot{r} = r(1 - r), \tag{41}$$

$$\dot{\theta} = r(1 - \cos\theta). \tag{42}$$

There are stationary points at $(0, 0)$ and $(1, 0)$. All solutions except the stationary solution at the origin converge to $(0, 1)$ for $t \to \infty$ but this point is not

stable. These claims are not proved here since the techniques which would be needed to do so have not yet been introduced.

One way of investigating the stability of a stationary solution x^* is to linearize the system about x^*. Consider the Taylor expansion of f about x^*, $f_i(x) = \frac{\partial f_i}{\partial x_j}(x^*)(x_j - x_j^*) + o(|x - x^*|)$. The matrix $\frac{\partial f_i}{\partial x_j}(x^*)$ will be denoted by A. The linearized equation about x^* is obtained by omitting the remainder term in the Taylor expansion which is supposed to be small. This leads to the equation $\frac{d\hat{x}}{dt} = A\hat{x}$. The hope is that under suitable conditions the solutions of the linearized equation approximate solutions of the original equation near x^*. When we consider the qualitative behaviour of solutions close to some point x^* we do not distinguish between systems which are topologically conjugate by a mapping which leaves x^* fixed. For this reason we are only interested in properties of A which are invariant under similarity transformations. Since any matrix is similar to its Jordan normal form these can only be properties of the normal form. For a detailed discussion of linear differential equations the reader is referred to the first chapter of the book of Perko [9]. The reduction to normal form is in general only possible with the help of complex numbers, because the eigenvalues can be complex. For this reason we are sometimes forced here to consider complex linear ordinary differential equations although in the end we are only interested in real solutions of equations with real coefficients.

For a complex matrix A with an eigenvalue λ the vectors which satisfy $(A - \lambda I)^k x = 0$ for a natural number k are called the corresponding generalized eigenvectors. They form a vector space V_λ. When a matrix is in canonical form the non-vanishing elements belong to a sequence of blocks along the diagonal, the Jordan blocks. The diagonal elements in each block are equal to a number λ which is an eigenvalue of the matrix. The elements immediately above the diagonal are equal to one and all remaining elements are zero. Suppose that the sizes of the blocks are n_i. The vectors where only the first n_1 components are different from zero are generalized eigenvectors which belong to the first eigenvalue λ_1. The vectors where only the components from $n_1 + 1$ to $n_1 + n_2$ are different from zero belong to the second eigenvalue λ_2 and so on. If A is a general real matrix and λ is a real eigenvalue then the space V_λ is defined just as in the complex case. When λ is a complex eigenvalue then the definition is a little more complicated. In that case V_λ is the set of real parts of the complex solutions of $(A - \lambda I)^k x = 0$. Because A is real $\bar{\lambda}$ is also an eigenvalue and $V_{\bar{\lambda}} = V_\lambda$. The whole space is a direct sum of generalized eigenvectors.

When we want to solve the equation $\dot{x} = Ax$ we can put A into Jordan form, solve the equation and transform the solution back. The subspaces of generalized eigenvectors which belong to the eigevectors are invariant under the flow of the linearized system. Thus it is enough to consider the case where there is only one Jordan block. If this block is of size one, with eigenvalue λ, then the solution is of the form $ce^{\lambda t}$ for a constant c. In general the solution is the product of the function $e^{\lambda t}$ with a Matrix whose elements are each a constant times a power of t. In general these are complex exponential functions. Taking real and imaginary parts in order to get solutions of the real equation then

each element is a linear combination of expressions of the form t^k, $t^k e^{at}\cos bt$ or $t^k e^{at}\sin bt$ where $\lambda = a + bi$. When λ is real and positive then the solution, if it not identically zero, grows at least as fast as $e^{\lambda t}$ when t increases. When λ is compolex and $a > 0$ then the solution grows at least as fast as e^{at} along suitable sequences which tend to $+\infty$. When λ is real and negative then the solution decays at least as fast as $e^{(\lambda+\epsilon)t}$ when t increases where $\epsilon > 0$ is arbitrary. When λ is complex and $a < 0$ then it decays at least as fast as $e^{(a+\epsilon)t}$. Of course similar statements can be made for the other time direction.

For the study of the properties of linear ordinary differential equations the concept of the exponential of a matrix is very useful. It is defined by

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}. \tag{43}$$

For an arbitrary complex matrix A this series converges uniformly on each compact subset in the sense that all the entries of the corresponding matrix do so. The relevance of this definition to ordinary differential equations is that $e^{tA}x_0$ solves the equation $\dot{x} = Ax$ with $x(0) = x_0$. When the matrix A is in Jordan form the matrix e^{tA} is the direct sum of expressions for the individual Jordan blocks. In this way it is possible to obtain estimates for e^{tA} which correspond to the estimates for linear equations which were discussed above.

We see that for linear systems eigenvalues with positive real parts have to do with instability and eigenvalues with negative real parts have to do with stability. This observation motivates the following definitions. The space V_+ which is spanned by all generalized eigenvectors which belong to eigenvalues with positive real part is called the unstable subspace. The space V_- which is spanned by all generalized eigenvectors which belong to eigenvalues with negative real part is called the stable subspace. The space which is spanned by all generalized eigenvectors which belong to eigenvalues zero positive real part is called the centre subspace. The whole space is the direct sum $V_- \oplus V_c \oplus V_+$. These spaces are invariant under the flow. For a linear system the following statements hold. If all eigenvalues have negative real part then the origin is asymptotically stable. If at least one eigenvalue has positive real part then the origin is unstable. In what follows we will prove analogou statements for a stationary solution of a general nonlinear system. For that we need introduce some more ideas.

Before doing that we look at the linearization of the fundamental model of virus dynamics about the two stationary solutions. The linearization about an arbitrary point is

$$\frac{d\hat{x}}{dt} = (-d - \beta v)\hat{x} - \beta x\hat{v}, \tag{44}$$

$$\frac{d\hat{y}}{dt} = \beta v\hat{x} - a\hat{y} + \beta x\hat{v}, \tag{45}$$

$$\frac{d\hat{v}}{dt} = k\hat{y} - u\hat{v}. \tag{46}$$

In the case of the stationary point with $v = 0$ this expression simplifies considerably and it can be seen immediatey that $-d$ is an eigenvalue. The other two can then be determined by solving a quadratic equation. The result is

$$\mu = \frac{1}{2}[-(a+u) \pm \sqrt{(a+u)^2 - 4au(1-R_0)}]. \tag{47}$$

We see that there are always at least two negative eigenvalues and that the third is positive, zero or negative according to whether $R_0 > 1$, $R_0 = 1$ or $R_0 < 1$. According to the stability criteria which we have not yet proved this point is asymptotically stable for $R_0 < 1$ and unstable for $R_0 > 1$. Intuitively these statements have the following meaning. This solution represents the state of a healthy person. When an infection takes place this state is disturbed a little. For $R_0 < 1$ the infection is automatically eliminated. For $R_0 > 1$ the virus is able to establish itself in the body. The linearization about the other stationary point is

$$\frac{d\hat{x}}{dt} = -dR_0\hat{x} - \frac{au}{k}\hat{v}, \tag{48}$$

$$\frac{d\hat{y}}{dt} = d(R_0 - 1)\hat{x} - a\hat{y} + \frac{au}{k}\hat{v}, \tag{49}$$

$$\frac{d\hat{v}}{dt} = k\hat{y} - u\hat{v}. \tag{50}$$

This leads to the eigenvalue equation

$$\mu^3 + (a + u + dR_0)\mu^2 + dR_0(a+u)\mu + adu(R_0 - 1) = 0. \tag{51}$$

Only the case $R_0 > 1$ is of interest since only then is the stationary solution positive. In that case all coefficients in the polynomial are positive. To get information about the eigenvalues we use the Routh-Hurwitz criterion. For a third degree equation of the form

$$\mu^3 + a_1\mu^2 + a_2\mu + a_3 = 0. \tag{52}$$

this criterion says that all eigenvalues have negative real part precisely when $a_1 > 0$, $a_3 > 0$ and $a_1a_2 - a_3 > 0$. Thus in our case all eigenvalues have negative real parts if

$$(a + u + dR_0)dR_0(a+u) > adu(R_0 - 1). \tag{53}$$

If we multiply out this equation and sort the terms according to powers of t then we get

$$d^2(a+u)R_0^2 + R_0[d(a^2 + ad + u^2)] + adu > 0, \tag{54}$$

a condition which obviously holds. When $R_0 = 1$ the eigenvalues are $0, -(a+u)$ and $-d$.

To prove statements about stability we use ideas which go back to Lyapunov. Let $\dot{x} = f(x)$ be an autonomic dynamical system. Let V be a continuously

differentiable function let $\dot{V} = \frac{\partial V}{\partial x_i} f_i(x)$. By the chain rule $\dot{V} = \frac{d}{dt}(V(x(t))$. A function which satisfies $\dot{V} \leq 0$ is called a Lyapunov function.

Theorem 6 Let G be an open neighbourhood of a point x_0. Let f be a C^1 vector field with $f(x_0) = 0$. Let V be a C^1 function with $V(x_0) = 0$ and $V(x) > 0$ for $x \neq x_0$. If $\dot{V}(x) \leq 0$ for all $x \in G$ then x_0 is stable. When $\dot{V}(x) < 0$ for all $x \in G$ except x_0 then x_0 is asymptotically stable. When $\dot{V}(x) > 0$ for all $x \in G$ except x_0 then x_0 is unstable.

Beweis We can assume w.l.o.g. that $x_0 = 0$. Let $\epsilon > 0$ be small enough that $\bar{B}_\epsilon(0) \subset G$ and let m_ϵ be the minimum of the continuous function V on the sphere S_ϵ of radius ϵ about the origin. Then $m_\epsilon > 0$. Since V is continuous and $V(0) = 0$ there exists $\delta > 0$ with the property that $V(x) < m_\epsilon$ for $|x| < \delta$. Because $\dot{V} \leq 0$ the function V cannot increase along the integral curves of the vector field. Hence the flow ϕ of f satisfies the condition

$$V(\phi(t, x_0)) \leq V(x_0) < m_\epsilon \tag{55}$$

for all $x_0 \in B_\delta(0)$. Suppose that for $|x_0| < \delta$ there exists t_1 with $\phi(t_1, x_0) \in S_\epsilon$. In this case we would have $V(\phi(t_1, x_0)) \geq m_\epsilon$, a contradiction. For this reason $|x_0| < \delta$ implies that $|\phi(t, x_0)| < \epsilon$ for $t \geq 0$.

We next consider the case that $\dot{V}(x) < 0$ for all $x \in G$ except x_0. Then V is strictly decreasing along the integral curves of f. Let $x_0 \in B_\delta(0)$ where δ is as before. Then $\phi(t, x_0) \in B_\epsilon(0)$ for all $t \geq 0$. Let $\{t_k\}$ be a sequence with $t_k \to \infty$. Since $\bar{B}_\epsilon(0)$ is compact there exists a subsequence with the property that $\phi(t_k, x_0)$ converges to a point y_0 of $\bar{B}_\epsilon(0)$. We will show that for each sequence of this type the limit must be zero. It then follows that $\phi(t_k, x_0)$ converges to the origin along each subsequence. It follows that $\phi(t, x)$ converges to the origin. It remains to show that when $\phi(t, x_0) \to y_0$ it must be the case that $y_0 = 0$. V decreases strictly along an integral curve and satisfies $V(\phi(t, x_0) \to V(y_0)$. Hence $V(\phi(t, x_0) > V(y_0)$ for all $t > 0$. If $y_0 \neq 0$ then $V(\phi(s, y_0)) < V(y_0)$ for $s > 0$. It can then be conluded by continuity that $V(\phi(s, y)) < V(y_0)$ for y near enough to y_0. But then $V(\phi(s + t_n, x_0)) < V(y_0)$ for n large enough, a contradiction.

Consider finally the case that $\dot{V}(x) > 0$ for all $x \in G$ except x_0. Let M be the maximum of V on the set $\bar{B}_\epsilon(0)$. In this case V is strictly increasing along the integral curves. For abitrary $\delta > 0$ and $x_0 \neq 0$ in $B_\delta(0)$ the inequality $V(\phi(t, x_0)) > V(x_0) > 0$ holds for all $t > 0$. The set where $V(x) \geq V(x_0)$ is open and the subset of $\bar{B}_\epsilon(0)$ on which $V(x) \geq V(x_0)$ is compact. There $\dot{V}$ is positive and it has a positive minimum m. We have $\inf_{t \geq 0} \dot{V}(\phi(t, x_0)) \geq m > 0$. Hence $V(\phi(t, x_0) \geq V(x_0) + mt > M$ for t sufficiently large. Thus the instability has been proved.

The statements about stability or instability of stationary solutions will be proved with the help of Lyapunov functions which are constructed for this purpose. Here we follow the treatment of this subject in [3].

Lemma 3 Let A be a real $n \times n$ matrix. The matrix equation $A^T B + BA = -C$ has a solution for each positive definite matrix C if and only if all eigenvalues of A have negaeive real part.
Proof Consider the linear equation $\dot{x} = Ax$ and the real-valued function $V(x) = x^T Bx$ where B is a symmetric matrix. Then

$$\dot{V}(x) = x^T (A^T B + BA)x. \tag{56}$$

If the equation considered in this lemma holds then $\dot{V}(x) < 0$ for $x \neq 0$ and the solution of the differential equation converges to the origin for $t \to \infty$. It follows that the eigenvalues of A have negative real parts. If conversely these eigenvalues have negative real parts and C is a positive definite matrix then let

$$B = \int_0^\infty e^{A^T t} C e^{At} dt. \tag{57}$$

This integral is well-defined since there exist positive constants K and α with $\|e^{At}\| \leq Ke^{-\alpha t}$ for $t \geq 0$. In addition B is positive definite and

$$A^T B + BA = \int_0^\infty \frac{d}{dt}(e^{A^T t} C e^{At}) dt = -C. \tag{58}$$

It follows that when the solution is asymptotically stable for the linear system there exists a quadratic form which is strictly decreasing along the solutions of this equation. Next we consider the ninlinear equation $\dot{x} = Ax + g(x)$ where g is continuously differentiable and satisfies the conditions $g(0) = 0$ and $\frac{\partial g_i}{\partial x_j}(0) = 0$. If the real parts of the eigenvalues of A are negative we can consider the matrix B of the Lemma in the case $C = I$. Then

$$\dot{V} = -|x|^2 + g^T Bx + x^T Bg = -|x|^2(1 + o(1)). \tag{59}$$

It follows that the origin is asymptotically stable. Now a theorem about the instability for the nonlinear system will be proved. For this it is useful to remark that the third part of Theorem 6 can be generalized. It is assumed that the stationary point which is to be examined is the origin. We introduce an open subset U with the property that the origin is in the closure of U. Let $H = U \cap B_\epsilon(0)$. We suppose that the continuously differentiable function V has the following properties. $V(x) = 0$ on the part of the boundary of H which lies in $B_\epsilon(0)$ and $V(x) > 0$ at all other points of H. $\dot{V}(x) > 0$ on H except at the origin. The claim is now that under these circumstances the origin is unstable. As in the proof of Theorem 6 let $\delta > 0$ be arbitrary. Let $x_0 \neq 0$ be an arbitrary point of $B_\delta(0) \cap H$. Then $V(\phi(t, x_0)) > V(x_0) > 0$ for all $t > 0$. Hence the solution can only leave the set H through the boundary of $B_\delta(0)$. The subset of $\bar{H}$ where $V(x) \geq V(x_0)$ is compact and $\dot{V}$ has a minimum there. Thus it is possible to argue as in the proof of Theorem 6 that the solution reaches the boundary of $B_\delta(0)$ after a finite time and the instability is proved.

How this result will be applied to the case where a stationary solution has an eigenvalue with positive real part. We assume that the stationary solution is

at the origin and that the subspaces V_+, V_c and V_- are coordinate subspaces. A general point can be represented as (x, y, z) where x, y and z belong to the subspaces V_+, V_c and V_-, respectively. In the given situation the dimension of V_+ is positive. The equations are

$$\frac{dx}{dt} = A_+x + f(x, y, z), \tag{60}$$

$$\frac{dy}{dt} = A_cy + g(x, y, z), \tag{61}$$

$$\frac{dz}{dt} = A_-z + h(x, y, z) \tag{62}$$

where the eigenvalues of A_+ and $-A_-$ have positive real part and the functions f, g and h are all $o(\sqrt{|x|^2 + |y|^2 + |z|^2})$ in a neighbourhood of the origin. As a consequence of the lemma there exist positive definite matrices B_+ und B_- with the property that $A_+^TB_+ + B_+A_+ = -I$ and $A_-^TB_- + B_-A_- = I$. As the function V we take $x^TB_+x - y^Ty - z^TB_-z$. Then

$$\dot{V} = -x^Tx - z^Tz + o(|x|^2 + |y|^2 + |z|^2). \tag{63}$$

The set U is defined by the condition $V > 0$. On this region $\dot{V} = -x^Tx - z^Tz + o(|x|^2 + |z|^2)$. Hence for ϵ small enough all conditions are satisfied and the origin is unstable.

It follows from the results which have just been proved that for $R_0 > 1$ the stationary solution of the fundamental model of virus dynamics with $v > 0$ is asymptotically stable und the solution with $v = 0$ unstable. On the other hand for $R_0 < 1$ the solution with $v = 0$, which in this case is the only non-negative stationary solution, is stable. These statements concern the behaviour of solutions in a neighbourhood of the stationary solutions. A Lyapunov function can also help to prove global results. If a dynamical system is defined on a open set U und V satisfies the inequality $\dot{V} \leq 0$ then the ω-limit points of a solution in U which lie in U are points where $\dot{V} = 0$. The function $V(x(t))$ is monotone decreasing and non-negative. Thus it converges to a constant V_∞. It follows that $V(y) = V_\infty$ for any ω-limit point y of $x(t)$. Hence V is constant on the ω-limit set. If y is an ω-limit point then the solution with initial value y is contained in the ω-limit set. The function V is constant along this solution and thus $\dot{V}(y) = 0$. It follows from these considerations that when $\dot{V} < 0$ on U there are no ω-limit points in U.

Korobeinikov [5] used Lyapunov functions to determine the global qualitative behaviour of solutions of the fundamental system of virus dynamics. He denotes the stationary solution with $v > 0$ by (x^*, y^*, v^*) and the stationary solution with $v = 0$ by $(x_0, 0, 0)$. Consider first the function

$$V(x, y, v) = x^* \left(\frac{x}{x^*} - \log \frac{x}{x^*}\right) + y^* \left(\frac{y}{y^*} - \log \frac{y}{y^*}\right) + \frac{a}{k}v^* \left(\frac{v}{v^*} - \log \frac{v}{v^*}\right) \tag{64}$$

This funciton has a minimum at the point (x^*, y^*, v^*). The derivative is

$$\dot{V} = \left(1 - \frac{x^*}{x}\right)\dot{x} + \left(1 - \frac{y^*}{y}\right)\dot{y} + \frac{a}{k}\left(1 - \frac{v^*}{v}\right)\dot{v} \tag{65}$$

$$
\begin{aligned}
&= \lambda - dx - \frac{au}{k}v - \lambda\frac{x^*}{x} + \beta x^* v + dx^* \\
&-\beta x v\frac{y^*}{y} + ay^* - ay\frac{v^*}{v} + \frac{au}{k}v^* \\
&= \lambda + dx^* + ay^* + \frac{au}{k}v^* - dx + \left(\beta x^* - \frac{au}{k}\right)v \\
&-\lambda\frac{x^*}{x} - \beta x v\frac{y^*}{y} - ay\frac{v^*}{v} \\
&= dx^*\left(2 - \frac{x}{x^*} - \frac{x^*}{x}\right) + ay^*\left(3 - \frac{x^*}{x} - \frac{xvy^*}{x^*v^*y} - \frac{yv^*}{y^*v}\right). \qquad (66)
\end{aligned}
$$

It can be shown that $\dot{V} \leq 0$ by using the inequality between arithmetic and geometric means. This says that, for positive numbers $a_1, \ldots, a_n$, $(\prod_{i=1}^n a_i)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n a_i$ and that equality holds only when all a_i are equal. It follows that $\dot{V}$ can only be zero when $x = x^*$. If there is a positive ω-limit point then $\dot{V} = 0$ there. Hence $x = x^*$ on the whole solution which starts at that point and $\dot{x} = 0$. If then information is substituted into the equation for x it is seen that v is constant. The equation for v then implies that y is constant. We see that every positive ω-limit point is a stationary solution. It can only be the point (x^*, y^*, v^*). On the other hand ω-limit points where one of the variables vanishes are impossible since because V tends to infinity in the approach to a point of that kind. Hence there can be no ω-limit point other than (x^*, y^*, v^*). Thus it has been shown that for $R_0 > 1$ every positive solution converges to this stationary solution for $t \to \infty$.

In order to understand the case $R_0 \leq 1$ we consider the function

$$
U(x, y, v) = x_0\left(\frac{x}{x_0} - \log\frac{x}{x_0}\right) + y + \frac{a}{k}v. \qquad (67)
$$

In the region where all variables are non-negative and x positive this function has a minimum at the point $(x_0, 0, 0)$. The derivative is

$$
\begin{aligned}
\dot{U} &= \left(1 - \frac{x_0}{x}\right)\dot{x} + \dot{y} + \frac{a}{k}\dot{v} \\
&= \lambda\left(2 - \frac{x}{x_0} - \frac{x_0}{x}\right) + \frac{au}{k}(R_0 - 1)v. \qquad (68)
\end{aligned}
$$

This quantity is non-negative and vanishes only when $R_0 = 1$ and $x = x_0$. An ω-limit point of a positive solution cannot satisfy $x = 0$ since U tends to infinity for $x \to 0$. It can be argued as in the case $R_0 > 1$ that an ω-limit point with $x > 0$ must be a stationary solution. Hence every solution converges to the unique stationary solution as $t \to \infty$.

This example shows how a Lyapunov function can help to investigate the asymptotic behaviour of a dynamical system. Unfortunately there is no general method for finding Lyapunov functions. It is rather an art than a science. How did Korobeinikov find his Lyapunov function? In his paper he does not say much about this but he mentions a relation to models from epidemiology. We will

follow this track a bit. This is also an opportunity to make the acquaintance of some important epidemiological models. The models concerned were introduced by Kermack and McKendrick in 1927. Consider a population of humans (or animals) which are exposed to an infectious disease. Let S be the proportion of the population which is are susceptible to the disease, I the proportion which is infected (or infectious) and R the proportion which has recovered or been removed. Then $S + I + R = 1$. Suppose that the total population is constant so that S, I and R are proportional to the numbers in the different groups. In the simplest model the equations are

$$\dot{S} = -\beta SI \tag{69}$$

$$\dot{I} = \beta SI - \alpha I \tag{70}$$

$$\dot{R} = \alpha I. \tag{71}$$

This is known as the SIR model. It is immediate from these equations that $S + I + R =$ is constant. Since R can be computed from the other variables the equation for R can be omitted. In this model it is assumed that a person who is infected can immediately infect others, which is unrealistic for many diseases. Later we will get to know another alternative. The transition from I to S can take place by recovery forom the disease with resulting immunity, by spatial separation (quarantine during an epidemic) or by death. In this model births are not included and deaths which are not due to the disease also not. The idea is that the model should only be valid for time periods where these effects play no role. Immunity after an illness occurs for the infection with many viruses, not however in the case of HIV.

In order to understand solutions of the SIR model, which is now two-dimensional, we can procede as follows. On an interval where S is monotone we can consider I as a function of S and derive the equation

$$\frac{dI}{dS} = -1 + \frac{\alpha}{\beta S} \tag{72}$$

In fact S is always strictly decreasing, since $\dot{S} < 0$. An integration shows that $I + S - \frac{\alpha}{\beta} \log S$ is constant along the integral curves. Thus it can be seen that each solution converges to a point with $I = 0$ as $t \to \infty$. This conserved quantity is useful for the investigation of the SIR model. Here we want to draw attention to the formal similarity with the Lyapunov functions of Korobeinikov. In the case of the SIR model the time derivative of the function is zero instead of negative.

Diseases without immunity, which include many of those caused by bacteria or helminths can be described by the SIS model where an individual who recovers instead of coming into the group R returns to the group S. The equations are

$$\dot{S} = -\beta SI + \gamma I \tag{73}$$

$$\dot{I} = \beta SI - \gamma I. \tag{74}$$

The quantity $S + I$ is constant and can, if we work with proportions of the population, be set to one. The variable S can be eliminated with the result

$$\dot{I} = \beta I(1 - I) - \gamma I = (\beta - \gamma)I \left(1 - \frac{I}{1 - \frac{\gamma}{\beta}}\right). \tag{75}$$

If $\gamma > \beta$ then $\dot{I} < 0$ and the solutions converge to zero as $t \to \infty$. If $\gamma > \beta$ then the solutions converge to $1 - \frac{\gamma}{\beta}$. In this model $R_0 = \frac{\beta}{\gamma}$ plays the role of the fundamental reproductive ratio.

The SIR model can be modified by introducing a new group of people who are infected but not yet infectious, the exposed group E. The people in the group I are infectious. In addition demographic effects (birth and death) are modelled. The new model (SEIR model with birth and death) is

$$\dot{S} = \mu - \beta SI - \mu S \tag{76}$$

$$\dot{E} = \beta SI - (\theta + \mu)E \tag{77}$$

$$\dot{I} = \theta E - (\delta + \mu)I. \tag{78}$$

The quantity R has once again been omitted since the equation for it decouples. This time we have a three-dimensional system. It was studied in [6]. The system in the paper was a little more complicated because it also included vertical transmission, i.e. transmission from the mother to the unborn child. Here we only include normal horizontal transmission. The observation of Korobeinikov is that this system is up to notation identical with the fundamental system of virus dynamics. We only need to identify the group S with the non-infected cells, the group E with the infected cells and the group I with the virus particles. There exists a SEIS model whose dynamics was studied by Korobeinikov with the help of a Lyapunov function. In general the question of whether a disease can maintain itself in a population is determined by a parameter R_0. Vaccination of children can be used to lower the effective value of R_0 and thus to combat the progress of the disease. If this leads to $R_0 < 1$ we talk of herd immunity. This is not easy to achieve. For measles it has been estimated that in developed countries it requires between 85 and 90 per cent vaccination in rural populations and well over 90 per cent in urban populations. In developing countries things are quite different since measles is often fatal.

The equations which have been considered here also have a similarity to the famous Lotka-Volterra equations for predator-prey systems. In that case the equations are

$$\dot{x} = x(\lambda - by) \tag{79}$$

$$\dot{y} = y(-\mu + cx) \tag{80}$$

and there exists the conserved quantity

$$cx - \mu \log x + by - \lambda \log y. \tag{81}$$

Here is the interpretation that x is the population of prey (e.g. hares) and y the population of predators (e.g. lynx).

6 The Arzela-Ascoli theorem

In the next section the existence of invariant manifolds is proved. For this we need the Arzela-Ascoli theorem. Since we do not wish to assume familiarity with this theorem it will be proved here. This theorem is also valuable in many other contexts in the theory of dynamical systems. It will first be proved in a relatively general setting.

Definition Let X be a metric space with metric ρ and $\mathcal{F}$ a set of real-valued functions on X. $\mathcal{F}$ is called *equicontinuous* if for each $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all $f \in \mathcal{F}$ and all x and y with $\rho(x, y) < \delta$. (In particular each function in $\mathcal{F}$ is uniformly continuous.)

$\mathcal{F}$ is called pointwise bounded if for each $x \in X$ there is an $M(x)$ such that $|f(x)| \leq M(x)$ for all $f \in \mathcal{F}$.

Theorem (Arzela-Ascoli) Let $\mathcal{F}$ be a pointwise bounded equicontinuous set of real-valued functions on a metric space X and suppose that there exists a countable dense subset $E \subset X$. Then each sequence $\{f_n\}$ of functions in $\mathcal{F}$ has a subsequence which converges uniformly on each compact subset of X.

Proof Let $x_1, x_2, x_3, \ldots$ be an enumeration of the points of E. Let S_0 be the set of natural numbers. Let $k \geq 1$ and suppose that an infinite subset S_{k-1} of S_0 has been chosen. Since $\{f_n(x_k) : n \in S_{k-1}\}$ is a bounded set of real numbers it has a convergent subsequence. In other words there exists an infinite set $S_k \subset S_{k-1}$ with the property that $\lim_{n\to\infty} f_n(x_k)$ exists for $n \in S_k$. In this way we obtain infinite sets $S_0 \supset S_1 \supset S_2 \supset \ldots$ with the property that the limit of $f_n(x_k)$, for $1 \leq j \leq k$, when $n \to \infty$ within S_k exists. Let r_k be the kth element of S_k and let S be the sequence $r_1, r_2, r_3, \ldots$. For each value of k there are at most $k-1$ elements of S which do not lie in S_k. Hence the limit $\lim f_n(x)$ exists for all $x \in E$ when $n \to \infty$ within S.

Let $K \subset X$ be compact and $\epsilon > 0$. Because $\mathcal{F}$ is equicontinuous there exists a $\delta > 0$ so dass $\rho(p, q) < \delta$ implies that $|f_n(p) - f_n(q)| < \epsilon$ for all n. There is a covering of K by finitely many open balls $B_1, B_2, \ldots, B_M$ of radius $\frac{\delta}{2}$. Since E is a dense subset of X there is a point p_i of E in each B_i, $1 \leq i \leq M$. The limit of $f_n(p_i)$ within S exists. Thus there exists an N with the property that $|f_m(p_i) - f_n(p_i)| < \epsilon$ for $1 \leq i \leq M$ when $m > N$, $n > N$ and m and n are in S. Now let $x \in K$. Then there is an i with $x \in B_i$ and $\rho(x, p_i) < \delta$. Because of the choice of N and δ we have

$$\begin{aligned} |f_m(x) - f_n(x)| &\leq |f_m(x) - f_m(p_i)| + |f_m(p_i) - f_n(p_i)| + |f_n(p_i) - f_n(x)| \\ &< \epsilon + \epsilon + \epsilon = 3\epsilon \end{aligned} \tag{82}$$

when $m > N$, $n > N$, $m \in S$ and $n \in S$.

The analogous statement for functions with values in $\mathbb{R}^k$ can be proved by the same method. The case which will interest us most in what follows is that where X is $\mathbb{R}^m$ with the Euclidean metric. In this case the subset E can be defined to be that of points with rational coordinates.

7 Invariant Manifolds

In the last section we proved results about the behaviour of the solutions of a dynamical system near a stationary solution under certain conditions. When all eigenvalues of the linearized system have negative real parts then soluitons which start near the stationary solution converge to it for $t \to \infty$. By replacing t by $-t$ we get analogous statements in the case that all eigenvalues have positive real parts. If the real parts of the eigenvalues have both signs then we still have little information. In this section we want to learn more about that case by studying invariant manifolds. An invariant manifold is a submanifold which is invariant under the flow. In this context it is useful to consider the restriction of the flow $\phi(t, x)$ to a fixed time t. This gives a local diffeomorphism between subsets of $\mathbb{R}^m$. An invariant manifold of a mapping of this type is a submanifold which is mapped into itself by the diffeomorphism.

For a real number t let T^t be a continuous mapping of a neighbourhood G_t of the origin in $\mathbb{R}^m$ onto a neighbourhood of the origin in the same space with $T^t(0) = 0$. A set S is called invariant with respect to $\{T^t\}$ if $T^t(G_t \cap S) \subset S$ for all t. It is called locally invariant if there exists $\epsilon > 0$ with the property that if $x \in S$ and $|T^t(0)| < \epsilon$ for $0 \leq t_0$ then $|T^t(x)| \in S$. We consider now a linear system $\dot{x} = Ax$ and the perturbed system $\dot{x} = Ax + F(x)$ where F is a continuously differentiable mapping which is $o(|x|)$ as $x \to 0$. The second property is equivalent to the conditions that $F(0) = 0$ and $\frac{\partial F_i}{\partial x_j}(0) = 0$. The first system is the linearization of the second system at the origin. We define $T^t(x) = \phi(t, x)$ on the region where this expression exists. Here ϕ is the flow of the nonlinear system.

If S is invariant then the intersection of S with a ball is locally invariant. If conversely S is locally invariant then the union of the sets $T^t(S \cap G_t)$ is invariant. Thus there is a close connection between invariant and locally invariant sets. This is useful for the following reason. If F is changed outside of a small ball and an invariant manifold for the new equation can be found then a locally invariant set for the original equation can be obtained. Another advantage of locally invariant sets is that it is to be expected that they are simpler that globally invariant sets. For instance there can exist solutions which tend to the origin for $t \to \infty$ and $t \to -\infty$ and form loops in between.

Suppose that the Matrix of first derivatives can be written in the form $A = [P, Q]$ where the eigenvalues p_j of P satisfy the inequality Re $p_j \leq \alpha < 0$ and the eigenvalues q_k of Q the inequality Re $q_k \geq \beta > \alpha$. The subspace where the coordinates z_j vanish is the union of all solutions whose distance to the origin is bounded by $e^{(\beta-\epsilon)t}$ for an $\epsilon > 0$. The question arises whether a similar statement holds for a nonlinear system. Consider the system

$$\dot{y} = Py + F_1(y, z), \quad \dot{z} = Qy + F_2(y, z) \tag{83}$$

where $F = (F_1, F_2)$ has the same properties as before. Is there a locally invariant manifold S of the form $z = g(y)$ which consists of all solutions which are bounded by $e^{(\beta-\epsilon)t}$ for t large and some $\epsilon > 0$? We will show that this is the case.

The solution with initial value zero is identically zero and exists globally in time. It folloes that for arbitrarily small initial data the solution exists for an arbitrarily long time. The identity $T^{t_1+t_2} = T^{t_1}T^{t_2}$ holds wherever both sides are defined. This property follows from the unique determination of solutions by initial data. The flow ϕ is C^1 and its first derivative $H(t,x)$ with respect to x satisfies the linear equation

$$\dot{H}(t,x) = \left[A + \frac{\partial F}{\partial x}\right] H(t,x), \quad H(0,x) = I. \tag{84}$$

In particular the equations $\dot{H}(t,0) = AH(t,0)$ and $H(0,x) = I$ hold. Hence $H(t,0) = e^{At}$. It follows that

$$\phi(t,x) = e^{At}x + \Xi(t,x), \tag{85}$$

where

$$\Xi(t,0) = 0 \quad \text{und} \quad \frac{\partial \Xi}{\partial x}(t,0) = 0. \tag{86}$$

In the construction of invariant manifolds technical difficulties can arise due to the fact that the solutions are not globally defined. To avoid this we replace the function F by one which is identical to F for x small, e.g. for $|x| \leq \frac{1}{2}s$ and vanishes for $|x| \geq s$. If we still denote the new function by F then the flow of our equation is globally defined. Then the family T^t is a group.

Lemma 4 Let $F(x)$ be a vector-valued function of class C^1 which is defined for $|x|$ small and satisfies the conditions $F(0) = 0$ and $\frac{\partial F}{\partial x}(0) = 0$. Let $\theta > 0$ be arbitrary. Then there exists a number $s = s(\theta) > 0$ (which tends to zero with θ) and a function $G(x)$ of class C^1 which is defined for all x with the properties that $G(x) = F(x)$ for $|x| \leq \frac{1}{2}s$, $G(x) = 0$ for $|x| \geq s$ and $\|\frac{\partial G}{\partial x}\| \leq \theta$ for all x.

Proof Let $s > 0$ be so small that $\|\partial F/\partial x\| \leq \theta/8$ and hence $|F(x)| \leq \theta\|x\|/8$ for $|x| \leq s$. Let $\psi(t)$ be a smooth real-valued function of t for $t \geq 0$ with $\psi(t) = 1$ for $t \leq (\frac{1}{2}s)^2$, $0 < \psi(t) < 1$ for $(\frac{1}{2}s)^2 < t < s^2$, $\psi(t) = 0$ for $t > s^2$ and $0 \leq -\frac{d\psi}{dt} \leq \frac{2}{s^2}$ for all $t \geq 0$. Let $G(x) = F(x)\psi(|x|^2)$ or $G(x) = 0$, according to whether $|x| \leq s$ or $|x| \geq s$. Then $\partial G/\partial x = 0$ for $|x| \geq s$. For $|x| \leq s$ we have $\frac{\partial G_i}{\partial x_j} = \frac{\partial F_i}{\partial x_j}\psi + 2F_i x_j \frac{d\psi}{dt}$ and hence $|\partial G/\partial x| \leq \theta/8 + 2(\theta\|x\|^2/8)(2/s^2) \leq \theta$. This completes the proof of the lemma.

Because of this result it is possible, when considering solutions close to the origin, to assume w.l.o.g. that F is globally defined and C^1, $\|\partial F/\partial x\| \leq \theta$ for all x and $F(x) = 0$ for $|x| \geq s$. Here s is allowed to depend on θ.

Next it will be shown that there exist $s_0 = s_0(s,\theta) > 0$ and $\theta_0 = \theta_0(s,\theta)$ with the property that s_0 and θ_0 tend to zero with s and θ and that ϕ is written as in (85)

$$\Xi(t,x_0) = 0, \qquad 0 \leq t \leq 1, \ |x_0| \geq s_0 \tag{87}$$

$$\|\partial \Xi/\partial x_0(t,x_0)\| \leq \theta_0, \qquad 0 \leq t \leq 1, x_0 \text{ beliebig.} \tag{88}$$

To prove this we can first use the fact that the condition on the derivative of F implies that $|F(x)| \leq \theta|x|$ and that as a consequence the solution of the

differential equation satifies $|\dot{x}| \leq c_0|x|$ where $c_0 = \|A\| + \theta$. Hence

$$\frac{d}{dt}(e^{2c_0 t}|x(t)|^2) \geq 0 \tag{89}$$

and $|x(t)| \geq |x_0|e^{-c_0 t}$. Therefore if $s_0 = se^{c_0}$ and $|x_0| \geq s_0$ then $|x(t)| \geq s$ for $0 \leq t \leq 1$. In this case the equation reduces on the interval $[0, 1]$ to $\dot{x} = Ax$. The solution with $x(0) = x_0$ is thus $x(t) = e^{At}x_0$. Hence $\Xi(t, x_0) = 0$ for $0 \leq t \leq 1$ and $|x_0| \geq s_0$.

The relation $\Xi(t, x_0) = \phi(t, x_0) - e^{At}x_0$ implies that $\frac{\partial \Xi}{\partial x_0} = H(t, x_0) - e^{At}$ or

$$\frac{\partial \Xi}{\partial x_0} = e^{At}[K(t, x_0) - I], \tag{90}$$

where $K(t, x_0) = e^{-At}H(t, x_0)$. The derivative of K is

$$\dot{K} = e^{-At}(\dot{H} - AH) = e^{-At}\frac{\partial \phi}{\partial x_0}e^{At}K. \tag{91}$$

In addition we have $K(0, x_0) = I$. The quantity $\|e^{-At}(\partial\phi/\partial x_0)e^{At}\|$ is bounded by $c_1\theta$ where $c_1 = (e^{\|A\|})^2$. For this reason $\|K(t, x_0)\|$ can be bounded by $e^{c_1\theta}$ for $0 \leq t \leq 1$. It follows that $\|\dot{K}(t, x_0)\| \leq c_1\theta e^{c_1\theta}$ and that $\|K(t, x_0) - I\| \leq c_1\theta e^{c_1\theta}$ for $0 \leq t \leq 1$. These inequalities lead to

$$\|\partial\Xi/\partial x_0\| \leq e^{\|A\|}c_1\theta e^{c_1\theta}. \tag{92}$$

Thus we have the desired condition with $\theta_0 = e^{\|A\|}c_1\theta e^{c_1\theta}$.

Now we consider the system (83) again. Let $B = e^P$ und $C = e^Q$. The eigenvalues of the matrices B and C are e^{p_j} and e^{q_k}. The norms of B and C^{-1} can be bounded by $e^{\alpha+\epsilon}$ and $e^{-\beta+\epsilon}$, respectively. To see this we use the fact that these are solutions of linear ordinary differential equations. We assume that ϵ is so small that $b = \|B\|$ and $1/c = \|C^{-1}\|$ satisfy the inequalities $b < c$ and $b < 1$. It will be supposed that F_1 and F_2 are C^1, that F_1, F_2 and their first order derivatives vanish at the origin and that the norms of these derivatives are no larger than θ for all x. In addition it is assumed that F_1 and F_2 vanish for $|x|^2 \geq s^2 > 0$. Then for each value of t the flow defines a mapping T^t of (y_0, z_0) onto (y, z) of the form

$$y(t, y_0, z_0) = e^{Pt}y_0 + Y(t, y_0, z_0), \tag{93}$$

$$z(t, y_0, z_0) = e^{Qt}z_0 + Z(t, y_0, z_0) \tag{94}$$

where Y, Z and their first order derivatives vanish at the origin, the norms of the first derivatives are no greater than θ_0 for $0 \leq t \leq 1$ and Y and Z vanish for $|y|^2 + |z|^2 \geq s_0^2$ and $0 \leq t \leq 1$.

Now a result about invariant manifolds of a mapping will be proved. It will then be applied to auf T^t with $t = 1$ to get a statement about invariant manifolds of a flow.

Lemma 5 Let B and C matrices with the properties which were listed above. Let T be a mapping of (y_0, z_0) onto (y_1, z_1) with

$$y_1 = By_0 + Y(y_0, z_0), \tag{95}$$
$$z_1 = Cz_0 + Z(y_0, z_0) \tag{96}$$

where Y and Z are C^1 and fulfil the conditions listed above. Then there exists a C^1 mapping $z = g(y)$ with $g(0) = 0$, $(\partial g/\partial y)(0) = 0$ such that the mapping R with

$$R(y, z) = (u, v) = (y, z - g(y)) \tag{97}$$

has the following properties. The mapping RTR^{-1} which maps (u_0, v_0) to (u_1, v_1) is of the form

$$u_1 = Bu_0 + U(u_0, v_0), \tag{98}$$
$$v_1 = Cv_0 + V(u_0, v_0) \tag{99}$$

where U, V and their first order derivatives vanish at the origin and $V(u_0, 0) = 0$. The last condition means that the subspace $v_0 = 0$ is invariant under the mapping RTR^{-1} and that the manifold $z = g(y)$ is locally invariant under T. If we carry out the reduction we get a globally invariant manifold for the transformed system. It can be assumed that $\theta_0 < \min\left(\frac{c-b}{4}, \frac{1-b}{2}\right)$.
Proof of the lemma If R exists the relations

$$u_1 = Bu_0 + Y(u_0, v_0 + g(u_0)), \tag{100}$$
$$v_1 = Cv_0 + Cg(u_0) + Z(u_0, v_0 + g(u_0))$$
$$-g(Bu_0 + Y(u_0, v_0 + g(u_0))). \tag{101}$$

hold. If we substitute for v_1 in the second equation we get

$$V(u, v) = Cg(u) + Z(u, v + g(u)) - g(Bu + Y(u, v + g(u))) \tag{102}$$

and $V(u, 0) = 0$ holds precisely when

$$g(u) = C^{-1}[g(Bu + Y(u, g(u))) - Z(u, g(u))]. \tag{103}$$

To prove the lemma we must show that the functional equation (103) has a C^1 solution. For this purpose a sequence g_n will be defined recursively. Let $g_0(u) = 0$ and if $g_{n-1}(u)$ has already been defined let

$$g_n(u) = C^{-1}[g_{n-1}(Bu + Y(u, g_{n-1}(u))) - Z(u, g_{n-1}(u))]. \tag{104}$$

To abbreviate this expression let $g_{n-1} = g_{n-1}(u)$, $Y^0 = Y(u, g_{n-1}(u))$, $g^0_{n-1} = g_{n-1}(Bu + Y^0)$ and $Z^0 = Z(u, g_{n-1}(u))$. It is clear that g_n is well-defined and C^1 for all n. If ∂g_n is the derivative of g_n then

$$\partial g_n = C^{-1}[\partial g^0_{n-1}(B + \partial_y Y^0 + (\partial_z Y^0)\partial g_{n-1})$$
$$-(\partial_y Z^0 + (\partial_z Z^0)\partial g_{n-1})]. \tag{105}$$

Let $\sigma = \frac{\theta_0}{c-b-3\theta_0}$, so that $0 < \sigma < 1$. It will be shown by induction that $\|\partial g_n(u)\| \leq \sigma$. The statement is obvious for $n = 0$. Suppose now that the statement holds when n is replaced by $n-1$.

$$\|\partial g_n\| \leq c^{-1}[\sigma(b + \theta_0 + \theta_0\sigma) + (\theta_0 + \theta_0\sigma)] \tag{106}$$
$$\leq c^{-1}[\sigma(b + 3\theta_0) + \theta_0]. \tag{107}$$

The last expression is equal to σ and thus the inductive step is complete.

Next it will be shown that the ∂g_n are equicontinuous. For each function $f(u)$ or $f(y, z)$ let $\Delta f = f(u + \Delta u) - f(u)$ or $\Delta f = f(y + \Delta y, z + \Delta z)$. Let $h_1(\delta) = \sup \|\Delta \partial_{y,z} Y, Z\|$ for $\|\Delta y\|, \|\Delta z\| \leq \delta$ where $\partial_{y,z} Y, Z$ denotes any of the derivatives $\partial_y Y, \partial_z Y, \partial_y Z, \partial_z Z$. It will be shown by induction that $\|\Delta \partial g_n\| \leq h(\delta)$ for $\|\Delta u\| \leq \delta < 1$, where

$$h(\delta) = \frac{4h_1(\delta)}{c - b - 4\theta_0}. \tag{108}$$

For $n = 0$ the statement is obvious. Suppose now that the statement holds when n is replaced by $n-1$. Now $\|\Delta g_{n-1}(u)\| \leq \sigma \|\Delta u\| \leq \|\Delta u\|$. It follows that

$$\|\Delta \partial_{y,z} Y^0, Z^0\| \leq h_1(\|\Delta u\|) \tag{109}$$

u and

$$\|\Delta(Bu + Y(u, g_{n-1}(u)))\| \leq (b + 2\theta_0)\|\Delta u\| \leq \|\Delta u\|. \tag{110}$$

Now we apply the general relation

$$\Delta(f_1(u) f_2(u)) = f_1(u + \Delta u)\Delta f_2 + \Delta f_1 f_2(u) \tag{111}$$

to the expression for ∂g_n.

$$\|\Delta \partial g_n\| \leq c^{-1}[h(\delta)(b + 2\theta_0) + (h_1(\delta) + h_1(\delta) + \theta_0 h(\delta))$$
$$+(h_1(\delta) + h_1(\delta) + \theta_0 h(\delta))] = c^{-1}[h(\delta)(b + 4\theta_0) + 4h_1(\delta)]. \tag{112}$$

The right hand side is $h(\delta)$. Now it will be shown that the sequence g_n converges uniformly on each bounded subset. This will hold if there exist M and r with $0 < r < 1$, so that for $n \geq 1$

$$\|g_n(u) - g_{n-1}(u)\| \leq M\|u\| r^n. \tag{113}$$

For $n = 1$ the inequlity holds provided $Mr = \sigma$. Suppose now that it holds when n is replaced by $n-1$. The quantity $c\|g_n(u) - g_{n-1}(u)\|$ can be bounded by

$$\|g_{n-1}(Bu + Y(u, g_{n-1}(u))) - g_{n-2}(Bu + Y(u, g_{n-2}(u)))\| +$$
$$\|Z(u, g_{n-1}(u)) - Z(u, g_{n-2}(u))\|. \tag{114}$$

The first term is no greater than

$$\|g_{n-1}(Bu + Y(u, g_{n-1}(u))) - g_{n-2}(Bu + Y(u, g_{n-1}(u)))\| +$$
$$\|g_{n-2}(Bu + Y(u, g_{n-1}(u))) - g_{n-2}(Bu + Y(u, g_{n-2}(u)))\|. \tag{115}$$

It follows that $c\|g_n(u) - g_{n-1}(u)\|$ is no greater than

$$M\|Bu + Y(u, g_n)\|r^{n-1} + \sigma\theta_0 M\|u\|r^{n-1} + \theta_0 M\|u\|r^{n-1}, \qquad (116)$$

which in turn is no greater than $Mr^{n-1}\|u\|(b + 4\theta_0)$. With the choice $r = (b + 4\theta_0)$ this completes the inductive step.

It follows that $g_n(u)$ converges to a limit $g(u)$, uniformly on each compact subset. It is possible to pass to the limit and see that $g(u)$ satusfies the functional equation. Since the sequence ∂g_n is pointwise bounded and equicontinuous the Arzela-Ascoli theorem implies the existence of a subsequence which converges on each compact subset. It follows that g is continuously differentiable. There exist corresponding statements when C^1 is replaced by C^r, with r finite or infinite. The condition $b < 1$ is not necessary. These sharper statements will not be proved here.

Corollary 1 Let T, $g(y)$ and θ_0 be as in the lemma. For given (x_0, y_0) we define a sequence recursively by $(y_{n+1}, z_{n+1}) = T(y_n, z_n)$. If $z_0 = g(y_0)$ then $\|(y_n, z_n)\| = O((b + \theta_0)^n)$ for $n \to \infty$. It is also the case that if $y_0 \neq 0$ then $y_n \neq 0$ for all n, $\|z_n\|/\|y_n\| \to 0$ and $\limsup n^{-1} \log \|(y_n, z_n)\| \leq \alpha$. If $z_0 \neq g(y_0)$ then $(c - 2\theta_0)^n = O(\|(y_n, z_n)\|)$ for $n \to \infty$.

If $c > 1$, so that $b < 1 < c$, then it is possible to characterize the points (y_0, z_0) of the manifold $z = g(y)$ by three alternative conditions. The first condition is that, with $(y_n, z_n) = T^n((y_0, z_0))$, $\|(y_n, z_n)\|$ converges exponentially to zero for $n \to \infty$. The second is that it converges to zero for $n \to \infty$. The third is that it stays in a neighbourhood of $(0, 0)$. In this case the manifold $z = g(y)$ is called the stable manifold of T. The corresponding manifold when n is replaced by $-n$ is called the unstable manifold of T.

Proof of the corollary $z_0 = g(y_0)$ is equivalent to the condition $v_0 = 0$. In this case $v_n = 0$ for all n. Correspondingly $u_n = Bu_{n-1} + U(u_{n-1}, 0)$ so that $\|u_n\| \leq (b + \theta_0)\|u_{n-1}\|$ and $\|u_n\| \leq (b + \theta_0)^n\|u_0\|$. In particular u_n converges to the origin. It follows that for an arbitrary $\epsilon > 0$ there exists an N with the property that $\|u_n\| \leq (b + \epsilon)\|u_{n-1}\|$ for $n \geq N$ and $\|u_{n+N}\| \leq (b + \epsilon)^n\|u_N\|$ for $n \geq 0$. Since $y_n = u_n$ und $z_n = g(u_n)$ it can be concluded that $\|(y_n, z_n)\| \leq (1 + \sigma)\|u_n\|$. In addition $\limsup n^{-1} \log \|(y_n, z_n)\| \leq \log b$. By a suitable change of variables $\log b$ can be brought arbitarily close to α and hence $\log b$ can be replaced by α in this relation. Differentiating the equation for $V(u, v)$ with respect to v gives

$$\partial_v V(u, v) = \partial_z Z(u, v + g(u)) - \partial g(Bu + Y(u, v + g(u)))\partial_z Y(u, v + g(u)). \quad (117)$$

Hence $\|\partial_v V\| \leq \theta_0 + \sigma\theta_0 \leq 2\theta_0$ and $\|V(u, v)\| \leq 2\theta_0\|v\|$. It then follows from the equation $v_n = Cv_{n-1} + V(u_n, v_n)$ that $\|v_n\| \geq (c - 2\theta_0)\|v_{n-1}\|$ and $\|v_n\| \geq (c - 2\theta_0)^n\|v_0\|$. It is also true that

$$\|(y_n, z_n)\| \geq \|(u_n, v_n)\| - \|g(u_n)\| \geq (1 - \sigma)\|(u_n, v_n)\|. \qquad (118)$$

This completes the proof of the corollary.

Theorem 7 Let T be a mapping of a neighbourhood of the origin in $\mathbb{R}^m$ to $\mathbb{R}^m$ of the form

$$x_1 = T(x_0) = \Gamma x_0 + \Xi(x_0), \tag{119}$$

where Ξ is C^1 with $\Xi(0) = 0$, $\partial\Xi/\partial x_0(0) = 0$ and Γ a matrix with d, e_0 and e eigenvalues whose modulus is less than, equal to and greater than one, respectively. Then there exists a continuously differentiable mapping R with invertible derivative and the property that RTR^{-1} is of the following form

$$\begin{aligned} u_1 &= Au_0 + U(u_0.v_0, w_0), \\ w_1 &= Bw_0 + W(u_0.v_0, w_0), \\ v_1 &= vu_0 + V(u_0.v_0, w_0). \end{aligned} \tag{120}$$

Here A, B and C are matrices which are $d \times d$, $e_0 \times e_0$ and $e \times e$ respectively and whose eigenvalues have modulus < 1, $= 1$ and > 1 respectively. U, V and W and their first order derivatives vanish at the origin and

$$V = 0, W = 0 \quad \text{wenn} \quad v_0 = 0, w_0 = 0 \tag{121}$$

$$U = 0, W = 0 \quad \text{wenn} \quad u_0 = 0, w_0 = 0. \tag{122}$$

These equations mean that the planes $v_0 = 0$, $w_0 = 0$ and $u_0 = 0$, $w_0 = 0$ are invariant planes of dimension d and e. If $e_0 = 0$ then the variables w_0 and w_1 are absent.

Proof The Lemma 5 provides a mapping R such that after the transformation with R_0 the first condition is satisfied. If we consider the mapping T^{-1} then Lemma 5 provides a mapping R_1 such that after the transformation with R_1 both conditions are satisfied.

Corollary 2 Let T^t be a group of mappings which are defined by the equations (93)-(94). Let g be the function from the Lemma 5 with $T = T^1$. Then RT^tR^{-1} has the form

$$u(t, u_0, v_0) = e^{Pt}u_0 + U(t, u_0, v_0), \tag{123}$$

$$v(t, u_0, v_0) = e^{Qt}u_0 + V(t, u_0, v_0), \tag{124}$$

where $V(t, u_0, 0) = 0$ for all t and u_0. In addition, when $y_0 \neq 0$ and $z_0 = g(y_0)$ then $z(t) = g(y(t))$ for all t, $y(t) \neq 0$ for all t, $\|z(t)\|/\|y(t)\| \to 0$ and $\limsup t^{-1} \log \|y(t)\| \leq \alpha$ for $t \to \infty$.

Proof First it will be shown that when $n \leq t \leq n+1$ there are positive constants c_1 and c_2 with

$$c_1\|(y(n), z(n))\| \leq \|(y(t), z(t))\| \leq c_2\|(y(n), z(n)). \tag{125}$$

It should be notived that $T^t = T^{t-n}T^n$. Hence

$$\|y(t) - e^{P(t-n)}y(n)\| \leq \theta_0(\|y(n)\| + \|z(n)\|) \tag{126}$$

and a similar inequality holds for $z(t)$. The inequalities (125) follow. Let $z_0 = g(y_0)$. Then the behaviour of $(y(n), z(n))$ für n large is described by Corollary

1. There results the inequality $\limsup t^{-1} \log \|(y(t), z(t))\| \leq \alpha$. If $z(t)$ is not equal to $g(y(t))$ for some t, say t_0, then $(c - 2\theta_0)^n = O(\|(y(n + t_0), y(n + t_0))\|)$ for $t \to \infty$. This would be a contradiction. Hence $z(t) = g(y(t))$ for all t. If $y(t) = 0$ for some t then $z(t) = g(y(t))$ implies that $\|z(t)\| \leq \sigma \|y(t)\| = 0$. But then $(y(t), z(t)) = 0$ for all t as a consequence of the group property of T^t.

Having proved the existence of invariant manifolds for mappings we can now return to differential equations.

Theorem 8 In the dynamical system

$$\dot{x} = Ax + F(x) \tag{127}$$

let F be continuously differentiable, $F(0) = 0$ and $\partial F/\partial x(0) = 0$. Suppose that A has r eigenvalues with negative real parts $\alpha_1 < \alpha_2 < \ldots < \alpha_r < 0$ and that the other eigenvalues, if there are any, have positive real parts. Let d be the sum of the dimensions d_i of the spaces of generalized eigenvectors corresponding to the eigenvalues α_i. If $0 < \epsilon < -\alpha_r$ then there exist solutions which satisfy the condition $\|x(t)\|e^{\epsilon t} \to 0$ for $t \to \infty$ und each such solution satisfies $\lim t^{-1} \log \|x(t)\| = \alpha_i$ for a certain i. For $\epsilon > 0$ small enough the point $x = 0$ and the set of points on solutions $x(t)$ with $\lim t^{-1} \log \|x(t)\| \leq \alpha_i$ for fixed i form a continuously differentiable locally invariant manifold of dimension $d_1 + \ldots + d_i$. The corresponding set which is defined by the condition $\limsup t^{-1} \log \|x(t)\| < 0$ is a continuously differentiable manifold of dimension d.

Proof If lim is replaced by limsup the last part of the theorem follows from Corollary 2. In addition, for $t \to \infty$ the condition $\liminf t^{-1} \log \|x(t)\| < \alpha_{i+1}$ implies the condition $\limsup t^{-1} \log \|x(t)\| \leq \alpha_i$, where α_{r+1} is interpreted as zero. Hence the condition $\limsup t^{-1} \log \|x(t)\| = \alpha_i$ implies that $\liminf = \limsup$.

Similar results are obtained when t is replaced by $-t$. The arguments used in the proofs of Theorem 8 and Corollary 2 give

Theorem 9 Let A and F be as in the last Theorem. Suppose that there are also e eigenvalues with positive real part. Let $x(t)$ be the solution of the system with $x(0) = x_0$ and T^t the corresponding mapping. Let $\epsilon > 0$. There is a mapping R with invertible derivative such that RT^tR^{-1} has the following form

$$u(t) = e^{Pt}u_0 + U(u_0, v_0, w_0),$$
$$w(t) = e^{P_0 t}w_0 + W(u_0, v_0, w_0), \tag{128}$$
$$v(t) = e^{Qt}u_0 + V(u_0, v_0, w_0). \tag{129}$$

U, V and W their first order derivatives vanish at the origin. If $v_0 = w_0 = 0$ then $V = W = 0$ and if $u_0 = w_0 = 0$ then $U = W = 0$. Moreover $\|e^P\| < 1$, $\|e^{-Q}\| < 1$ and the eigenvalues of P_0 have modulus one. The mapping R transforms the equation for x into

$$\dot{u} = Pu + F_1(u, v, w), \tag{130}$$

$$\dot{w} = P_0 w + F_2(u, v, w), \tag{131}$$
$$\dot{v} = Qv + F_3(u, v, w), \tag{132}$$

where the F_i do not have to be C^1. The planes $v_0 = w_0 = 0$ and $u_0 = w_0 = 0$ are locally invariant manifolds and are called stable and unstable manifolds. In the case that A has no eigenvalues with vanishing real part they are characterized by the condition that they consist of those solutions which converge to the origin as $t \to +\infty$ or $t \to -\infty$.

To illustrate these general ideas we can consider the fundamental model of virus dynamics with $R_0 \neq 1$. The only case in this example where a stationary solution has eigenvalues with real parts of both signs is the solution with $v = 0$ in the case $R_0 > 1$. In that case the stable manifold is two-dimensional and the unstable manifold is one-dimensional. The line $v = y = 0$ is invariant and lies in the stable manifold. The stable manifold cannot intersect the positive region due to the theorem of Korobeinikov. To understand something about the unstable manifold in this case we consider the positive eigenvalue of the linearization, call it μ_+. Let $(\hat{x}, \hat{y}, 1)$ be the components of a corresponding eigenvector. Then the relations $\hat{y} = \frac{\beta x}{\mu_+ + a}$ and $\hat{x} = -\frac{\beta x}{\mu_+ + d}$ hold. Because the second and third component of this vector are positive it is geometrically clear that one half of the unstable manifold lies in the positive region.

8 Centre manifolds

In the last section the existence of stable and unstable manifolds was proved. They both have equivalent properties and therefore we concentrate here on the stable manifold. It is an invariant manifold containing a stationary solution x_0 whose tangent space at x_0 agrees with the stable subspace in that point. In a sufficiently small neighbourhood of x_0 it is uniquely determined by this property. If the system is C^k for a natural number $k \geq 1$ the the manifold is C^k. If the system is C^∞ or analytic then the manifold has the corresponding property.

In view of these facts it is natural to ask whether there is an analogue of these invariant manifolds for the centre subspace V_c. In other words, is there an invariant manifold M_c through x_0 whose tangent space there agrees with V_c? In fact there does exist such a manifold, the centre manifold. It would be too much to prove this result in this course. Instead we limit ourselves to formulating statements about such manifolds and showing how this existence statement can be useful in concrete applications. Further information about centre manifolds can be found in [1].

If x_0 is a stationary solution of a dynamical system and V_c is non-trivial then a centre manifold exists. It does not, however, have to be unique, as can be seen in the following simple example.

$$\dot{x} = -x, \tag{133}$$
$$\dot{y} = y^2. \tag{134}$$

The origin is a stationary solution. The centre subspace is the set $x = 0$ and it is also a centre manifold in this case. It is not the only one. All solutions of this equation except the stationary one can be written in the form $x = ae^{-t}$, $y = -1/(t+b)$. If we solve for y as a function of x we get $x = Ce^{1/y}$. For $y > 0$ there exist invariant manifold which are tangent to the centre subspace to all orders. It is in fact the case that although the centre manifold is not uniquely determined its derivatives at the point x_0 are always uniquely determined. If the system is C^k, with k finite then there exists a centre manifold which is C^k. If the system is C^∞ then there does not always exist a centre manifold which is C^∞, although in this case there is a centre manifold of class C^k for each finite value of k. For an analytic system there does not necessarily exist an analytic centre manifold.

How can it be that a centre manifold is useful in applications when it is not known explicitly and when it is not even unique? As we will see later the dynamics of the restriction of the system to a centre manifold determines the whole dynamics close to a stationary solution. At the same time it is often possible to determine the dynamics on the centre manifold without knowing that manifold.

As a first example consider the system

$$\dot{x} = xy + ax^3, \tag{135}$$

$$\dot{y} = -y + cx^2. \tag{136}$$

The centre manifold is of the form $y = \phi(x)$ where $\phi(x) = O(x^2)$. It follows from the equation $\dot{y} = \phi'(x)\dot{x} = O(x^3)$ that $y = cx^2 + O(x^3)$ on the centre manifold. It can be concluded that $\dot{x} = (a + c)x^3 + O(x^4)$. The origin is unstable for $a + c > 0$. For $a + c < 0$ the coordinate x decreases along the centre manifold. We will see later that in this case the asymptotic stability of the origin follows.

In the case of the fundamental model of virus dynamics with $R_0 = 1$ the stationary solution at the point $(\lambda/d, 0, 0)$ has a one-dimensional centre manifold. The tangent space to this manifold at this point is spanned by the vector $(-au, du, dk)$. Along the centre manifold we have

$$x = \frac{\lambda}{d} - \frac{au}{dk}v + \psi_1(v), \tag{137}$$

$$y = \frac{u}{k}v + \psi_2(v). \tag{138}$$

The functions ψ_1 and ψ_2 are $O(v^2)$. If we differentiate the equation for y we get $\dot{y} = \left(\frac{u}{k} + \psi_2'\right)\dot{v} = u\psi_2 + O(v^3)$. On the other hand the evolution equation for y gives

$$\dot{y} = -\frac{\beta au}{dk}v^2 - a\psi_2 + O(v^3). \tag{139}$$

Hence $\psi_2(v) = -\frac{\beta au}{(a+u)dk}v^2 + O(v^3)$ and v decreases along the centre manifold. As in the last example it is possible to conclude the asymptotic stability of the stationary solution.

9 The Grobman-Hartman theorem

A stationary solution of a dynamical system where the linearization has no purely imaginary eigenvalues is called *hyperbolic*. Theorem 9 shows that a system can be simplified by a transformation in a neighbourhood of a hyperbolic stationary solution so that it looks more like the linearized system. But can a system be made completely linear ba a transformation in this situation? Consider the system

$$\dot{x} = Ax + F(x) \tag{140}$$

in the case that no eigenvlaue of A has vanishing real part. Does there exist a diffeomorphism R of class C^1 such that $y = R(x)$ satisfies the equation $y = Ay$? In general the answer to this question is in the negative, even in the two-dimensional case. This statement will not be proved here. If instead of a C^1 mapping we only ask for a continuous mapping then things look better. The following theorem holds.

Theorem (Grobman-Hartman) Suppose that in the system (140) none of the eigenvalues of the matrix A has vanishing real part and that F is C^1, with $F(0) = 0$ and $\partial F/\partial x(0) = 0$. Let ϕ and ψ be the flows of (140) and the system $y = Ay$, respectively. Then there exists a continuous injective mapping R from a neighbourhood of $x = 0$ to R^m such that $\mathbb{R}(\phi(t, x_0)) = \psi(t, R(x_0))$ for x_0 in a neighbourhood of the origin and t small. In particular, R maps solutions into solutions while preserving the parametrization. Thus the systems are topologically conjugate.

The corresponding statement does not always hold when there are purely imaginary eigenvalues. A generalization to that case will be presented later. As with the invariant manifolds the proof of the theorem about flows is preceded by a corresponding theorem about mappings.

Lemma 6 Let B and C be invertible matrices which are $d \times d$ and $e \times e$ and satisfy the inequalities $b = \|B\| < 1$ and $c^{-1} = \|C^{-1}\| < 1$. Let T be a mapping of the form

$$T(y_0, z_0) = (By_0 + Y(y_0, z_0), Cz_0 + Z(y_0, z_0)) \tag{141}$$

where Y and Z are functions of class C^1 which vanish at the origin together with their first order partial derivatives. Then there exists a continuous injective mapping $R(u, v) = (\Phi(u, v), \Psi(u, v))$ of a neighbourhood of the origin in $\mathbb{R}^m$ onto another which transforms T into the linear mapping $A = RTR^{-1}$ where $A(u_0, v_0) = (u_1, v_1) = (Bu_0, Cv_0)$.

As in previous proof we can cut off the function F and the mapping R for the cut-off system exists on the whole of $\mathbb{R}^m$. Before proving Lemma 6 we need two other statements.

Lemma 7 Let L be an invertible $m \times m$ matrix and let $l_1 = \|L^{-1}\|$. Let S be a mapping of the form $x_1 = S(x_0) = Lx_0 + X(x_0)$, where X is defined on the

whoile of $\mathbb{R}^m$ and satisfies a Lipschitz condition

$$\|X(x_0 + \Delta x_0) - X(x_0)\| \leq \theta_1 \|\Delta x_0\| \tag{142}$$

where $\theta_1 l_1 < 1$. Then S is injective and surjective on$\mathbb{R}^m$. If, in addition, $\|X(x_0)\| \leq c_0$ for all x_0 and the inverse of S is of the form $L^{-1} + X_1$ then $\|X_1(x_1)\| \leq l_1 c_0\|$ for all x_1.

Proof L satisfies the equation

$$\|L^{-1}[X(x_0 + \Delta x_0) - X(x_0)]\| \leq l_1 \theta_1 \|\Delta x_0\| \tag{143}$$

To see that S is injective it is enough to prove that $L^{-1}S$ is injective. If x_0 and $x_0 + \Delta x_0$ have the same image under $L^{-1}S$ then

$$0 = \|x_0 + \Delta x_0 + L^{-1}X(x_0 + \Delta x_0) - x_0 - L^{-1}X(x_0)\| \geq (1 - l_1\theta_1)\|\Delta x_0\|, \tag{144}$$

which implies that $\Delta x_0 = 0$. To see that S is surjective it is enough to prove that $L^{-1}S$ is surjective, i.e. that for given x_1 there is an x_0 with $x_1 = x_0 + L^{-1}X(x_0)$. To prove the existence of x_0 we use an iteration. Let Sei $x^0 = 0$ and $x^n = x_1 - L^{-1}X(x^{n-1})$ für $n \geq 1$. For $n \geq 2$

$$\|x^n - x^{n-1}\| \leq \|L^{-1}[X(x^{n-1}) - X(x^{n-2})]\| \leq l_1\theta_1\|x^{n-1} - x^{n-2}\|. \tag{145}$$

It follows that $\|x^n - x^{n-1}\| \leq (l_1\theta_1)^{n-1}\|x^1 - x^0\|$ for $n \geq 1$. Since $0 < b_1\theta_1 < 1$ the sequence $\{x^n\}$ has a limit for $n \to \infty$, say x_0. This shows that S is surjective. Hence S defines a bijection between x_0 and $x_1 = S(x_0)$. We have

$$X^1(x_1) = x_0 - L^{-1}x_1 = L^{-1}(Lx_0 - x_1) = -L^{-1}X(x_0) \tag{146}$$

and this completes the proof of Lemma 7.

Now we need some terminology. The matrix B of Lemma 6 has an inverse. Let $b_1 = \|B^{-1}\|$. Let a_1, θ_1 and θ be constants which satisfy the inequalities $a_1 = \max\{b, 1/c\} > 0$, $0 < b_1\theta_1 < 1$ and

$$\theta = \theta_1(1 + c) + \max\{b, c\} < 1. \tag{147}$$

If c_0 is a positive constant let $\Omega(\theta_1, c_0)$ be the set of all pairs of functions $(Y(y_0, z_0), Z(y_0, z_0))$ which satisfy the following conditions for all (y_0, z_0). $Y(0, 0) = 0$, $Z(0, 0) = 0$, $\|Y(y_0, z_0)\| + \|Z(y_0, z_0)\| \leq c_0$,

$$\|\Delta Y\|, \|\Delta Z\| \leq \frac{1}{2}\theta_1(\|\Delta y_0\| + \|\Delta z_0\|), \tag{148}$$

where Δ again denotes a difference. Lemma 6 is contained in the case $Y_1 = 0$, $Z_1 = 0$ of the following result.

Lemma 8 Let B and C be as in Lemma 6, (Y, Z) and (Y_1, Z_1) a pair of elements of $\Omega(\theta_1, c_0)$ and

$$(y_1, z_1) = T(y_0, z_0), y_1 = By_0 + Y(y_0, z_0), z_1 = By_0 + Z(y_0, z_0), \tag{149}$$
$$(y_1, z_1) = U(y_0, z_0), y_1 = By_0 + Y_1(y_0, z_0), z_1 = By_0 + Z_1(y_0, z_0). \tag{150}$$

Then there exists a unique continuous mapping

$$(u, v) = R_0(y, z), u = y + \Lambda(y, z), v = z + \Theta(y, z), \tag{151}$$

defined for all (y, z) with $\Lambda(0, 0) = 0$, $\Theta(0, 0) = 0$, Λ and Θ bounded and $R_0T = UR_0$. Moreover, R_0 is a bijection.

Proof According to Lemma 7 the mapping T has an inverse which is defined for all (y_1, z_1), say

$$T^{-1}(y_1, z_1) = (B^{-1}y_1 + Y^1(y_1, z_1), C^{-1}z_1 + Z^1(y_1, z_1)). \tag{152}$$

$\|Y^1(y_1, z_1)\|$ and $\|Z^1(y_1, z_1)\|$ can be bounded by b_1c_0. The equation $R_0T = UR_0$ is equivalent to the equations

$$By + Y + \Lambda(By + Y, Cz + Z) = B(y + \Lambda) + Y_1(y + \Lambda, z + \Theta), \tag{153}$$
$$Cz + Z + \Theta(By + Y, Cz + Z) = C(z + \Theta) + Z_1(y + \Lambda, z + \Theta) \tag{154}$$

where (y, z) is the argument of Y, Z, Λ and Θ. The first of these equations can be rewritten as

$$y+\Lambda = B[B^{-1}y+Y^1+\Lambda(T^{-1})]+Y_1(B^{-1}y+Y^1+\Lambda(T^{-1}), C^{-1}z+Z^1+\Theta(T^{-1})). \tag{155}$$

This is in fact the first component of the equation $R_0 = UR_0T^{-1}$. Hence the equation $R_0T = UR_0$ is equivalent to the following equations.

$$\Theta = C^{-1}[Z - Z_1(y + \Lambda, z + \Theta) + \Theta(By + Y, Cz + Z)], \tag{156}$$
$$\Lambda = B[Y^1 + \Lambda(T^{-1})]$$
$$+Y_1(B^{-1}y + Y^1 + \Lambda(T^{-1}), C^{-1}z + Z^1 + \Theta(T^{-1})). \tag{157}$$

The existence of R_0 is proved by showing that the last two equations have a solution using an iteration. Let $\Lambda^0 = 0$, $\Theta^0 = 0$ and

$$\Theta^n = C^{-1}[Z - Z_1(y + \Lambda^{n-1}, z + \Theta^{n-1}) + \Theta^{n-1}(By + Y, Cz + Z)], \tag{158}$$
$$\Lambda^n = B[Y^1 + \Lambda^{n-1}(T^{-1})]$$
$$+Y_1(B^{-1}y + Y^1 + \Lambda^{n-1}(T^{-1}), C^{-1}z + Z^1 + \Theta^{n-1}(T^{-1})) \tag{159}$$

for $n \geq 1$. Λ^n und Θ^n are well defined and continuous for all n. They are also bounded. It is clear that Λ^0, Θ^0, Λ^1 und Θ^1 are bounded and with the definition

$$r_n = |||\Lambda^n - \Lambda^{n-1}||| + |||\Theta^n - \Theta^{n-1}||| \tag{160}$$

where $|||\ |||$ denotes the supremum of $\|\ \|$ it follows that

$$|||\Theta^n - \Theta^{n-1}||| \leq c[\theta_1 r_{n-1} + |||\Theta^{n-1} - \Theta^{n-2}|||] \tag{161}$$
$$|||\Lambda^n - \Lambda^{n-1}||| \leq [b|||\Lambda^{n-1} - \Lambda^{n-2}||| + \theta_1 r_{n-1}]. \tag{162}$$

The sum of these two inequalities gives

$$r_n \leq [\theta_1(c + 1) + \max\{b, c\}]r_{n-1} = \theta r_{n-1}. \tag{163}$$

It follows from this that $r_n \leq r_1 \theta^{n-1}$ for all $n \geq 1$. Hence the sequences Λ^n and Θ^n converge uniformly to limits Λ and Θ which are continuous and bounded. These quantities satisfy the functional equations. Uniqueness can be proved by the usual method. To complete the proof it only remains to show that R_0 is a bijection. We denote the unique solution R_0 of the equation $R_0 T = U R_0$ by R_{TU}, so that $R_{TU} T = U R_{TU}$. If we interchange the roles of T and U we see that there exists a unique solution R_{UT} of the equation $R_{UT} U = T R_{UT}$. It follows that

$$R_{TU} R_{UT} U = R_{TU} T R_{UT} = U R_{TU} R_{UT}, \tag{164}$$

$$T R_{UT} R_{TU} = R_{UT} U R_{TU} = R_{UT} R_{TU} T. \tag{165}$$

Because of uniqueness it follows that $R_{TU} R_{UT} = R_{UU} = I$ and $R_{UT} R_{TU} = R_{TT} = I$. Hence R_{TU} and R_{UT} are bijections.

Proof of the Grobman-Hartman theorem It is assumed that A has $d > 0$ eigenvalues with positive real part and $e > 0$ eigenvalues with negative real part. The general casde can easily be obtained by adding artificial extra components. First all quanities are normalized as in Lemma 6. Let R_0 be the mapping which the lemma provides for T^1 so that $R_0 T^1 R_0^{-1} = L$. Here L is the mapping which is given by the flow of the linearized equation for time $t = 1$, i.e. $L = e^A$. Let

$$R = \int_0^1 L^{-s} R_0 T^s ds. \tag{166}$$

Then

$$L^t R = \left(\int_0^1 L^{t-s} R_0 T^{s-t} ds \right) T^t. \tag{167}$$

By introducing $s - t$ as a new integration variable the integral becomes

$$\int_{-t}^0 L^{-s} R_0 T^s ds + \int_0^{1-t} L^{-s} R_0 T^s ds. \tag{168}$$

In the first of these integrals it is possible to use the relation $L^{-s} R_0 T^s = L^{-1-s} R_0 T^{s+1}$. Hence

$$L^t R = \left(\int_0^1 L^{-s} R_0 T^s ds \right) T^t = R T^t. \tag{169}$$

To complete the proof it is enough to show that $R = R_0$. This follows from Lemma 8 with $U = L$.

The difficulties which arise when trying to replace the continuous mapping in the Grobman-Hartman theorem by a mapping of higher differentiability have to do the phenomenon of resonances. To illustrate this point we consider the simple system

$$\dot{x} = -x, \tag{170}$$

$$\dot{y} = -2y + x^2. \tag{171}$$

The general solution of this system is

$$x(t) = ae^{-t}, \tag{172}$$

$$y(t) = a^2te^{-2t} + be^{-2t}. \tag{173}$$

If there existed a C^2 diffeomorphism which tranformed the solution of the linearized system into the solution of the nonlinear system then the solution of the nonlinear system would be of the form

$$x(t) = ae^{-t} + be^{-2t} + o(e^{-2t}), \tag{174}$$

$$y(t) = ce^{-2t} + o(e^{-2t}). \tag{175}$$

However this is not the case. The problem here is that exponent on the right hand side of the second equation which arises when the expression for x is substituted in is equal to the coefficient -2 on the left hand side. In general there can be problems when one eigenvalue can be written as a linear combination of the others with integer coefficients. There is a theorem of Sternberg which says that when the coefficients of the system are C^∞ and there are no resonances the mapping in the Grobman-Hartman theorem can be chosen to be C^∞. A corresponding statement in the analytic case (under certain additional assumptions) was already proved by Poincaré in 1879.

Let x_0 be a hyperbolic stationary solution where m_+ eigenvalues of the linearization have positive real parts and m_- negative real parts. The system $\dot{x} = Ax$ is then the model for the nonlinear system if we consider toplogically equivalent systems. It is possible to ask further when linear systems are topologically equivalent to each other. It turns out that this is the case when they have the same values of m_+ and m_-. Thus we can take as model the *standard saddle*, which is defined by $\dot{x} = x$, $\dot{y} = -y$ for $x \in \mathbb{R}^{m_+}$ and $y \in \mathbb{R}^{m_-}$. This statement will not be proved here but it is possible to understand the central idea in the case of the 2×2 matrix $-I$ and a matrix with eigenvalues $-1 \pm i$. In the one case the solution curves are radial while in the other case they are spiral. It not difficult to see that the spirals can be straightened out by a homeomorphism which rotates the circles centred at the origin by different amounts. More information on this topic can be found in [7], chapter 2.

With the Grobman-Hartman theorem and the result about linear systems just mentioned it can be concluded that in a neighbourhood of a hyperbolic stationary point a dynamical system is always topologically equivalent to a standard saddle. The case where all eigenvalues have positive real parts is called a hyperbolic source and the case where they all have negative real parts is called a hyperbolic sink. A source x_0 can never be an ω-limit point of another solution. Each solution which is near enough to x_0 at some time converges to x_0 as $t \to -\infty$. There are corresponding statments for a sink. Now let x_0 be a hyperbolic stationary solution which is neither a source nor a sink. If x_0 is in the ω-limit set of a solution $x(t)$ then $x(t)$ also has ω-limit points in the stable and unstable manifolds of x_0. For a general stationary solution we have

Theorem (Shoshitaishvili) Let x_0 be a stationary solution of a dynamical system. Then in a neighbourhood of x_0 the system is topologically equivalent

to the product of the restriction of the system to a centre manifold of x_0 with a standard saddle.

It follows in particular that the restrictions of the system to two different centre manifolds of x_0 are topologically equivalent. Hence the non-uniqueness of the centre manifold is not a problem. With this theorem it is possible to prove the statments about asymptotic stability which were mentioned in the discussion of the examples of centre manifolds. There is a theorem of Takens which is a common generalization of the theorems of Sternberg and Shoshitaishvili. When there no resonances in a suitable sense the homeomorphism in the theorem of Shoshitaishvili can be shosen to be a diffeomorphism.

10 Poincaré-Bendixson theory

Having spent a long time discussing the local behaviour of solutions close to a stationary solution we now turn to global properties. One-dimensional dynamical systems are easy to analyse. Systems of dimension at least three can present great difficulties. Typical themes are chaos and strange attractors. Between there two there is dimension two which can be relatively well controlled with the help of Poincaré-Bendixson theory. This theory is the main subject of this section but defore we come to that we make some remarks about the one-dimensional case. This is the case $\dot{x} = f(x)$ where x is a scalar quantity. The stationary solutions are the zeroes of f. The set on which f is non-zero is a union of open intervals U_i. On U_i for fixed i each solution is stricly monotone. Let U_i be the interval (x_-, x_+) where the endpoints are allowed to be infinite. In each time direction the solution must tend to a limit (finite of infinite) and this limit can only be an endpoint of the interval. Suppose that the solution is monotone increasing. Then there are only three possibilities for the asymptotics in the future. The solutions tends to infinity in finite time, it exists globally and tends to infinity as $t \to \infty$ or it exists globally and tends to $x_+ < \infty$ or $t \to \infty$. There are corresponding possibilities for monotone decreasing solutions. The ω-limit set of a bounded solution is always a stationary solution. Suppose that a system is defined on an interval I and that there exist two stationary solutions x_1 und x_2 with $x_1 < x_2$ and that f does not vanish indentically on the interval $[x_1, x_2]$. Then there exists an unstable stationary solution x_3 with $x_1 < x_3 < x_2$, a fact that can be seen as follows. There exists a point $x_4 \in (x_1, x_2)$ with $f(x_4) \neq 0$. We can assume that $f(x_4) > 0$ since otherwise we could replace x by $-x$. Let (x_5, x_6) be the maximal open interval containing x_4 on which f is positive. Then we can choose $x_3 = x_5$.

Now we come to two-dimensional systems. A central tool here is the Jordan curve theorem. A Jordan curve is the set of points x in the plane of the form $x = x(t)$, $a \leq x \leq b$, where $x(t)$ is continuous, $x(a) = x(b)$ and $x(s) \neq x(t)$ for $a \leq s < t \leq b$.

Theorem (Jordan curve theorem) If J is a Jordan curve then its comple-

ment in the plane is the union of two disjoint open subsets E_1 and E_2 with $\partial E_1 = \partial E_2 = J$. One of these regions is bounded, is called the interior of J and is simply connected.

A topological space X is called simply connected if for every continuous mapping $\gamma : S^1 \to X$ there exists a continuous mapping $H : [0,1] \times S^1 \to X$ with $H(0,x) = \gamma(x)$ and $H(1,x) = \gamma(0)$ for all $x \in S^1$. Intuitively this means that each closed curve can be continuously deformed to a point.

Consider a continuous mapping $x : [a,b] \mapsto \mathbb{R}^2$ with image J. Let $\eta(t)$ be a consinuous mapping from $[a,b]$ to $\mathbb{R}^2 \setminus \{0\}$. Intuitively this is a continuous nowhere vanishing vector field on J. For $x \in \mathbb{R}^2 \setminus \{0\}$ let $\pi(x) = x/\|x\|$. This defined a mapping $\pi : \mathbb{R}^2 \setminus \{0\} \to S^1$. Let $\phi(t)$ be the angle from the positive x_1 direction to $\eta(t)$. Then $\cos\phi = \eta_1/\|\eta\|$ and $\sin\phi = \eta_2/\|\eta\|$. These formulas determine ϕ up to an integer multiple of 2π. If it is required that ϕ is continuous and its value is fixed at one point, say a, then ϕ is uniquely determined. In other words $\pi \circ \eta$ is a mapping from $[a,b]$ to S^1. The assignment $\phi \mapsto (\cos\phi, \sin\phi)$ defines a continuous mapping from $\mathbb{R}$ to S^1. We are thus looking for a mapping $\tilde{\eta}$ with the property that $p \circ \tilde{\eta} = \pi \circ \eta$. A mapping of this type exists and is unique up to an additive constant. Let $j_\eta(J)$ be defined by $2\pi j_\eta(J) = \phi(b) - \phi(a)$. If J is made by joining two curves J_1 and J_2 then $j_\eta(J) = j_\eta(J_1) + j_\eta(J_2)$. We are interested in this definition in the case that J is a Jordan curve. Here only those Jordan curves are considered which are piecewise C^1 and it will always be assumed that they are positively oriented in the sense that $(-dx_2/dt, dx_1/dt)$ always points into the interior of J. It is clear that $j_\eta(J)$ is an integer. It is called the index of J.

Theorem (Umlaufsatz) Let J be a positively oriented Jordan curve of class C^1 which is defined on $[0,1]$ and $\eta(t)$ the corresponding tangent vector field. Then $j_\eta(J) = 1$.
Proof On the triangle Δ defined by $0 \le s \le t \le 1$ a function η with values in S^1 will be defined. When $s \ne t$ and $(s,t) \ne (0,1)$ let $\eta(s,t) = [x(t) - x(s)]/\|x(t) - x(s)\|$. This function has a unique continuous extension to Δ. $\eta(t,t) = x'(t)/\|x'(t)\|$ and $\eta(0,1) = -\eta(0,0)$. Suppose that $x(0)$ is chosen in such a way that the tangent in this point is parallel to the x_1-axis and no point of the curve lies below the tangent. There is a unique continuous function $\tilde{\eta} : \Delta \to \mathbb{R}$ with $p \circ \tilde{\eta} = \eta$ and $\tilde{\eta}(0,0) = 0$. We also write ϕ instead of $\tilde{\eta}$. Then $2\pi j_\eta(J) = \phi(1,1) - \phi(0,0)$, as can be seen by consideration of $\phi(t,t)$. Now $0 \le \phi(0,t) \le \pi$ and $\phi(0,1) = k\pi$ where k is an odd integer. Hence $\phi(0,1) = \pi$. The number $\phi(s,1)$ is always between π and 2π and $\phi(1,1) = k\pi$ where k is an even integer. Hence $\phi(1,1) = 2\pi$. Since $j_\eta(J) = \phi(1,1) - \phi(0,0)$ the proof of the theorem is complete.

The essential idea of this theorem is contained in the following lemma.

Lemma 9 Let J be a Jordan curve and $\xi(t)$ and $\eta(t)$ two vector fields on J which can be deformed into each other without ever vanishing. Then $j_\xi(J) = j_\eta(J)$.

To say that the vector field can be deformed means that there exists a continuous vector field $\eta(s,t)$ for all $a \leq t \leq b$ and $0 \leq s \leq 1$ with $\eta(t,0) = \xi(t)$, $\eta(t,1) = \eta(t)$, $\eta(a,s) = \eta(b,s)$ and $\eta(t,s) \neq 0$.

Proof Let $j(s)$ be the index of $\eta(t,s)$ for fixed s. Then $j(s)$ is a continuous function of s. Since, however, $j(s)$ is also an integer it must be constant. In particular $j(0) = j(1)$.

Next we define the index of a stationary solution. Let J be a positively oriented Jordan curve on which a vector field f never vanishes. Then $j_f(J)$ is called the index of f with respect to J where $j_f(J) = j_\eta(J)$ and $\eta(t) = f(x(t))$. As in Lemma 9 it is possible to show that if J_0 and J_1 are two Jordan curves which can be deformed into each other without meeting a stationary solution of f then $j_f(J_0) = j_f(J_1)$ holds. We now consider a dynamical system which is defined on a region G. Let J be a Jordan curve in G with the property that the interior of J is also contained in G and that the vector field is non-vanishing on J and its interior. Then $j_f(J) = 0$. Since the interior of J is simply connected the curve J can be continuously deformed to a curve J_1 which is a small circle around a point x_0. Since $f(x_0) \neq 0$ the angle between $f(x)$ and the positive x_1 direction is almost constant. Since $j_f(J)$ is an integer it can only be zero. For a point x_0 the index is equal for all Jordan curves J with the properties that x_0 lies in the interior of J and there are no stationary points in the interior of J except possibly x_0 itself. This number is called the index of x_0 with respect to f. If x_0 is not a stionary solution then this index is zero. If there are only finitely many stationary solutions in the interior of J, which in turn lies in G then $j_f(J) = j_f(x_1) + \ldots j_f(x_n)$. We will not give a complete proof of this statement but the basic intuitive idea of the proof is easy to understand. The curve is deformed into a curve with the following properties. It first almost goes almost all the way around a stationary point on a small circle and then moves to a circle about another stationary point. In this way each stationary point is visited once. After that they are visited in the reverse order, where the return path between the two points is very close to the original one. In the end the curve is again close to its starting point. The summands in the formula are provided by the circles. The contributions of the path from one stationary point to another and the corresponding return path are almost opposite.

Theorem 10 Let f be a continuous function on an open set G and let $x(t)$ be a periodic solution of the equation $\dot{x} = f(x)$ with period p. If $x(t)$, $0 \leq t \leq p$ is a Jordan curve whose interior I is contained in G then I contains a stationary solution.
Proof If the Jordan curve is positively oriented then according to the Umlaufsatz we have $j_f(J) = 1 \neq 0$. Thus it cannot be the case that there are no stationary solutions in I.

Now we come to the Poincaré-Bendixson theorem.

Theorem 11 (Poincaré-Bendixson) Let f be a C^1 function on an open

subset G of $\mathbb{R}^2$ and let $x(t)$ be a solution of $\dot{x} = f(x)$ for $t \geq 0$ which is contained in a compact subset of G and which is not periodic. If there are no stationary solutions in the ω-limit set of $x(t)$ then the ω-limit set is the image of a periodic solution $y(t)$.

It will follow from the proof of Theorem 11 that when the assumptions of the theorem hold there is a monotonically increasing sequence $\{t_n\}$ with the properties that $x(t + t_n) \to y(t)$ for $n \to \infty$, uniformly on $[0, p]$ and $t_{n+1} - t_n \to p$ for $n \to \infty$. Here p is the minimal period of $y(t)$.

Proof of Theorem 11 A closed and bounded line segment L is called a transversal to the equation $\dot{x} = f(x)$ if $f(x) \neq 0$ for all $x \in L$ and der direction of $f(x)$ is not parallel to L at any point of L. Then the solution always crosses L in the same direction. The proof is divided into five steps (a)-(e).

(a) Let $x_0 \in G$, $f(x_0) \neq 0$ and let L be a line segment through x_0 which is transversal to f. It follows from the local existence theorem that that there is a neighbourhood G_0 of x_0 and $\epsilon > 0$ such that for $x_1 \in G_0$ the solution with $x(0) = x_1$ exists for $|t| \leq \epsilon$ and meets L only once. For given an arbitrary $\delta > 0$ the neighbourhood G_0 and the number ϵ can be chosen such that the difference between $x(t)$ and $x_1 + tf(x_1)$ is not greater than $\delta|t|$ for $|t| \leq \epsilon$. In particular it follows that $x(t)$ can only meet the the segment L a finite number of times for t in a bounded interval.

(b) Let L be a segment which is transversal to f and which contains the point x_0. We can assume w.l.o.g. that L is a subset of the x_2 axis. Suppose that $x(t)$ meets the segment L for values $t_1 < t_2 \ldots$ of t. Then $x_2(t_n)$ is a monotone function of n. To see this let us assume that x_1 increases at crossings of L. Consider w.l.o.g. the case that $x_2(t_1) < x_2(t_2)$. The set consisting of the curve $y(t)$, $t_1 \leq t \leq t_2$, and the segment of the x_2 axis with $x_2(t_1) \leq x_2 \leq x_2(t_2)$ is a Jordan curve J. For $t > t_2$ the point $x(t)$ is always in the interior of J or never in the interior. This follows from the fact that the solution always crosses L in one direction. It is then clear that $x_2(t_3) > x_2(t_2)$ and the argument can be repeated. The sequence $\{x_2(t_n)\}$ is monotonically increasing.

(c) Now it will be shown that the ω-limit set of $x(t)$ contains at most one point of L. If y is a point of this type then as a consequence of (a) $x(t)$ meets the segment infinitely often. As a consequence of (b) the intersection points with L converge monotonically to x_0.

(d) The ω-limit set of $x(t)$ is not empty. Let y_0 be a point of this set. The solution $y(t)$ with $y(0) = y_0$ is contained in the ω-limit set of $x(t)$. The ω-limit set of $y(t)$ is also contained in the ω-limit set of $x(t)$ and is not empty. Let z_0 be a point of this set. It follows from the assumptions of the theorem that z_0 is not a stationary solution. Thus there is a segment L_0 through z_0 which is tranverse to f. $y(t)$ crosses L_0 infinitely often. z_0 and every crossing point lie in the ω-limit set of $x(t)$. It follows from (c) that all these points coincide. Hence there exist $t_1 < t_2$ with $y(t_1) = y(t_2) = z_0$. Hence $\dot{x} = f(x)$ has a periodic solution with period $p = t_2 - t_1$. It may be assumed that p is the minimal period.

(e) Now it will be shown that the ω-limit set of $x(t)$ coincides with its subset, the image of $y(t)$. If this statement were false then the complement Z of the image of $y(t)$ in the ω-limit set of $x(t)$ would be non-empty. The image of $y(t)$ would also have to contain a cluster point x_1 of Z since the ω-limit set of $x(t)$ is connected. Let L_1 be a segment through x_1 which is tranverse to f. Each small ball about x_1 contains a point $x_2 \in Z$. Let $w(t)$ be the solution of $\dot{w} = f(w)$ with $w(0) = x_2$. The image of w is contained in the ω-limit set of $x(t)$. When x_2 is close enough to x_1 then $w(t)$ crosses L_1. This crossing can only happen at x_1 as a consequence of (c). Since x_2 is not in the image of $y(t)$ we get a contradiction.

Now the statement following the theorem will be proved. Let $y(t)$ be the periodic solution with $y(0) = y_0$ and let L_0 be a segment through y_0 which is transverse to f. Let $t_1 < t_2 < \ldots$ the consecutive crossings of L_0 by the solution $x(t)$. Then $x(t_n)$ converges monotonically along L_0 to y_0. By continuous dependence of solutions on intitial data $x(t + t_n)$ converges to $y(t)$, uniformly on $[0, p]$. For $n \to \infty$ the quantity $x(t_n+p)$ converges to $y(p) = y_0$. Thus for $\epsilon > 0$ and n large $x(t)$ crosses L_0 in the interval $[t_n + p - \epsilon, t_n + p + \epsilon]$. Hence $t_{n+1} \leq t_n + p + \epsilon$. At the same time $|x(t_n + t) - y(t)|$ is small for n large and $0 < \epsilon \leq p$. Hence there exists $\delta > 0$ with the property that $\|x(t_n + t) - y_0\| \geq \delta$ for $0 < \epsilon \leq t \leq p - \epsilon$. In particular, there is no crossing of L_0 for $\epsilon \leq t \leq p - \epsilon$. Hence $t_{n+1} \geq t_n + p - \epsilon$ for n large and the claim is proved.

Theorem 12 Let f and $x(t)$ be as in Theorem 11 except for the fact that there is a finite number n of stationary solutions in the ω-limit set of $x(t)$. If $n = 0$ then Theorem 11 can be applied. If $n = 1$ and the ω-limit set of $x(t)$ is a point then the solution converges to this point for $t \to \infty$. If $n \geq 1$ and the ω-limit set of $x(t)$ contains more than one point then the ω-limit set consists of stationary solutions $x_1, \ldots, x_n$ and a finite or infinite, but countable, set of solutions $y(t)$ on $\mathbb{R}$ with the following properties. The solution $y(t)$ has limits for $t \to +\infty$ and $t \to -\infty$ and these limits are among the points x_i.

Proof We consider the case that $n \geq 1$ and the ω-limit set of $x(t)$ contains more than one point. Since the ω-limit set is connected it contains a point y_0 which is not a stationary solution. The solution $y(t)$ with $y(0) = y_0$ is conteined in the ω-limit set of $x(t)$. Consider the case that the ω-limit set of $y(t)$ contains a point z_0 which is not a staionary solution. Then it follows from part (d) of the proof of Theorem 11 that the solution $y(t)$ is periodic. In addition, there is a neighbourhood of $y(t)$ which contains no other ω-limit points of $x(t)$. Since the ω-limit set of $x(t)$ is connected the set would have to consist of the periodic solution, a contradiction. Hence the ω-limit set of $y(t)$ is one of the points x_i. The same argument applies to α-limit points. It remains to show that the set of solutions $y(t)$ is countable. Suppose this set were uncountable. Then there would exist points x_i and x_j, not necessarily distinct, which are connected by an uncoutable set of solutions. Each of these curves, or each pair of these curves, defines a Jordan curve. If a set of these Jirdan curves is such that for each

pair the interior of one does not intersect the interior of the other then the set must be countable. Hence there must be two Jordan curves in the family whose interiors intersect. One of these, call it J_1 must be in the closure of the interior of the other, call it J_2. The interior of J_1 must also intersect the interior of another curve J_3 which is distinct from J_2. The image of the solution $x(t)$ lies between J_1 and J_2. Hence it cannot be that J_3 is in the closure of the interior of J_1. It can also not be the case that J_3 lies in the complement of the closure of the closure of the interior of J_2. Hence J_3 must lie between J_1 and J_2 and the image of $x(t)$ must lie between one of these curves and J_3. But then the other cannot be in the ω-limit set of $x(t)$, a contradiction.

The Poincaré-Bendixson theorem can often be used to prove the existence of periodic solutions. There is a simple criterion which can often be used to rule out the existence of periodic solutions of two-dimensional systems. Let $\dot{x} = f(x)$ be a two-dimensional dynamical system and let g be a real-valued function. If $\mathrm{div}(gf) \geq 0$ and $\mathrm{div}(gf)$ does not vanish identically then g is called a Dulac function. If a Dulac function exists then the system has no periodic solution whose interior lies in the domain of definition of f. In this case the integral of the component of gf in the normal direction along the closed curve is equal to the integral over the interior of $\mathrm{div}(gf)$ due to Stokes' theorem. Since the first integral vanishes and the second is strictly positive this gives a contradiction.

11 Oscillators

When a dynamical system has a periodic solution a phenomenon which is described by this solution will exhibit persistent oscillations. A situation of this kind is often called an oscillator. In order that this behaviour can actually be observed in reality it should have a certain stability. The definitions of stability are similar to those in the case of a stationary solution. Let $x(t)$ be a periodic solution with image γ. The solution is called orbitally stable if for each neighbourhood U of γ there exists a neighbourhood V of γ so that each solution which starts in V remains forever in U. The solution is called orbitally asymptotically stable if it is orbitally stable and there exists a neighbourhood U of γ with the property that for every solution that starts in U the distance of $x(t)$ from γ tends to zero for $t \to \infty$.

First some well-known examples will be discussed briefly. In the 1920's Balthasar van der Pol studied an electric circuit with nonlinear damping which leads to oscillations. A dynamical system which describes this circuit is known as the van der Pol oscillator. The original equation is a scalar second order equation. By introducing new variables it can be reduced to a first order system for two variables which belongs to the class of Liénard systems. This system has a stable periodic solution. Its discoverer called this phenomenon a relaxation oscillation. We will come back to this concept later. The conduction of electrical signals by nerve cells can be described by a four-dimensional dynamical system. This system played a central role in understanding this biological phenomenon.

Hodgkin and Huxley received the Nobel prize for medicine for their work on this subject. A simplified version of this model which is two-dimensional and still retains essential features of the full system is the Fitzhugh-Nagumo model. The Fitzhugh-Nagumo model is closely related to the van der Pol oscillator and also has a stable periodic solution. It was believed for a long time that chemical reactions always tend to equilibrium so that persistent ocillations in chemical systems were not possible. Later this idea was proved wrong by experiments. This had to do with the Belousov-Zhabotinski reaktion. The real reaction is very complicated but it is possible to construct simplified mathematical models for it. A well-known example is the Field-Noyes model, also known as the Oregonator. This system is three-dimensional so that Poincaré-Bendixson theory cannot be applied to it. A further simplification leads to the two-dimensional Brusselator.

Now a concrete example from biology will be investigated. In our bodies energy is released by the chemical processing of sugar molecules. This process is called glycolysis. The mechanism can be best investigated in simple organisms, for instance baker's yeast, *Saccharomyces cerevisiae*. This single-celled organism obtains energy from sugar and in the process produces alcohol which is used for the production of alcoholic drinks. A simple experiment is as follows. We have yeast cells in a solution and add glucose at a constant rate k_0. If k_0 is small enough alcohol is produced at a constant rate. If, however, k_0 is increased the rate of alcohol production begins to oscillate. This phenomenon was studied by Higgins and Selkov. These authors introduced a two-dimensional dynamical system to describe this experiment and today it is known as the Higgins-Selkov oscillator. If the cells are broken up and the contents extracted the oscillations are still seen. This suggests that that this is a purely chemical phenomenon which is independent of complicated structures in the cell.

Suppose we have a chemical reaction with a substrate S and a product P. The concentration of the substrate satisfies the equation $\dot{S} = k_0 - k_1 S P^2$. The substrate is supplied at the constant rate k_0 and consumed at a rate which increases with the concentration of the product. In the case of glycolysis the enzyme is phosphofructokinase (PFK) which converts fructose 6-phosphate into fructose 1,6-bisphosphate while ATP is converted to ADP. ATP is considered to be freely available and is therefore not included in the model. ADP increases the activity of PFK and therefore gives rise to a positive feedback. In the model S plays the role of the concentration of glucose and P that of the concentration of ADP. The equation for P is $\dot{P} = k_1 S P^2 - k_2 P$. The constants k_i are all positive. Due to their interpretation the quantities S and P should be positive. If they start positive they remain positive. The proof like in the case of Lemma 1. The sum of the equations gives $(S + P)^{\cdot} = k_0 - k_2 P$. Thus a solution is bounded on each interval of the form $[0, t_1)$ and the solutions exist globally in the future.

The Higgins-Selkov oscillator has a unique stationary solution, which is given by $P = \frac{k_0}{k_2}$ and $S = \frac{k_2^2}{k_0 k_1}$. The linearization at this point is

$$\frac{d\hat{S}}{dt} = -\frac{k_0^2 k_1}{k_2^2}\hat{S} - 2k_2\hat{P}, \tag{176}$$

$$\frac{d\hat{P}}{dt} = \frac{k_0^2 k_1}{k_2^2}\hat{S} + k_2\hat{P}. \tag{177}$$

The determinant of the linearization is $\frac{k_0^2 k_1}{k_2} > 0$ and the trace is $-\frac{k_0^2 k_1}{k_2^2} + k_2$, an expression which changes sign when $k_0^2 = \frac{k_2^3}{k_1}$. If both eigenvalues are real they have the same sign and this sign is determined by the trace. If the eigenvalues are complex they have the same real part and the sign of the real part is determined by the trace. We see that the stationary solution is stable when $k_0^2 > \frac{k_2^3}{k_1}$ and unstable when $k_0^2 < \frac{k_2^3}{k_1}$.

If we could show that in the case of the Higgins-Selkov oscillator with parameters for which the stationary solution is unstable there exists at least one non-stationary solution which is bounded in the future then we could conclude from the theorem of Poincaré-Bendixson that a periodic solution exists. The boundedness is, however, apparently hard to show. In fact the solution has a periodic solution for parameters of this kind but there are also unbounded solutions [12]. For this reason we now consider a different related model, the Schnakenberg model [11], which is easier to analyse. In this other model the equation for P is replaced by $\dot{P} = k_1 S P^2 - k_2 P + k_3$ where it is assumed that $k_3 < k_0$. The solutions exist globally in the future by the same argument as used in the case of the Higgins-Selkov oscillator. In the case of the Schnakenberg modell there is a unique stationary solution with $P = \frac{k_0 - k_3}{k_2}$ und $S = \frac{k_0 k_2^2}{(k_0-k_3)^2 k_1}$. The linearization at this point is

$$\frac{d\hat{S}}{dt} = -\frac{(k_0-k_3)^2 k_1}{k_2^2}\hat{S} - 2\frac{k_0 k_2}{k_0 - k_3}\hat{P}, \tag{178}$$

$$\frac{d\hat{P}}{dt} = \frac{(k_0-k_3)^2 k_1}{k_2^2}\hat{S} + \frac{(k_0+k_3)k_2}{k_0 - k_3}\hat{P}. \tag{179}$$

The determinant of the linearization is $\frac{(k_0-k_3)^2 k_1}{k_2} > 0$ and the trace is of the form $-\frac{(k_0-k_3)^2 k_1}{k_2^2} + \frac{(k_0+k_3)k_2}{k_0-k_3}$. We obtain statements about the stability of the stationary solution similar to those in the case of the Higgins-Selkov oscillator, where the boundary between the stable and unstable cases is given for the Schnakenberg model by

$$\frac{(k_0-k_3)^3}{k_0+k_3} = \frac{k_2^3}{k_1}. \tag{180}$$

Now it will be shown that the solutions of the Schnakenberg model are bounded. It follows that when the stationary solution is unstable the ω-limit set of each non-stationary solution is a periodic solution. The first step is to show that in each solution P is bounded below by a positive constant for t sufficiently large. The inequality $\dot{P} \geq k_3 - k_2 P$ holds. Integrating this gives

$$P(t) \geq \frac{k_3}{k_2} + \left(P(0) - \frac{k_3}{k_2}\right) e^{-k_2 t}. \tag{181}$$

This implies the desired statement for an arbitrary value of $P_- < \frac{k_3}{k_2}$. We can then use the inequality $\dot{S} \leq k_0 - k_1 P_-^2 S$ to see that S is bounded by a constant S_+. It follows that

$$\frac{d}{dt}(P+S) = -k_2 P + k_0 + k_3 \leq -k_2(P+S) + k_2 S_+ + k_0 + k_3. \tag{182}$$

Hence $P+S$ is bounded. Thus from Poincaré-Bendixson there exist periodic solutions of the Schnakenberg model.

We have now seen how the existence of a periodic solution of a given model can be proved. This does not, however give much information about where the solution lies. There is a possiblity, under certain circumstance, to localize the solution better for certain parameter values. This involves the relaxation oscillations mentioned previously. These ideas will now be explained in the case of the van der Pol oscillator. This oscillator is defined by the equation $\ddot{u} - k(1-u^2)\dot{u} + u = 0$ where $k > 0$ is a parameter. We replace this equation by the system

$$\epsilon \dot{u} = v - u^3/3 + u = v - G(u), \tag{183}$$

$$\dot{v} = -\epsilon u \tag{184}$$

where $\epsilon = k^{-1}$. If (u, v) satisfies the system (183)-(183) then u satisfies the original second order equation. For this system it is possible to develop the following intuitive picture. If ϵ is small and a solution is far from the curve $v = G(u)$ then the derivative of u is large and that of v is small. Thus under certain circumstances the solution has the tendency to make jumps in a horizontal direction. Otherwise it moves almost on the curve $v = G(u)$. In particular we get a picture of what a periodic solution could look like for ϵ small. This picture can be turned into a proof. The function G has the symmetry property $G(-u) = -G(u)$. It has a unique maximum at $u = -1$ and a unique minimum at $u = 1$. The maximum is $\frac{2}{3} > 0$ and the minimum is $-\frac{2}{3} < 0$. The zeroes of G are at 0 and $\pm\sqrt{3}$. The only stationary solution of the system is at the origin.

Theorem 13 Let J be the Jordan curve which consists of the following pieces: the horizontal piece which starts on $v = G(u)$ and ends at the local maximum of that curve, the horizontal piece which starts on $v = G(u)$ and ends at the local minimum of that curve and the parts of the curve $v = G(u)$ which join the ends of these lines. For ϵ small enough the van der Pol oscillator has a periodic solution with the property that for $\epsilon \to 0$ its image converges to J.

Proof For this a region H will be constructed whose points are not further from J than a given constant und which has the property that for ϵ small enough H is invariant under the flow. Moreover, this region contains no stationary solutions. As a consequence of the Poincaré-Bendixson theory this region contains a periodic solution und the definition of the region implies the desired convergence. Let h be a positive constant. Let x_1 be the point $\left(0, \frac{2}{3} + 2h\right)$. Let x_2 be the unique point on $v = G(u)$ with $v = \frac{2}{3} + 2h$. Let x_3 be the unique point

on $v = G(u) - h$ with the same u coordinate as x_2. Let x_4 be the point of $v = G(u)$ such that the tangent to the curve at that point passes through the point $\left(0, -\frac{2}{3} - 2h\right)$. The u coordinate of x_4 is $\left(\frac{2+3h}{2}\right)^{\frac{1}{3}}$. We can use the symmetry of the curve $v = G(u)$ to produce further points x_5 to x_8 from the points x_1 to x_4. Now these points will be used to construct a Jordan curve J_1. x_1 is joined to x_2 by a horizontal line, x_2 is joined to x_3 by a vertical line, x_3 is joined to x_4 by part of the curve $v = G(u) - h$ and x_4 is joined to x_5 by a straight line. Curves joining the remaining point are then determined by the symmetry. Let x_9 and x_{10} be the points $(-1, 2/3)$ und $(0, 2/3)$, let x_{11} be a point to be determined in the interior of J with $u > 1$ and $G(u) < v < 2/3$, let x_{12} be the point on the graph of $v = G(u)$ with the same u coordinate as x_{11}. Points x_{13} to x_{16} are determined by the symmetry. A Jordan curve J_2 is constructed using these points. The point x_9 to x_{12} are joined by straight lines and x_{12} is joined with x_{13} by part of the graph of G. The other points are joined using the symmetry. The curves J_1 and J_2 do not meet and J_2 lies in the interior of J_1. Let H be the closed region between J_1 and J_2. It remains to show that H has the desired properties. The origin lies in the interior of J_2 which guarantees that there are no stationary solutions in H. For $h \to 0$ the maximum distance from J of a point of J_1 converges to zero. If we assume that the distance from x_{11} to $(G^{-1}(2/3), 2/3)$ is arbitrarily small then the maximum distance from J of a point of J_2 also arbitrarily small. It remains to show that the vector field points inward everywhere on the boundary of H. On the horizontal and vertical lines this condition is satisfied, except at the endpoints. At the endpoints the vector field is tangent to the boundary but it is nevertheless the case that no solution can escape through these points. On the part of J_2 between x_{12} and x_{13} the condition is also satisfied. It remains to check three parts of the boundary. Consider first the part of the curve J_1 between x_3 and x_4. Let $g(u)$ be the slope of this curve at the point u. If for a solution v is written as a function of u then $\frac{dv}{du} = -\frac{\epsilon^2 u}{v - G(u)}$. On the part of the curve being considered at the moment the relation $v - G(u) = -h$ holds and hence $\frac{dv}{du} = \frac{\epsilon^2 u}{h} < \frac{\epsilon^2 u(x_3)}{h}$. For ϵ small enough this quantity is smaller than $g(x_4) < g(u)$. There $\dot{v} < 0$ and thus the desired condition holds. Next consider the part of J_1 between x_4 und x_5. There $|v - G(u)| > h$ and it follows that $|\frac{dv}{du}| \leq \frac{\epsilon^2 u(x_4)}{h}$. For ϵ small enough this quantity is smaller than $g(x_4)$, the slope of the line. Finally the part of J_2 between x_{10} and x_{11} will be examined. Let K be the length of the straight line from x_{11} to x_{12}. Then $|v - G(u)| > K$ on this segment. There $|\frac{dv}{du}| \leq \frac{\epsilon^2 u(x_{11})}{K}$ and this quantity tends to zero as $\epsilon \to 0$. For ϵ small enough the vector field points into the interior of H since $\dot{u} > 0$ there.

The Poincaré-Bendixson theory only applies to two-dimensional systems. How can the existence of oscillations in higher dimensional systems be proved? One possibility is the theory of monotone systems, which will now be discussed briefly. The system $\dot{x} = f(x)$ is called monotone (or cooperative) if $\frac{\partial f_i}{\partial x_j} > 0$ for all $i \neq j$. The name 'cooperative' comes from the case that the x_i are popula-

tion densities of different organisms. Then the condition means that a higher population density of one species increases the growth rate of the populations of all other species. In practise it is more common in population dynamics that $\frac{\partial f_i}{\partial x_j} < 0$ for all $i \neq j$. Then the system is called competitive. The transformation $t \to -t$ can be used to convert a cooperative system into a competitive one and conversely. In a certain sense monotone systems have a simpler asymptotic behaviour than general dynamical systems. Roughly speaking the solutions have a stronger tendency to converge to stationary solutions. It is also the case that monotone systems of n equations are not more complicated than general systems of $n-1$ equations. A detailed treatment of these ideas with precise statements cannot be given in this course. However a theorem will be proved which plays a central role in the theory and explains the name 'monotone'. Certain properties of the solutions of monotone systems can be transferred to those of competitive systems by the transformation $t \to -t$.

Theorem 14 (Müller-Kamke) Let $\dot{x} = f(x)$ be a cooperative dynamical system on a convex subset G of $\mathbb{R}^m$ and let x_0 and $\tilde{x}_0$ be points of G with $x_{0,i} \leq \tilde{x}_{0,i}$ for all i. Let $x(t)$ and $\tilde{x}(t)$ be solutions with $x(0) = x_0$ and $\tilde{x}(0) = \tilde{x}_0$ which are defined on a common time interval $[0, t_1)$. Then $x_i(t) \leq \tilde{x}_i(t)$ for all $t \in [0, t_1)$ and all i.

Proof Let $y_\epsilon(t)$ be the solution of $\dot{y}_\epsilon = f(y_\epsilon) + \epsilon$ with $y_\epsilon(0) = \tilde{x}_0$. Let t_* be the supremum of all $t < t_1$ with the property that $x_i(t) \leq y_{\epsilon,i}(t)$ for these values of t and all i. Either t_* is the upper limit of the maximal interval of existence of the solution $y_\epsilon(t)$ or there exists at least one j with the property that $x_j(t_*) = y_{\epsilon,j}(t_*)$. In the second case

$$\frac{d}{dt}(y_{\epsilon,j} - x_j) = f_j(y_\epsilon) - f_j(x) + \epsilon > 0 \tag{185}$$

for $t = t_*$. For

$$\begin{aligned} &f_j(y_\epsilon) - f_j(x) = f_j(y_\epsilon) - f_j(x_1, y_{\epsilon,2}, ..., y_{\epsilon,n}) + \dots \\ &+ f_j(x_1, x_2, \dots, x_{n-1}, y_{\epsilon,n}) - f_j(x_1, x_2, \dots, x_{n-1}, x_n). \end{aligned} \tag{186}$$

Each summand on the right hand side is non-negative by the fundamental theorem of calculus. That the intermediate points over which it is necessary to integrate are contained in G is guaranteed by the convexity. Hence the inequality $x_j(t) < y_{\epsilon,j}(t)$ holds for t slightly larger than t_* and indices j of this type. The corresponding inequality holds for all other indices by continuity. Thus we get a contradiction unless $t_* = t_1$. It follows that $x_i(t) \leq y_{\epsilon,i}(t)$ as long as the solution $y_\epsilon(t)$ exists. It can be concluded by continuous dependence of the solution on parameters that $y_\epsilon(t)$ exists for $t < t_1$ and ϵ sufficiently small and that $y_\epsilon(t)$ converges to $\tilde{x}(t)$. This proves the theorem.

It has already been mentioned that mathematical modelling played a major role in the advances which took place in the understanding of HIV in the 1990's. In that period two influential papers appeared at the same time which

used different mathematical models for the same biological system. One is the fundamental system of virus dynamics whose asymptotics we have already studied. The we call the alternative system in order to have a name for it. It differs from the fundamental model by the fact that there is an additional term in the equation for $\dot{x}$ of the form $px\left(1-\frac{x}{\bar{x}}\right)$. Here p and $\bar{x}$ are positive constants. The interpretation is that in the alternative system the fact is taken into account that the population of non-infected T cells can increase by cell division. This system can be made into a competitive system by means of a coordinate transformation. Let $x_1 = x$, $x_2 = -y$ and $x_3 = v$. Then the system for (x_1, x_2, x_3) is competitive and it is possible to use statements which are similar to those of Poincaré-Bendixson theory. It was shown in [2] that the asymptotics of solutions for $t \to \infty$ is determined by a number R_0. When $R_0 \leq 1$ all solutions converge to a stationary solution on the boundary, with $v = 0$. When $R_0 > 1$ all non-staionary solutions converge to a positive stationary solution for $t \to \infty$ for certain values of the parameters and to a nontrivial periodic solution for other values of the parameters. These two classes are distinguished by the stability of the unique positive stationary solution. It is natural to ask how two models for the same biological system can give different results. In fact the parameter values which lead to periodic solutions do not lie in the biologically reasonable range.

12 Bifurcation theory

Since it is difficult or impossible to determine the asymptotic behaviour of the most general solutions of dynamical systems of dimension three or more it makes sense to look for methods which at least allow the global dynamics to be analysed in certain limited cases. One method of this kind uses the concept of bifurcation. Suppose that a system $\dot{x} = f(x, \lambda)$ is given which depends on a parameter λ and that the equation is easy to analyse for $\lambda = 0$. Under what circumstances is the system for λ small but non-zero topologically equivalent to the system for $\lambda = 0$ and when this is not the case what is the relation between the equivalence classes of the two systems? If the systems are not equivalent we say that there is a bifurcation at $\lambda = 0$.

Consider the case that the system $\dot{x} = f(x, \lambda)$ has a stationary solution at $x = 0$ for $\lambda = 0$, i.e. $f(0, 0) = 0$. A particularly simple case is that where this stationary solution is hyperbolic. Then, in particular, the the derivative $Df(0)$ is invertible and we can apply the implcit function theorem. It follows that for x and λ sufficiently small there exists a unique solution of $x = g(\lambda)$ of $f(x, \lambda) = 0$ for fixed λ. The function g is as smooth as f. The matrix $Df(g(\lambda), \lambda)$ depends smoothly on λ. Its eigenvalues depend continuously on λ. Hence this stationary solution is hyperbolic for λ sufficiently small and the number of eigenvalues with positive real part is constant. A sink stays a sink, a source stays a source and a true saddle remains a true saddle. Thus it is seen that in this case $\lambda = 0$ is not a bifurcation point.

We now consider bifurcations in one-dimensionalen dynamical systems. The

results are also useful for systems of higher dimensions due the possibility of reduction to a centre manifold. The starting system can have an arbitrary dimension provided the centre manifold of a stationary solution is of dimension one. Let $\dot{x} = f(x, \lambda)$ be an equation for a real-valued function f. We write f' for the derivative $\frac{\partial f}{\partial x}$ and an analogous notation for derivatives of higher order. The case of a stationary solution without a bifurcation is that where $f(0, 0) = 0$ und $f'(0, 0) \neq 0$. This case is in the following sense generic. If a function of this kind is defined on a set $[-a, a] \times [-a, a]$ we can suppose that, possibly after reducing the size of a, that f' is non-vanishing on the whole set. If $\epsilon > 0$ is given we consider all functions g with the property that

$$\sup(|g(x, \lambda) - f(x, \lambda)| + |g'(x, \lambda) - f'(x, \lambda)|) \tag{187}$$

is not greater than ϵ on this set. For ϵ small enough there is a unique point x in $(-a, a)$ with $g(x, 0) = 0$ and $g'(x, 0) \neq 0$. Thus the absence of a bifurcation is stable. On the other hand for any function f on the given set with $f(0, 0) = 0$ is possible to find a function g satisfying the inequality with $g(0, 0) = 0$ and $g'(0, 0) \neq 0$. These statements can also be formulated by saying that the set of functions without a bifurcation form an open and dense subset in a suitable topology. This type of formulation will, however, not be pursued further here and the concept 'generic' will be applied in an intuitive way from this point on.

The possibilities for bifurcations are explored by assuming a certain number of conditions and then considering cases which are generic within this class. In the simplest case we assume that $f'(0, 0) = 0$, so that a bifurcation is present, and that the conditions $f''(0, 0) \neq 0$ and $\partial f/\partial\lambda(0, 0) \neq 0$ hold. This bifurcation is called a fold (or saddle node). A model for it is $f(x, \lambda) = x^2 - \lambda$. We could also take the equivalent form $f(x, \lambda) = -x^2 + \lambda$. The other similar expression $f(x, \lambda) = x^2 + \lambda$ is not topologically equivalent to the other two since the definition of topological equivalence requires the direction of time to be preserved. In this example there are no stationary solutions for $\lambda < 0$, exactly one stationary solution for $\lambda = 0$ and two stationary solutions for $\lambda > 0$, one of which is asymptotically stable and the other unstable. These features are always present for a fold. Intuitively the situation can be described as follows. When λ increases there exists a critical parameter value for which two stationary solutions (one stable and one unstable) appear out of nothing. Alternatively it can be said that when λ decreases two stationary solutions (one stable and one unstable) collide and annihilate each other.

Suppose now that a system of the form $\dot{x} = x^2 - \lambda + O(x^3)$ is given. Then it is locally topologically equivalent to the model system $\dot{x} = x^2 - \lambda$. The proof makes use of the fact that in one dimension a homeomorphism which maps stationary solutions to stationary solutions the images of solutions which connect them onto each other. Consider the system $\dot{y} = y^2 - \lambda + \psi(y, \lambda)$ where ψ is smooth and satisfies the condition $\psi(y, \lambda) = O(y^3)$. The implicit function theorem implies that the set of stationary solutions is a manifold of the form $\lambda = g(y)$ where $g(y) = y^2 + O(y^3)$. Thus for λ sufficiently small and positive there exist exactly two stationary solutions close to $x = 0$. Next a parameter-dependent homeomorphism will be constructed for small λ which defines the

equivalence. For $\lambda < 0$ the mapping h_λ is the identity. Für $\lambda > 0$ it is a linear function $h_\lambda(x) = a(\lambda) + b(\lambda)x$ where the coefficients a and b are chosen so that $-\sqrt{\lambda}$ und $\sqrt{\lambda}$ are mapped onto the two stationary solutions.

The generic fold, i.e. a system for which at the origin $f = 0$, $f' = 0$, $f'' \neq 0$ and $\partial f/\partial\lambda \neq 0$ hold is locally topologically equivalent to the model system. To prove this statement it is enough to show that the system is topologically equivalent to a system of the special form which has just been treated. A Taylor expansion of the right hand side in x gives

$$f(x, \lambda) = f_0(\lambda) + f_1(\lambda)x + f_2(\lambda)x^2 + O(x^3). \tag{188}$$

This equation can be simplified by a translation in x, $\xi = x + \delta$, where the translation δ depends on λ. Substituting into the dynamical system gives

$$\dot{\xi} = f_0(\lambda) + f_1(\lambda)(\xi - \delta) + f_2(\lambda)(\xi - \delta)^2 + O((\xi - \delta)^3). \tag{189}$$

Sorting the terms according to powers of ξ gives

$$\begin{aligned}\dot{\xi} &= [f_0(\lambda) - f_1(\lambda)\delta + f_2(\lambda)\delta^2 + O(\delta^3)] \\ &+[f_1(\lambda) - 2f_2(\lambda)\delta + O(\delta^2)]\xi \\ &+[f_2(\lambda) + O(\delta)]\xi^2 + O(\xi^3).\end{aligned} \tag{190}$$

The condition that the coefficient of ξ vanishes is that

$$F(\lambda, \delta) = f_1(\lambda) - 2f_2(\lambda)\delta + \psi(\lambda, \delta)\delta^2 = 0 \tag{191}$$

for a smooth function ψ. We have $F(0, 0) = 0$, $\frac{\partial F}{\partial \delta}(0, 0) = -2f_2(0) \neq 0$ and $\frac{\partial F}{\partial \lambda}(0, 0) = f_1'(0)$. The implicit function theorem implies that there exists a smooth function $\delta = \delta(\lambda)$ with $\delta(0) = 0$ and $F(\lambda, \delta(\lambda))(0) = 0$. It also follows that $\delta(\lambda) = \frac{f_1'(0)}{2f_2(0)}\lambda + O(\lambda^2)$. After the transformation the equation for $\dot{\xi}$ contains no terms that are linear in ξ. Consider a new parameter $\mu = \mu(\lambda)$ where the right hand side is the coefficient of ξ^0 of the expansion in powers of ξ. Then $\mu(\lambda) = f_0'(0)\lambda + \lambda^2\phi(\lambda)$ with a smooth function ϕ. $\mu(0) = 0$ and $\mu'(0) = f_0'(0) = \frac{\partial f}{\partial \lambda}(0, 0)$. By the inverse function theorem there is a smooth inverse $\lambda = \lambda(\mu)$ with $\lambda(0) = 0$. Hence the equation for $\dot{\xi}$ becomes $\dot{\xi} = \mu + b(\mu)\xi^2 + O(\xi^3)$. Here b is a smooth function with $b(0) = f_2(0) \neq 0$. Let $\eta = |b(\mu)|\xi$ und $\beta = |b(\mu)|\mu$. Then $\dot{\eta} = \beta + s\eta^2 + O(\eta^3)$ where $s = \pm 1$ corresponds to the sign of $b(0)$.

Next the case is considered that at the origin f, f' und f'' vanish but f''' does not vanish. It turns out that to ensure that a model for all parameter values close to zero is obtained it is necessary to consider the case with two parameters. Thus in this case we need a function $f(x, \lambda_1, \lambda_2)$. To get the generic case it is assumed that at the origin $\frac{\partial^2 f}{\partial x \partial \lambda_1}\frac{\partial f}{\partial \lambda_2} - \frac{\partial^2 f}{\partial x \partial \lambda_2}\frac{\partial f}{\partial \lambda_1} \neq 0$. This bifurcation is called a generic cusp. A model for this bifurcation is $\lambda_1 + \lambda_2 x - x^3$. It contains points where there is a fold. They are points away from the origin where the two equations $\lambda_1 + \lambda_2 x - x^3 = 0$ and $\lambda_2 - 3x^2 = 0$ hold. It is possible to eliminate x from these equations, with the result $4\lambda_2^3 = 27\lambda_1^2$. When the parameters lie

outside the region bounded by the cusp the system has exactly one stationary solution and it is stable. Inside the cusp there are three stationary solutions, two stable and one unstable.

A system which contains a generic cusp will now be simplified by a procedure which is similar to that used in the case of the fold. To make the notation more concise let $\lambda = (\lambda_1, \lambda_2)$. A Taylor expansion in x gives

$$f(x, \lambda) = f_0(\lambda) + f_1(\lambda)x + f_2(\lambda)x^2 + f_3(\lambda)x^3 + O(x^4). \tag{192}$$

The conditions $f_1(0) = 0$ and $f_2(0) = 0$ hold. As in the case of the fold we try to achieve a simplification by a translation $\xi = x + \delta$. Substituting into the dynamical system gives

$$\begin{aligned}\dot{\xi} &= [f_0(\lambda) - f_1(\lambda)\delta + \omega(\lambda, \delta)\delta^2] + [f_1(\lambda) - 2f_2(\lambda)\delta + \phi(\lambda, \delta)\delta^2]\xi \\ &+[f_2(\lambda) - 3f_3(\lambda)\delta + \psi(\lambda, \delta)\delta^2]\xi^2 + [f_3(\lambda)\delta + \theta(\lambda, \delta)]\xi^3 + O(\xi^4)\end{aligned} \tag{193}$$

for smooth functions ω, ϕ, ψ and θ. Because $f_2(0) = 0$ it is not possible in this case as for the fold to use the implicit function theorem to eliminate the linear term in ξ. We can, however, eliminate the quadratic term in ξ. To do this let $F(\lambda, \delta) = f_2(\lambda) - 3f_3(\lambda)\delta + \psi(\lambda, \delta)\delta^2$ and notice that $F(0, 0) = 0$ and $\frac{\partial F}{\partial \delta}(0, 0) = -3f_3(0) \neq 0$. Consider new parameters μ_1 and μ_2 with

$$\mu_1(\lambda) = f_0(\lambda) - f_1(\lambda)\delta(\lambda) + \delta^2(\lambda)\omega(\lambda, \delta(\lambda)), \tag{194}$$

$$\mu_2(\lambda) = f_1(\lambda) - 2f_2(\lambda)\delta(\lambda) + \delta^2(\lambda)\phi(\lambda, \delta(\lambda)). \tag{195}$$

The quanitity $\mu = (\mu_1, \mu_2)$ satisfies $\mu(0) = 0$. The new parameters can be introduced if the Jacobian determinant $\det(\partial\mu/\partial\lambda)$ is not zero. This condition is equivalent to the second condition for the generic cusp. Then the inverse function theorem can be used. We get a smooth inverse $\lambda = \lambda(\mu)$ with $\alpha(0) = 0$. After the transformation to the new parameters the equation is of the form

$$\dot{\xi} = \mu_1 + \mu_2\xi + c(\mu)\xi^3 + O(\xi^4) \tag{196}$$

where $c(\mu) = f_3(\lambda(\mu)) + \delta(\lambda(\mu))\omega(\lambda(\mu), \delta(\lambda(\mu))$ is a smooth function of μ and $c(0) = f_3(0) = \frac{1}{6}f'''(0, 0) \neq 0$. Finally the linear scaling $\eta = \sqrt{|c(\mu)|}\xi$ is carried out and new parameters $\beta_1 = \sqrt{|c(\mu)|}\mu_1$ und $\beta_2 = \mu_2$ are defined. The result is

$$\dot{\eta} = \beta_1 + \beta_2\eta + s\eta^3 + O(\eta^4). \tag{197}$$

It has now been shown that in the case of the cusp an approximate normal form can be obtained. It is also possible, as in the case of the fold, to go further and get an exact normal form. However the proof will not be given here. There is a special case of the cusp, the pitchfork, in which the system has the symmetry $x \mapsto -x$. This bifurcation is generic in the class of systems with this symmetry.

It was already mentioned that statements about bifurcation theory in one dimension can be applied to problems in higher dimensions. Now this remark will be developed further. For this we consider a two-dimensional dynamical

system $\dot{x} = f(x, \lambda)$ with $f(0, 0) = 0$ as an example. Suppose that the rank of the mapping $Df(0, 0)$ is one and that the non-vanishing eigenvalue of this metrix is negative. The dimension of the centre manifold is one. Now consider the extended system

$$\dot{x} = f(x, \lambda), \tag{198}$$

$$\dot{\lambda} = 0. \tag{199}$$

In this system the parameter has become an unknown. The extended system has a stationary point at $(0, 0, 0)$ and the centre manifold at that point is two-dimensional. Call it M. The intersections M_λ of the λ coordinate planes with M are invariant manifolds for the systems in the parameter-dependent setting. The manifold M_0 is a centre manifold of the origin in the system for $\lambda = 0$. If the systems for $\lambda \neq 0$ have stationary solutions which are close enough to the origin in the (x, λ) space then they must lie on M_λ. The manifold M_λ is invariant but is in general not a centre manifold for the stationary solutions which it contains. Let us now get back to the two-dimensional system and let us suppose that that the restriction of the system to M has a fold bifurcation. We now how this situation can be characterized for the restriction. Using the theorem of Shoshitaishvili we then also know what the qualitative behaviour of the original system looks like in a neighbourhood of the origin. For $\lambda < 0$ there are no stationary solutions. For $\lambda > 0$ there are two. One is asymptotically stable while the other is a saddle. For $\lambda = 0$ there is exactly one stationary solution and it is a saddle node. This means that it has a neighbourhood which is divided by the stable manifold into two regions, one which looks like a node (hyperbolic sink) and the other like a saddle. This is the origin of the alternative name for this bifurcation. Whether the system on the centre manifold has the desired properties can be checked by direct calculations with the full system. Here these conditions will be stated without proof. In order that there is a fold at $(0, 0)$ the following conditions must hold. $f(0, 0) = 0$, the rank of $Df(0, 0)$ is one. The condition $l(D^2 f(r, r)) \neq 0$ holds, where l and r are left and right eigenvectors of the matrix $Df(0, 0)$ with eigenvalue zero. Finally the condition $l(\partial f/\partial\lambda) \neq 0$ holds. There is a similar characterization of the cusp in higher dimensions which, however, looks a bit more complicated than one might guess. The conditions are $f(0, 0) = 0$, the rank of $Df(0, 0)$ is one, $l(D^2 f(r, r)) = 0$, $l(D^3 f(r, r, r)) - 3D^2 f(r, z) \neq 0$ for a certain vector z whose definition will not be given here. The correction in the condition for the third derivative is related to the fact that it is necessary to work on the centre manifold and not on the centre subspace. The condition containing derivatives with respect to the parameters must be generalized appropriately. Further details on these matters can be found in the book of Kuznetsov [7].

We have now considered some aspects of the case where the linearization of the system at the bifurcation point has zero eigenvalue. We next consider the case that there is a purely imaginary eigenvalues which are not zero. For this it is of course necessary to have a system of dimension at least two. The model

system in this case is

$$\dot{x}_1 = \lambda x_1 - x_2 - x_1(x_1^2 + x_2^2), \tag{200}$$
$$\dot{x}_2 = x_1 - \lambda x_2 - x_2(x_1^2 + x_2^2). \tag{201}$$

The system has a stationary point at the point $(0, 0)$ for all λ and the Jacobian matrix there is

$$\begin{bmatrix} \lambda & -1 \\ 1 & \lambda \end{bmatrix} \tag{202}$$

with eigenvalues $\lambda \pm i$. In this form the structure of the system is not very transparent. It can be seen better in polar coordinates. It is helpful for the calculations to introduce the complex quantity $z = x_1 + ix_2$. Then $|z|^2 = x_1^2 + x_2^2$ and

$$\dot{z} = (\lambda + i)z - z|z|^2. \tag{203}$$

With $z = \rho e^{i\phi}$ we get

$$\dot{\rho} = \rho(\lambda - \rho^2), \tag{204}$$
$$\dot{\phi} = 1. \tag{205}$$

These equations are decoupled. The first equation is of course only relevant for $\rho \geq 0$. For $\lambda < 0$ the stationary solution at $\rho = 0$ is asymptotically stable and hyperbolic. For $\lambda = 0$ it is still stable but no longer hyperbolic. For $\lambda > 0$ the stationary solution at $\rho = 0$ is unstable and a new stable stationary solution appears at $\rho = \sqrt{\lambda}$. The second equation describes a rotation with constant speed. If we combine these facts we obtain the following picture of the dynamics of the two-dimensional system. For $\lambda < 0$ there is a hyperbolic sink at the origin. Solutions away from the origin approach it while spiralling as $t \to \infty$. The system for $\lambda = 0$ is topologically equivalent to the system for $\lambda < 0$. For $\lambda > 0$ there is a stable periodic solution which encircles the origin. The bifurcation which has just been described is the Hopf bifurcation. There is a similar bifurcation in which the sign in the nonlinear term in the model system is reversed. In that case the periodic solution is unstable. These two cases are called supercritical and subcritical, respectively.

As in the case of the bifurcations considered up to now we would like to relate more general systems with the model system. If a term is added to the right hand side of the model system which is $O(|x|^4)$ then the resulting system is topologically equivalent to the model system. We now consider generic Hopf bifurcations. If a two-dimensional system has a stationary solution at the origin and the eigenvalues there are purely imaginary but non-zero then it follows from the implicit function theorem that for λ small but non-zero there exists exactly one stationary solution close to the origin. After doing a λ-dependent coordinate transformation we can assume that $f(0, \lambda) = 0$ for all λ. In order that the bifurcation is generic and the system topologically equivalent with the model it suffices that two conditions are satisfied. The first says that the real part of the eigenvalue of the Jacobian matrix at the origin moves through zero

with positive velocity for $\lambda = 0$. The second condition is that the first Lyapunov coefficient does not vanish. This quantity is a combination of the derivatives of f of second and third order of f at the origin. The supercritical and subcritical cases are distinguished by the sign of the Lyapunov coefficient. By means of centre manifold reduction it is also possible to define Hopf bifurcations in higher dimensions.

Having investigated the van der Pol oscillator in the limit $k \to \infty$ we will now consider the case $k \to 0$ with the aim of finding a Hopf bifurcation. It turns out that in order to do this it is best to rescale the equations [13]. Starting from the first order system for u and v we define $x = k^{\frac{1}{2}}u$ and $y = k^{\frac{3}{2}}v$. The transformed system is

$$\dot{x} = y - x^3/3 + kx = 0, \tag{206}$$

$$\dot{y} = -x. \tag{207}$$

The eigenvalues at $(0,0)$ are $\frac{1}{2}(k \pm \sqrt{k^2 - 4})$. The first condition for a generic Hopf bifurcation is fulfilled. To check the second condition first notice that the linear part of the system for $k = 0$ is already in standard form. Under these circumstances it is easy to apply a formula for the Lyapunov coefficient which is given in [9]. The coefficient is $-\frac{3\pi}{2}$ and this is a supercritical Hopf bifurcation. It follows that there is a stable stationary solution close to the origin for $k > 0$. The amplitude of the oscillation, measured in the transformed variables is proportional to $\sqrt{k}$ in leading order. In the original variable u the amplitude is independent of k in leading order, so that there could not be a generic Hopf bifurcation in the original variables.

13 The Lorenz system

In this course we have stayed in relatively calm waters. At the same time it is important to know that in the sea of dynamical systems there are many storms. The Lorenz system is a system of three ordinary differential equations which depends on three parameters. At first sight it looks harmless:

$$\dot{x} = \sigma(y - x), \tag{208}$$

$$\dot{y} = rx - y - xz, \tag{209}$$

$$\dot{z} = xy - bz. \tag{210}$$

The meteorologist Edward Lorenz obtained this system as a model system for convective rolls in the atmosphere. The parameters σ and r have direct physical interpretations as the Prandtl number and the Rayleigh number, respectively. The parameter b does not have a special name. All parameters are assumed positive. The system is symmetric under the transformation $(x, y, z) \mapsto (-x, -y, z)$. The origin is a staionary solution for all values of the parameters. All stationary solutions satisfy $x = y$, $x^2 = bz$ and $(r-1)x = b^{-1}x^3$. It follows that there are no stationary solutions away from the origin for $r < 1$. For $r > 1$ there are two stationary solutions with $x = y = \pm\sqrt{b(r-1)}$ and $z = r - 1$.

Now the stability of the stationary solutions will be investigated. In linearization at the origin the equation for z decouples. The solution of the linearized equation for z decays exponentially. The linearization in x and y gives a matrix with determinant $\sigma(1-r)$. When $r > 1$ there is one positive and one negative eigenvalue and the origin is a saddle point. The satble manifold is two-dimensional and the unstable manifold one-dimensional. The trace is $-\sigma - 1$ and hence the origin is a hyperbolic sink for $r < 1$ and thus asymptotically stable. We can say more with the help of a Lyapunov function. Let $V(x,y,z) = \frac{1}{\sigma}x^2 + y^2 + z^2$. This function vanishes only at the origin where it has a global minimum.

$$\begin{aligned}\frac{1}{2}\dot{V} &= (r+1)xy - x^2 - y^2 - bz^2 \\ &= -\left[x - \frac{r+1}{2}y\right]^2 - \left[1 - \left(\frac{r+1}{2}\right)^2\right]y^2 - bz^2. \end{aligned} \tag{211}$$

$\dot{V} \leq 0$ and so V is a Lyapunov function when $r < 1$. In addition $\dot{V}$ vanishes only at the origin. Thus in this case the origin is globally asymptotically stable. For the Jacobian matrix of the other two stationary solutions the characteristic equation is

$$\lambda^3 + (\sigma + b + 1)\lambda^2 + (r+\sigma)b\lambda + 2b\sigma(r-1) = 0. \tag{212}$$

For r a little greater than one the Routh-Hurwitz criterion implies that all eigenvalues have negative real part. Hence the real parts of the eigenvalues stay negative as long as they do not meet the imaginary axis away from the origin. When $\lambda = i\omega$ with ω real this equation becomes the conditions $\omega^2 = (r+\sigma)b$ and $(\sigma + b + 1)\omega^2 = 2b\sigma(r-1)$. We can eliminate ω from these equations with the result that $r = r_H = \sigma\left(\frac{\sigma+b+3}{\sigma-b-1}\right)$. This formula only gives a relevant solution when the result is positive. At $r = r_H$ there is a Hopf bifurcation where the two stable solutions become unstable. If this bifurcation were supercritical it could provide a candidate for the ω-limit set of other solutions. In fact it is subcritical.

What is the long-time beahviour of solutions of the Lorentz system? We have now excluded some simple candidates for ω-limit sets. Could it be that the solutions tend to infinity for $t \to \infty$? It will now be shown that this alternative can also be excluded. Suppose that at some time a solution satisifies the inequality $2\sigma x^2 + 2y^2 + b^2z^2 \geq C_1$ for a constant C_1.

$$\begin{aligned}&\frac{d}{dt}[x^2 + y^2 + (z - r - \sigma)^2] = 2[-\sigma x^2 - y^2 - bz^2 + b(r+\sigma)z] \\ &\leq -2\sigma x^2 - 2y^2 - bz^2 + b(r+\sigma)^2 \\ &\leq b(r+\sigma)^2 - C_1. \end{aligned} \tag{213}$$

It follows that when $C_1 > b(r+\sigma)^2$ the quantity $x^2 + y^2 + (z - r - \sigma)^2$ is decreasing at a positive rate. If for a constant $C_2 > 0$ the ball K defined by

the inequality $x^2 + y^2 + (z - r - \sigma)^2 \leq C_2$ contains the region E defined by the inequality $2\sigma x^2 + 2y^2 + b^2 z^2 \leq C_1$ then a solution which starts outside K must reach K in finite time and stay there at all later times. This proves in particular that the solutions of the Lorenz system are bounded. The divergence of the vector field defining the Lorenz system is $-\sigma - 1 - b < 0$. The divergence is a constant. If we consider the volume v of a region like K then we see, using Stokes theorem, that $\dot{v} = -(\sigma + 1 + b)v$. The volume tends exponentially to zero as $t \to \infty$.

In his original paper Lorenz chose the parameter values $\sigma = 10$, $b = \frac{8}{3}$ and $r = 28$. Later many authors explored the parameter space, usually varying r and fixing the other two parameters. For the given values of the other parameters the Hopf bifurcation we already mentioned takes place at a value of r which is approximately 24.74. The unstable manifold of the origin is one-dimensional and it can be asked where it goes. For a certain value of r, approximately 13,926 it comes back to the origin and there is a homoclinic solution. If when starting from this value r is increased the unstable periodic solutions which end at the Hopf bifurcation originate from the homoclinic solution. The bifurcations which we discussed up to now are local bifurcations. In that case everything happens close to a stationary solution. In contrast the homoclinic bifurcation in the Lorenz system is a global bifurcation. The picture which has just been described is based on numerical simulations but it is apparently at least proved that a homoclinic solution exists. There is a theory which classifies homoclinic bifurcations. In some cases of the classification there is very complicated behaviour near the homoclinic solution.

For the parameter values of Lorenz computer simulations show that the solutions converge to a set which is known as the Lorenz attractor. On the attractor itself the solution cannot be localized. They jump back and forth between two pieces of the attractor which look like disks. The disks are not manifolds but rather like many layers lying on top of each other, a bit like puff pastry. Some of these statements were proved by Warwick Tucker. It was a computer-assisted proof. This means that the proof used a computer but is nevertheless a rigorous proof. The technique used is interval arithmetic.

The Lorentz attractor is often called a 'strange attractor'. In this context there are several problems. The first is the definition of attractor. There is more than one definition in the literature and it is not clear if one of them is the best. The sam is true of the definition of 'strange' in this context. The third problem is, when a definition has been fixed, to show that the Lorenz system (or another system) contain a set which has these properties. Here we will only try to sketch the intuitive ideas which play a role in the consideration of these questions. One possible definition of an attractor is as follows. It is a set A which is invariant under the flow, which has an open neighbourhood U with property that every solution which starts in U converges to A as $t \to \infty$ and which is minimal in the sense that it is not the union of two other sets which have the first two properties. The attractor is called strange if it shows sensitive dependence of the solutions on initial data. The last condition means that that for sufficiently many initial data on the attractor the maximal Lyapunov exponent of the solution is positive.

This last quantity measures how fast neighbouring solutions move awy from the given solution. When the attractor is a point this exponent is the largest real part of an eigenvalue of the linearization.

14 Virus dynamics in practise

How are the models of virus dynamics used in practise? A stationary solution of the model corresponds to a person who is infected with HIV. Suppose that this person is treated with a drug which prevents new cells being infected. To model the state during treatment we set the parameter β to zero. Then the equations for y and v decouple from the equation for x. We then have the system

$$\dot{y} = -ay, \tag{214}$$
$$\dot{v} = ky - uv. \tag{215}$$

These equations are linear and can be solved explicitly, with the result

$$y(t) = y^* e^{-at} \tag{216}$$
$$v(t) = \frac{v^*(ue^{-at} - ae^{-ut})}{u - a}. \tag{217}$$

The population of infected cells decays exponentially and after a certain time the number of virions does the same. Let us make the plausible assumption that the free virions are eliminated faster than the infected cels die, i.e. $u > a$. Then the number of virions is approximately proportional to e^{-at}.

A somewhat more complicated case is obtained if a drug is considered which has the effect that neuly produced virions are defective and are not able to infect new cells. Let the population of defective virions be denoted by w while v is the population of functional virions. Then the following equations hold

$$\dot{y} = \beta xv - ay, \tag{218}$$
$$\dot{v} = -uv, \tag{219}$$
$$\dot{w} = ky - uw. \tag{220}$$

These equations are no longer decoupled from the equation for x. If, however, we consider a time interval on which x changes very little then we can replace x by a constant. Then the equations obtained in this way can be solved explicitly, with the result

$$y(t) = \frac{y^*(ue^{-at} - ae^{-ut})}{u - a}, \tag{221}$$
$$v(t) = v^* e^{-ut}, \tag{222}$$
$$w(t) = v^* \left[(e^{-at} - e^{-ut}) \frac{u}{u - a} - ate^{-ut} \right] \frac{u}{u - a}. \tag{223}$$

With the assumption that $u > a$ the total number of virus particles $v + w$ is approximately proportional to e^{-at} after a certain time. The half-life of the

infected cells is the same for both types of drug. The reverse transcriptase inhibitors are of the first type and the protease inhibitors of the second. Both possibilities were tried in the fundamental experiments on this subject.

In one case the experiment was as follows. Twenty HIV-infected patients were treated with a protease inhibitor and the concentration of virus was measured approximately every four days. (Before treament the concentration was approximately constant.) If this concentration decayed exponentially then $\log v$ would be a linear function of t and it would be possible to read off the exponent from the slope of this line. In the data a linear dependence is found and it turns out that the half-life of the infected cells is about two days. Under these assumptions the half-life of the virus must be even less. Later experiments were done where shortly after the beginning of treatment measurements were made much more frequently (every two hours). In this way it was possible to measure the parameter u. This gives an estimate of how fast virions were being eliminated from the system before treatment and, correspondingly, how fast new virions were being produced. This gives a rate of 10^9 per day.

15 Sources

Most themes which occur in these notes are treated in many places in the literature. Here we list some of the sources which we relied on most when preparing these lectures. The main source for sections 2, 4, 7, 9 and 10 is [4]. The biggest difference here is that we concentrate on the case where uniqueness holds. This makes many proofs simpler and, it seems to the author, more transparent. The main source for parts of sections 5 and 11 is [3]. The main source for section 6 is [10]. The main source for sections 12, 13 and 14 are [7], [13] and [8], respectively.

I would like to thank Gerrit Pfluger for pointing out a mistake in an earlier version of this manuscript.

References

[1] Carr, J. 1981 Applications of centre manifold theory. Springer, Berlin.

[2] De Leenheer, P. und Smith, H. 2003 Virus dynamics: a global analysis. SIAM J. Appl. Math. 63, 1313–1327.

[3] Hale, J. K. 2009 Ordinary Differential Equations. Dover, Mineola.

[4] Hartman, P. 1982 Ordinary Differential Equations. Birkhäuser, Basel.

[5] Korobeinikov, A. 2004 Global properties of basic virus dynamics models. Bull. Math. Biol. 66, 879–883.

[6] Korobeinikov, A. 2004 Lyapunov functions and global properties for SEIR and SEIS epidemic models. Math. Medicine and Biol. 21, 75–83.

[7] Kuznetsov, Y. A. 2010 Elements of Applied Bifurcation Theory. Springer, Berlin.

[8] Nowak, M. A. und May, R. A. 2000 Virus Dynamics. Oxford University Press, Oxford.

[9] Perko, L. 2001 Differential equations and dynamical systems. Springer, Berlin.

[10] Rudin, W. 1987 Real and complex analysis. McGraw-Hill, New York.

[11] Schnakenberg, J. 1979 Simple chemical reaction systems with limit cycle behaviour. J. Theor. Biol. 81, 389–400.

[12] Selkov, E. E. 1968 Self-oscillations in glycolysis. I A simple kinetic model. Eur. J. Biochem. 4, 79–86.

[13] Strogatz, S. H. 1994 Nonlinear dynamics and chaos. Perseus, Cambridge.