Dynamical Systems and Nonlinear Ordinary Differential Equations

Lecture Notes

Christian Schmeiser¹

Contents

1. Introduction	1
2. Linear systems	7
2.1. Inhomogeneous linear ODE systems	9
3. Scalar ODEs – stability	10
4. Hyperbolic stationary points – linearization	11
5. Scalar ODEs – bifurcations	15
5.1. The fold	15
5.2. The transcritical bifurcation	16
5.3. The pitchfork bifurcation	16
5.4. The spruce budworm – the cusp bifurcation	17
6. Scalar iterated maps – bifurcations and chaos	18
7. Invariant regions – Lyapunov functions	20
8. Limit cycles	24
8.1. Multiple scales	24
8.2. The Poincaré map	27
8.3. Relaxation oscillations	27
8.4. The Hopf bifurcation	29
8.5. The Poincaré-Bendixson theorem	30
9. The Lorenz equations	33
10. Hamiltonian mechanics	36
Appendix 1 – second order Taylor remainders	43
Appendix 2 – Young's inequality	44
References	44

1. INTRODUCTION

Most of the material for this course can also been found in the books [2, 4, 5], and we do not give detailed references to these in the following.

We consider models for the time evolution of systems, whose state can be described by a finite number of parameters.

- Therefore states will always be points in \mathbb{R}^n , $n \in \mathbb{N}$, the state space.
- Time will be assumed to either evolve *continuously or in discrete steps*.

 $^{^{1}\}mathrm{Fakultät}$ für Mathematik, Universität Wien, Oskar-Morgenstern-Platz 1, 1090 Wien, Austria. christian.schmeiser@univie.ac.at

- We shall also assume that the state at a certain time completely determines all later states, and finally
- we assume that the environment for our system does not change with time.

Starting with the case of discrete time $k \in \mathbb{Z}$, the assumption that the state $u_k \in \mathbb{R}^n$ at time k determines the state at time k + 1 means that there is a map $f_k : \mathbb{R}^n \to \mathbb{R}^n$ such that $u_{k+1} = f_k(u_k)$. Since we also assume that the environment does not change with time, the rule for the time step from k to k + 1, i.e. the map f_k should not depend on k. Furthermore we consider the possibility that not all points in \mathbb{R}^n are admissible states, and postulate for $f : \mathcal{M} \to \mathcal{M} \subset \mathbb{R}^n$ the evolution rule

(1)
$$u_{k+1} = f(u_k), \qquad k \in \mathbb{Z}$$

One particular forward trajectory is fixed by prescribing an initial state

(2)
$$u_0 = \overline{u} \in \mathcal{M}$$

The choice of k = 0 as initial time is not essential by the independence of f on k. Since the forward trajectory is obviously given by

$$u_k = f^k(\overline{u}) := \underbrace{f \circ \cdots \circ f}_{k \text{ times}} (\overline{u}), \qquad k \ge 0,$$

we talk about *iterated maps* in this situation. If the map f is invertible, the initial state also determines the states at negative times, and the above formula can also be used for k < 0 with the convention $f^{-k} := (f^{-1})^k$, k > 0.

For continuous time we consider *explicit first order autonomous systems of* ordinary differential equations

$$\dot{u}(t) = f(u(t))$$

with $u(t) \in \mathbb{R}^n$ for $t \in \mathbb{R}$, $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\dot{u} := du/dt$. The differential equations are *ordinary*, since the unknown function u only depends on one variable. They are *explicit*, since the derivatives of the components are given as functions of the state. Finally, *autonomous* means that f does not explicitly depend on t, which reflects the time independence of the environment.

Again we expect that prescribing the state at a certain time (w.l.o.g. chosen as t = 0) determines the subsequent evolution. We consider (3) subject to the initial condition

$$(4) u(0) = u_0$$

The *Picard-Lindelöf Theorem* shows that from t = 0 we can actually go forward and backward in time, at least a little:

Theorem 1. Let $u_0 \in \mathbb{R}^n$ and let f(u) be Lipschitz continuous in a neighborhood U of u_0 with values in \mathbb{R}^n . Then there exists T > 0 and a unique $u \in C^1((-T,T))$ solving (3), (4), for -T < t < T. The existence time T only depends on U, on $\sup_U |f|$, and on the Lipschitz constant of f in U.

 $\mathbf{2}$

The Picard theorem requires Lipschitz continuity of f. In the following it will be convenient to assume even more regularity. In order to avoid technicalities concerning precise smoothness assumptions, we shall assume from now on

(5)
$$f \in C^{\infty}(\mathbb{R}^n)^n$$

for the functions in both (1) and (3). This assumption will be used for the rest of the course, and it will not be repeated in each theorem.

The Picard-Lindelöf theorem is a local existence theorem guaranteeing existence only in a small enough time interval. The example

(6)
$$\dot{u} = u^2, \qquad u(0) = 1,$$

with the explicit solution $u(t) = (1-t)^{-1}$ shows that in general no better result can be expected. We observe that the maximal existence interval $(-\infty, 1)$ is open, and $\lim_{t\to 1^-} |u(t)| = \infty$ holds. The following result shows that 'nothing worse' can happen.

Theorem 2. Let (5) hold and let $u_0 \in \mathbb{R}^n$. Then the maximal existence interval I of the unique solution of (3), (4) is open, i.e. I = (a, b) with $-\infty \leq a < 0 < b \leq \infty$. In the cases $a > -\infty$ or $b < \infty$ we have

 $\lim_{t \to a+} |u(t)| = \infty \qquad or, \ respectively, \qquad \lim_{t \to b-} |u(t)| = \infty \,.$

Remark 1. The Euclidian norm in \mathbb{R}^n is denoted by $|\cdot|$ and the scalar product by a dot, *i.e.* $|u|^2 = u \cdot u$.

Proof: For $I = \mathbb{R}$ there is nothing to prove. Therefore we first assume $b < \infty$. If $\lim_{t \to b^-} |y(t)| = \infty$ does not hold, then there exists a sequence $t_n \to b^-$, such that the sequence $u(t_n)$ is bounded and therefore it contains a convergent subsequence $u(t_{n_k}) \to \overline{u}$ (by the Bolzano-Weierstrass theorem). Theorem 1 implies that for a neighborhood U of \overline{u} there exists T > 0 such that for all $\tilde{u} \in U$ the solution of the initial value problem (3) subject to $u(\tilde{t}) = \tilde{u}$ exists in the interval $(\tilde{t} - T, \tilde{t} + T)$. Since $(t_{n_k}, u(t_{n_k})) \to (b, \overline{u})$, there exists n_k , such that $u(t_{n_k}) \in U$ and $b - t_{n_k} < T$. The solution can therefore be extended up to the time $t_{n_k} + T > b$ in contradiction to b being the right end of the existence interval. It is an obvious consequence that the existence interval is open at b.

The left end is treated analogously. $\hfill\blacksquare$

This result often helps in proving *global existence* of solutions, i.e. existence for all times. A useful auxiliary result is the *Gronwall lemma*:

Lemma 1. a) Let $z : [0,T] \rightarrow [0,\infty)$ be continuous, $\lambda, z_0 \geq 0$, and let

$$z(t) \le z_0 + \lambda \int_0^t z(s) ds$$
, $0 \le t \le T$.

Then $z(t) \leq e^{\lambda t} z_0, \ 0 \leq t \leq T$.

b) Let $z: [0,T] \to [0,\infty)$ be continuously differentiable, $\lambda \in \mathbb{R}$, $z(0) = z_0 \ge 0$,

and let

$$\dot{z}(t) \le \lambda z(t), \qquad 0 \le t \le T.$$

Then $z(t) \leq e^{\lambda t} z_0, \ 0 \leq t \leq T$.

Proof: a) The function

$$v(t) := e^{-\lambda t} \int_0^t z(s) ds$$

satisfies

$$\dot{v}(t) = e^{-\lambda t} \left(z(t) - \lambda \int_0^t z(s) ds \right) \le e^{-\lambda t} z_0 \,.$$

By integration we obtain

$$v(t) \leq \frac{z_0}{\lambda} \left(1 - e^{-\lambda t} \right) \,.$$

Since $z(t) \leq z_0 + \lambda e^{\lambda t} v(t)$, the result follows. Note that $\lambda \geq 0$ is used in this last step.

b) The function $u(t) = e^{-\lambda t} z(t)$ satisfies $\dot{u} \leq 0$ and, thus, $u(t) \leq u(0) = z_0$.

The following theorem is a typical global existence result.

Theorem 3. Let the assumptions of Theorem 2 be satisfied and let the right hand side f have at most linear growth, i.e. there exist $\lambda, \mu \geq 0$ such that $|f(u)| \leq \lambda |u| + \mu$ for all $u \in \mathbb{R}^n$. Then for every $u_0 \in \mathbb{R}^n$ the solution of (3), (4) exists for all times.

For every $t \in \mathbb{R}$, u(t) depends Lipschitz continuously on the initial state u_0 .

Proof: We prove existence for all t > 0. The proof for negative t is analogous after $t \leftrightarrow -t$.

The formulation of the initial value problem as integral equation

$$u(t) = u_0 + \int_0^t f(u(s))ds$$

implies

$$|u(t)| \le |u_0| + \int_0^t (\lambda |u(s)| + \mu) ds$$

For $\lambda = 0$ this gives $|u(t)| \le |u_0| + t\mu$. For $\lambda > 0$ we use the Gronwall lemma with $z(t) = |u(t)| + \mu/\lambda$ and obtain

$$|u(t)| \le e^{\lambda t} |u_0| + \frac{\mu}{\lambda} \left(e^{\lambda t} - 1 \right)$$
.

In both cases |u(t)| cannot grow above all bounds in finite time. Thus the solution is global by Theorem 2.

For proving Lipschitz continuous dependence on the initial state, fix $u_0 \in \mathbb{R}^n$ and $t \in \mathbb{R}$. Then the estimates above show that for initial states in a bounded neighborhood U of u_0 , the trajectories between times 0 and t lie in a bounded

4

set. Denote the Lipschitz constant of f in this set by L, choose $v_0 \in U$ and let u(t), v(t) be the solutions of (3) with $u(0) = u_0, v(0) = v_0$. Then we have

$$\begin{aligned} |u(t) - v(t)| &\leq |u_0 - v_0| + \int_0^t |f(u(s)) - f(v(s))| ds \\ &\leq |u_0 - v_0| + L \int_0^t |u(s) - v(s)| ds \,, \end{aligned}$$

and, thus, by the Gronwall lemma,

$$|u(t) - v(t)| \le e^{Lt} |u_0 - v_0|.$$

Definition 1. Let \mathcal{M} be a metric space (the state space or phase space) and let the set of <u>times</u> \mathcal{T} be either \mathbb{R} , $[0, \infty)$, \mathbb{Z} , or \mathbb{N}_0 . A <u>deterministic dynamical</u> system is a map $\mathcal{T} \times \mathcal{M} \to \mathcal{M}$, $(t, u_0) \mapsto S_t(u_0)$, satisfying

- (1) $\forall u_0 \in \mathcal{M}$: $S_0(u_0) = u_0$,
- (2) $\forall u_0 \in \mathcal{M}, s, t \in \mathcal{T}: S_{s+t}(u_0) = S_s(S_t(u_0)),$
- (3) $\forall t \in \mathcal{T}: u_0 \mapsto S_t(u_0)$ is continuous.

In the cases $\mathcal{T} = [0, \infty)$ and $\mathcal{T} = \mathbb{N}_0$, S_t is called a <u>forward dynamical system</u>; for $\mathcal{T} = \mathbb{Z}$ or $\mathcal{T} = \mathbb{N}_0$ it is called a <u>discrete dynamical system</u>; and for $\mathcal{T} = \mathbb{R}$ or $\mathcal{T} = [0, \infty)$ it is called a <u>continuous dynamical system</u>. For fixed $u_0 \in \mathcal{M}$, the set $\{S_t(u_0) : t \in \mathcal{T}\}$ is called the <u>trajectory</u> through u_0 . The collection of all trajectories is called the phase portrait of the dynamical system.

Remark 2. Condition (1) just means that u_0 is the initial state. Condition (2) is called the semigroup property, since it induces a semigroup structure for forward dynamical systems. For forward-and-backward dynamical systems ($\mathcal{T} = \mathbb{R}$ or $\mathcal{T} = \mathbb{Z}$) it is actually a group structure with the inverse of S_t given by S_{-t} .

Finally, condition (3) means continuous dependence on the initial state. For continuous dynamical systems one typically also expects continuity with respect to time.

Remark 3. Under the assumptions of Theorem 3, (3) defines a continuous dynamical system on \mathbb{R}^n . Even if these assumptions are violated, a dynamical system might result from a reduction of the state space, e.g. the ODE in (6) defines a forward dynamical system on $\mathcal{M} = (-\infty, 0]$, but not on $\mathcal{M} = \mathbb{R}$.

The iteration (1) defines a discrete dynamical system on \mathbb{R}^n , whenever f is continuous. Discrete dynamical systems also result from the explicit Euler discretization

$$u_{k+1} = u_k + \Delta t f(u_k)$$

of (3) with time step Δt .

In this course we deal with both discrete and continuous dynamical systems on finite dimensional state spaces. The solution operators of partial differential equations and delay differential equations are typical examples for dynamical systems on infinite dimensional state spaces.

Remark 4. For continuous dynamical systems defined by ODEs, trajectories are either smooth curves or individual (stationary) points. By uniqueness of the solutions of initial value problems there is exactly one trajectory through each point. Thus, the phase portrait provides a simple covering of the phase space. Knowing this, the possible qualitative behaviors of trajectories are restricted, mainly by the dimension of the phase space. An application of these observations is the Poincaré-Bendixson theorem (Section 8.5).

Dynamical systems theory (and therefore also this course) is mostly concerned with the investigation of the long-time behavior of trajectories and how it changes with varying initial state and in dependence of parameters. In this context, a basic object of study are steady states and their stability.

Definition 2. Let S_t , $t \in \mathcal{T}$, be a dynamical system on (\mathcal{M}, d) . Every $\overline{u} \in \mathcal{M}$ satisfying $S_t(\overline{u}) = \overline{u}$ for all $t \in \mathcal{T}$ is called a <u>stationary point</u> or <u>steady state</u>. A steady state is called <u>stable</u>, if

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad d(u_0, \overline{u}) < \delta \implies \quad d(S_t(u_0), \overline{u}) < \varepsilon \quad \forall t > 0 \,.$$

In words: Trajectories stay arbitrarily close to \overline{u} , if they start close enough to it. If \overline{u} is not stable, it is called <u>unstable</u>.

A stable steady state \overline{u} is called (locally) asymptotically stable, if

$$\exists \delta > 0: \quad d(u_0, \overline{u}) < \delta \implies \lim_{t \to \infty} S_t(u_0) = \overline{u}.$$

In words: Trajectories converge to \overline{u} , if they start close enough to it.

An asymptotically stable steady state \overline{u} is called <u>globally asymptotically stable</u>, if

$$\forall u_0 \in \mathcal{M}: \quad \lim_{t \to \infty} S_t(u_0) = \overline{u}.$$

In words: All trajectories converge to \overline{u} .

Remark 5. The steady states \overline{u} of recursions $u_{k+1} = f(u_k)$ are the fixed points of f, i.e. $\overline{u} = f(\overline{u})$. The steady states \overline{u} of ODEs $\dot{u} = f(u)$ are the zeroes of f, i.e. $f(\overline{u}) = 0$. Their stability properties are not seen quite as easily.

Definition 3. Let S_t , $t \in \mathcal{T}$, be a dynamical system on (\mathcal{M}, d) . A set $A \subset \mathcal{M}$ is called positively invariant, if

$$\{S_t(u_0): u_0 \in A, t \in \mathcal{T} \cap (0, \infty)\} \subset A$$

Definition 4. Let S_t , $t \in \mathcal{T}$, be a forward dynamical system on (\mathcal{M}, d) and let $u_0 \in \mathcal{M}$. The <u>omega limit</u> $\omega(u_0)$ of u_0 is the set of all $u \in \mathcal{M}$ such that there exists a sequence $\{t_n\}_{n\in\mathbb{N}} \subset \mathcal{T}$ with $\lim_{n\to\infty} t_n = \infty$, such that $S_{t_n}(u_0) = u$.

b) Let $\mathcal{M} = \mathbb{R}^n$ and let $\{S_t(u_0) : t \in \mathcal{T} \cap (0, \infty)\}$ be bounded. Then $\omega(u_0)$ is nonempty and connected.

Proof: Proseminar.

2. Linear systems

A special case of (1) is a linear homogeneous recursion

$$(7) u_{k+1} = A u_k$$

with a quadratic matrix $A \in \mathbb{R}^{n \times n}$. In this case the solution of the initial value problem is given by $u_k = A^k u_0, k \ge 0$. With the help of a little linear algebra, this can be made more explicit. Particularly simple is the case of a *diagonalizable* matrix A, i.e. when there exists an invertible matrix R and a diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, such that

(8)
$$A = R\Lambda R^{-1}$$

In this case $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A and the columns of R are corresponding eigenvectors. It is easily shown that

(9)
$$A^{k} = R \Lambda^{k} R^{-1} = R \operatorname{diag} \left(\lambda_{1}^{k}, \dots, \lambda_{n}^{k}\right) R^{-1}$$

holds. This implies that the solution can be written as a linear combination of eigenvectors of A with coefficients $\lambda_1^k, \ldots, \lambda_n^k$. If for example $|\lambda_j| < 1$, $j = 1, \ldots, n$, then the solution converges to zero as $k \to \infty$ for arbitrary u_0 . An alternative way to obtain the result is by diagonalizing the recursion. With the transformation $u_k = Rv_k$, i.e. representing u_k in terms of the basis given by the eigenvectors, we obtain the equivalent formulation

$$v_{k+1} = \Lambda v_k$$
, i.e. $v_{k+1,j} = \lambda_j v_{k,j}$, $j = 1, ..., n$,

a decoupled system of scalar recursions with the obvious solution

$$v_{k,j} = \lambda_j^k v_{0,j}, \qquad k \ge 0, \quad j = 1, \dots, n.$$

The diagonalized form also shows that for $|\lambda_j| < 1$, j = 1, ..., n, $\overline{u} = 0$ is the only steady state, which is globally asymptotically stable.

A decomposition of the form (8) always exists, but in general the matrix Λ is not diagonal, but contains Jordan blocks. The eigenvalues of A are still important for the long-time behavior of solutions. We state the corresponding result without proof.

Theorem 5. Let $|\lambda| < 1$ for all eigenvalues λ of A. Then for every initial state u_0 the solution $u_k = A^k u_0$ of (7) converges to zero as $k \to \infty$. If $|\lambda| > 1$ for at least one eigenvalue λ of A, then there exists $u_0 \in \mathbb{R}^n$ such that $u_k = A^k u_0$ satisfies $\lim_{k\to\infty} |u_k| = \infty$.

Remark 6. The theorem does not cover the case, where the spectral radius of A is equal to 1. In this case no general statement is possible. The behavior is determined by the Jordan block structure of possible multiple eigenvalues with modulus 1.

Now we turn to the continuous case and consider a linear homogeneous version of (3):

$$\dot{u} = Au$$

again with a quadratic matrix $A \in \mathbb{R}^{n \times n}$. In this case the solution of the initial value problem is given by $u(t) = e^{At}u_0$, where the matrix exponential is defined by the power series

$$e^{At} := \sum_{j=0}^{\infty} \frac{(At)^j}{j!} \,,$$

whose convergence can be proven analogously to the convergence of the power series for the scalar exponential function. Also the proof of the semigroup property

$$e^{A(t+s)} = e^{At} e^{As}, \qquad \forall s, t \in \mathbb{R},$$

is analogous to the case n = 1. The validity of the differential equation can be shown by term-by-term differentiation of the power series. For a diagonizable matrix A, the matrix exponential can be computed explicitly with the help of (9):

$$e^{At} = R e^{\Lambda t} R^{-1} = R \operatorname{diag}\left(e^{\lambda_1 t}, \dots, e^{\lambda_n t}\right) R^{-1}$$

Again the ODE system can be decoupled by the transformation u(t) = Rv(t). We state a result on the long time behavior of trajectories also for possibly nondiagonizable matrices:

Lemma 2. Let $\operatorname{Re}(\lambda) < 0$ for all eigenvalues λ of A. Then there exists $\overline{\lambda} < 0$, such that for every initial state u_0 the solution $u(t) = e^{At}u_0$ of (10) satisfies $|u(t)| \leq e^{\overline{\lambda}t}|u_0|, t \geq 0$. If $\operatorname{Re}(\lambda) > 0$ for at least one eigenvalue λ of A, then there exists $u_0 \in \mathbb{R}^n$ such that $u(t) = e^{At}u_0$ satisfies $\lim_{t\to\infty} |u(t)| = \infty$.

Remark 7. As in the previous theorem not all cases are covered. If A has eigenvalues with non-positive real parts, then the Jordan block structure of multiple imaginary eigenvalues will be important for the stability properties of the steady state zero.

From our computations above it is easily seen that for diagonalizable matrices $A, \overline{\lambda}$ can be chosen as the maximum of the real parts of the eigenvalues of A. In the general case any value strictly bigger can be used.

Finally, let us consider the case n = 2, i.e.,

(11)
$$\dot{v}_1 = \lambda_1 v_1, \qquad \dot{v}_2 = \lambda_2 v_2,$$

with the assumption $\lambda_1 < 0 < \lambda_2$ on the eigenvalues. The positive and negative parts of the coordinate axes are trajectories, where the v_1 -axis is called the *stable* manifold, since it contains all initial values such that the solution converges to zero as $t \to \infty$. Similarly, the v_2 -axis is called the unstable manifold, since it contains all initial values such that the solution converges to zero as $t \to -\infty$. All other trajectories lie on curves with the equation

$$v_1|^{\lambda_2}|v_2|^{-\lambda_1} = c, \qquad c > 0.$$

This can be seen by either differentiating this equation or by using the explicit solutions of (11). The trajectories have the qualitative behavior of hyperbolas filling the (v_1, v_2) -plane. As $t \to \infty$ they approach the unstable manifold, and as $t \to -\infty$ the stable manifold. This picture is qualitatively the same in the original (u_1, u_2) -plane. However, the stable und unstable manifolds are now spanned by the eigenvectors of A.

2.1. Inhomogeneous linear ODE systems. For later reference we provide some results for inhomogeneous linear systems. Note that we permit time dependent inhomogeneities, i.e. non-autonomous equations. For a system of the form

(12)
$$\dot{u} = Au + h(t),$$

with a constant matrix $A \in \mathbb{R}^{n \times n}$ and a given inhomogeneity $h(t) \in \mathbb{R}^n$, particular solutions are given by the *variation of constants* formula

$$u(t) = \int_{t_0}^t e^{A(t-s)} h(s) ds \,,$$

where t_0 can be chosen arbitrarily. In particular, by the superposition principle, the solution of the initial value problem with $u(0) = u_0$ is given by

(13)
$$u(t) = e^{At}u_0 + \int_0^t e^{A(t-s)}h(s)ds$$

We now consider the situation, where h(t) is bounded in $[0, \infty)$, and look for bounded solutions of (12) in two different cases.

Lemma 3. Let $h: [0, \infty) \to \mathbb{R}^n$ be continuous and bounded.

a) Let all eigenvalues of A have negative real parts. Then all solutions of (12) can be written in the form (13) and are bounded on $[0, \infty)$.

b) Let all eigenvalues of A have positive real parts. Then there is exactly one bounded solution of (12), given by

(14)
$$u(t) = -\int_t^\infty e^{A(t-s)}h(s)ds$$

Proof: a) Clearly the set of all solutions can be parametrized by its state at t = 0 and, thus, all solutions are of the form (13). With $\overline{\lambda} < 0$ from Lemma 2 we

have

$$\begin{aligned} |u(t)| &\leq e^{\overline{\lambda}t} |u_0| + \int_0^t e^{\overline{\lambda}(t-s)} |h(s)| ds \leq |u_0| + \sup_{[0,\infty)} |h| \int_0^t e^{\overline{\lambda}(t-s)} ds \\ &\leq |u_0| + \frac{1}{|\overline{\lambda}|} \sup_{[0,\infty)} |h| \,. \end{aligned}$$

b) Since now the eigenvalues of -A have negative real parts, Lemma 2 can be applied to this matrix with a corresponding $\overline{\lambda} < 0$. This implies that for every $u_0 \neq 0$, the solution $u_h(t) = e^{At}u_0$ of the initial value problem for the homogeneous equation cannot be bounded, since otherwise the estimate

$$|u_0| = |e^{-At}u_h(t)| \le e^{\overline{\lambda}t} \sup_{[0,\infty)} |u_h|$$

would lead to a contradiction. Therefore it suffices to show that the solution given by (14) is bounded, since any other solution is obtained by adding an unbounded term of the form $e^{At}u_0$.

$$u(t)| \leq \int_t^\infty e^{\overline{\lambda}(s-t)} |h(s)| ds \leq \frac{1}{|\overline{\lambda}|} \sup_{[0,\infty)} |h| \,.$$

3. Scalar ODEs – stability

An ODE of the form $\dot{u} = f(u)$ with $f : \mathbb{R} \to \mathbb{R}$ can in principle be solved: The solution subject to the initial condition $u(0) = u_0$ is implicitly given by

$$\int_{u_0}^{u(t)} \frac{d\eta}{f(\eta)} = t$$

Typically, the qualitative behavior of solutions is not obvious from this formula. On the other hand, it can easily be seen directly from f. Because of its continuity the state space (i.e. the *u*-axis) is the union of the zeroes of f on the one hand, and open intervals where f is either positive or negative, on the other hand. Zeroes u_0 of f are steady states, since the solution of the initial value problem with $u(0) = u_0$ is the constant $u(t) = u_0$, $t \in \mathbb{R}$.

Suppose on the other hand that f is positive between two zeroes u_1 , u_2 and $u_1 < u_0 < u_2$. Then the solution starting at u_0 exists for all time and satisfies

$$\lim_{t \to -\infty} u(t) = u_1, \qquad \lim_{t \to \infty} u(t) = u_2.$$

As a third case let f be positive everywhere to the right of the zero u_1 , then for $u_0 > u_1$,

$$\lim_{t \to -\infty} u(t) = u_1, \qquad \lim_{t \to T} u(t) = \infty,$$

10

where $T \leq \infty$ is the right end of the existence interval. For all other possible cases similar statements hold. In particular, trajectories either converge as $t \to \pm \infty$ or they tend to ∞ or $-\infty$. The proofs are easy and left to the reader.

The stability of steady states is easily seen. If a steady state \overline{u} is isolated, and the sign of f changes at \overline{u} from positive to negative, then \overline{u} is asymptotically stable. Isolated steady states with any other behavior of f in the neighborhood are unstable. If a steady state lies in the interior of an interval, where f vanishes, then the steady state is stable, but not asymptotically stable.

Simple examples:

- $\dot{u} = 0$: Every $\overline{u} \in \mathbb{R}$ is a stable, but not asymptotically stable, steady state.
- $\dot{u} = -u$: $\overline{u} = 0$ is a globally asymptotically stable steady state.
- $\dot{u} = u, \, \dot{u} = \pm u^2 : \overline{u} = 0$ is an unstable steady state.
- $\dot{u} = u^3 u : \overline{u} = 0$ is a locally asymptotically stable steady state. $\overline{u} = \pm 1$ are unstable steady states.

4. Hyperbolic stationary points – linearization

Although usually not really necessary for scalar equations, it is a reasonable idea to study local stability properties by using local approximations of f, i.e. Taylor polynomials. If in the ODE the function f is approximated by its first order Taylor polynomial around a steady state \overline{u} ,

$$f(u) \approx f(\overline{u}) + f'(\overline{u})(u - \overline{u}) = f'(\overline{u})(u - \overline{u}),$$

the resulting linear ODE for $v \approx u - \overline{u}$,

(15)
$$\dot{v} = f'(\overline{u})v,$$

is called the *linearization* of (3) at \overline{u} . Obviously, the steady state v = 0 of (15) is asymptotically stable for $f'(\overline{u}) < 0$, stable for $f'(\overline{u}) = 0$, and unstable for $f'(\overline{u}) > 0$. Consequences for the underlying nonlinear ODE are easily seen:

Theorem 6. Let n = 1 and let \overline{u} be a steady state of (3). If $f'(\overline{u}) < 0$, then \overline{u} is locally asymptotically stable. If $f'(\overline{u}) > 0$, then \overline{u} is unstable.

The simple proof is left to the reader. There is no conclusion for $f'(\overline{u}) = 0$, because in this case the local behavior of f around \overline{u} , and therefore also its stability properties, depend on higher order terms in the Taylor expansion. The examples $\dot{u} = 0$, $\dot{u} = \pm u^3$ share the linearization at $\overline{u} = 0$ with $f'(\overline{u}) = 0$, but not the stability properties.

The linearization approach can also be used for systems. Then the linearized equation reads $\dot{v} = Df(\overline{u})v$ (the generalization of (15)) with the Jacobian matrix $Df(\overline{u})$. We generalize the assumptions of Theorem 6 to higher dimensions.

Definition 5. Let $n \ge 1$ and let \overline{u} be a steady state of (3). Then \overline{u} is called hyperbolic, if $\operatorname{Re}(\lambda) \ne 0$ for all eigenvalues λ of the Jacobian $Df(\overline{u})$.

The term hyperbolic can be motivated by the two-dimensional linear example in Section 2. We shall use the fact that for hyperbolic steady states the Jacobian can be block diagonalized, i.e.

(16)
$$Df(\overline{u}) = R\Lambda R^{-1}$$
, with $\Lambda = \begin{pmatrix} \Lambda_{-} & 0\\ 0 & \Lambda_{+} \end{pmatrix}$, $R = (R_{-}, R_{+})$,

where $\Lambda_{-} \in \mathbb{R}^{k \times k}$, $0 \le k \le n$, has only eigenvalues with negative real parts, and $\Lambda_{+} \in \mathbb{R}^{(n-k) \times (n-k)}$ has only eigenvalues with positive real parts. The columns of $R_{-} \in \mathbb{R}^{n \times k}$ are generalized eigenvectors corresponding to the eigenvalues with negative real parts, and the columns of $R_{+} \in \mathbb{R}^{n \times (n-k)}$ are generalized eigenvectors corresponding to the eigenvalues with positive real parts.

Theorem 7. (Stable manifold theorem) Let $\overline{u} \in \mathbb{R}^n$ be an hyperbolic steady state of the dynamical system S_t generated by (3). Then there is a neighborhood $U \subset \mathbb{R}^n$ of \overline{u} , such that

$$\mathcal{M}_s[\overline{u}] := \{ u_0 \in U : S_t(u_0) \in U, t \ge 0 \}$$

is a k-dimensional (referring to the diagonalization (16)) manifold in \mathbb{R}^n , called the stable manifold of \overline{u} . For all $u_0 \in \mathcal{M}_s[\overline{u}]$,

(17)
$$|S_t(u_0) - u_0| \le c e^{\lambda_- t}, \quad t \ge 0,$$

where $c \geq 0$ and $\lambda_{-} < 0$ is the constant $\overline{\lambda}$ for the matrix Λ_{-} from Lemma 2. The tangent space of $\mathcal{M}_{s}[\overline{u}]$ at \overline{u} is spanned by the columns of R_{-} from (16), i.e. by the eigenvectors corresponding to eigenvalues of $Df(\overline{u})$ with negative real parts.

Analogously, the set

$$\mathcal{M}_u[\overline{u}] := \{ u_0 \in U : S_t(u_0) \in U, t \le 0 \}$$

is a (n-k)-dimensional manifold in \mathbb{R}^n , called the <u>unstable manifold</u> of \overline{u} . For all $u_0 \in \mathcal{M}_s[\overline{u}]$,

$$|S_t(u_0) - u_0| \le c e^{\lambda_+ t}, \qquad t \le 0,$$

where $c \geq 0$ and $-\lambda_+ < 0$ is the constant $\overline{\lambda}$ for the matrix $-\Lambda_+$. The tangent space of $\mathcal{M}_u[\overline{u}]$ at \overline{u} is spanned by the columns of R_+ .

Remark 8. Obviously, hyperbolic steady states are either locally asymptotically stable (k = n) or unstable. For non-hyperbolic steady states, the linearization does not contain sufficient information for a complete characterization of the local behaviour. In particular, if all eigenvalues have non-positive real parts with at least one eigenvalue on the imaginary axis, then there is no conclusion concerning the stability properties of the steady state.

Proof: The right hand side of (3) can be written as $f(u) = Df(\overline{u})(u - \overline{u}) + Rr(R^{-1}(u - \overline{u}))$ with the second order remainder term, which we have chosen to write in this form with the matrix R from (16). We decouple increasing and decreasing modes by the transformation $u(t) = \overline{u} + Rv(t)$, and obtain

$$\dot{v} = \Lambda v + r(v) \,.$$

With the notation $v = (v_-, v_+)$, $r = (r_-, r_+)$, according to the block structure of Λ , this can be written as

$$\dot{v}_{-} = \Lambda_{-}v_{-} + r_{-}(v), \qquad \dot{v}_{+} = \Lambda_{+}v_{+} + r_{+}(v).$$

Let u_0 be in the set $\mathcal{M}_s[\overline{u}]$. Then $S_t(u_0)$, and therefore also $v(t) = R^{-1}(S_t(u_0) - \overline{u})$, and consequentially r(v(t)) are bounded for $t \ge 0$. Thus we can use the results of Lemma 3 to obtain

(18)
$$v_{-}(t) = e^{\Lambda_{-}t}p + \int_{0}^{t} e^{\Lambda_{-}(t-s)}r_{-}(v(s))ds,$$

(19)
$$v_{+}(t) = -\int_{t}^{\infty} e^{\Lambda_{+}(t-s)} r_{+}(v(s)) ds$$

for some $p \in \mathbb{R}^k$. In the following we shall prove that for given small enough p the integral equation problem (18), (19) has a unique solution. Therefore $v_+(0)$ is determined as a function of $p = v_-(0)$, and the stable manifold in the *v*-space is thus given as the graph of a function from \mathbb{R}^k to \mathbb{R}^{n-k} .

In order to prove the decay estimate (17) at the same time, we set $v(t) = e^{\lambda - t}w(t)$ in (18), (19):

(20)
$$w_{-}(t) = e^{(\Lambda_{-}-\lambda_{-})t}p + e^{-\lambda_{-}t} \int_{0}^{t} e^{\Lambda_{-}(t-s)}r_{-}(e^{\lambda_{-}s}w(s))ds$$

(21)
$$w_{+}(t) = -e^{-\lambda_{-}t} \int_{t}^{\infty} e^{\Lambda_{+}(t-s)} r_{+}(e^{\lambda_{-}s}w(s)) ds.$$

We shall use the Banach fixed point theorem for w in the space $C_B([0,\infty))^n$ (bounded continuous functions), noting that (20), (21) has the fixed point form w = F(w). Actually we shall restrict to a ball

$$\mathcal{B}_{\delta} := \{ w \in C_B([0,\infty))^n : \|w\|_{\infty} < \delta \},\$$

where $\|\cdot\|_{\infty}$ denotes the supremum norm on $[0,\infty)$. We shall always assume $(p,0) \in \mathcal{B}_{\delta/2}$.

Our first claim is that

(22)
$$F: \mathcal{B}_{\delta} \to \mathcal{B}_{\delta}$$
 for δ small enough.

Since by Lemma 2, $|e^{\Lambda_{-}t}u_0| \leq e^{\lambda_{-}t}|u_0|, t \geq 0$, and $|e^{\Lambda_{+}t}u_0| \leq e^{\lambda_{+}t}|u_0|, t \leq 0$, we have for $w \in \mathcal{B}_{\delta}$,

$$\begin{split} |F(w)_{-}(t)| &\leq |v_{-}(0)| + \int_{0}^{t} e^{-\lambda_{-}s} |r_{-}(e^{\lambda_{-}s}w(s))| ds \leq \frac{\delta}{2} + \delta^{2}L \int_{0}^{t} e^{\lambda_{-}s} ds \\ &\leq \frac{\delta}{2} + \frac{\delta^{2}L}{|\lambda_{-}|} \,, \\ |F(w)_{+}(t)| &\leq e^{(\lambda_{+}-\lambda_{-})t} \int_{t}^{\infty} e^{-\lambda_{+}s} |r_{+}(e^{\lambda_{-}s}w(s))| ds \\ &\leq \delta^{2}L \, e^{(\lambda_{+}-\lambda_{-})t} \int_{t}^{\infty} e^{(2\lambda_{-}-\lambda_{+})s} ds \leq \frac{\delta^{2}L}{\lambda_{+}-2\lambda_{-}} \,, \end{split}$$

where we have used the Lipschitz continuity estimate (52) from the Appendix. After summing these inequalities, (22) is obvious. It remains to prove that, again for δ small enough, F is a contraction: For $w_1, w_2 \in \mathcal{B}_{\delta}$,

$$\begin{aligned} |F(w_{1})_{-}(t) - F(w_{2})_{-}(t)| &\leq \int_{0}^{t} e^{-\lambda_{-}s} |r_{-}(e^{\lambda_{-}s}w_{1}(s)) - r_{-}(e^{\lambda_{-}s}w_{2}(s))| ds \\ &\leq \delta L \int_{0}^{t} e^{\lambda_{-}s} |w_{1}(s) - w_{2}(s)| ds \leq \frac{\delta L}{|\lambda_{-}|} ||w_{1} - w_{2}||_{\infty} \,, \\ |F(w_{1})_{+}(t) - F(w_{2})_{+}(t)| &\leq \delta L ||w_{1} - w_{2}||_{\infty} e^{(\lambda_{+} - \lambda_{-})t} \int_{t}^{\infty} e^{(2\lambda_{-} - \lambda_{+})s} ds \\ &\leq \frac{\delta L}{|\lambda_{+} - 2\lambda_{-}|} ||w_{1} - w_{2}||_{\infty} \,. \end{aligned}$$

Again, summing the inequalities immediately implies the contraction property for δ small enough. This implies that for each small enough p there exists a unique (in a small ball) solution of (18), (19) satisfying (17). Thus, $u_0(p) = \overline{u} + Rv(0; p)$ is a parametrization of $\mathcal{M}_s[\overline{u}]$ with parameter $p \in \mathbb{R}^k$. An extension of the Banach fixed point theorem for problems with parameters implies that the solution depends smoothly on p.

It remains to determine the tangent space at $u_0 = \overline{u}$, which obviously is obtained with p = 0, whence (18), (19) has the solution v = 0. The function $z_j(t) = \partial_{p_j} v(t) |_{p=0}$ satisfies

$$(z_j)_{-}(t) = e^{\Lambda_{-}t}e_j, \qquad (z_j)_{+}(t) = 0,$$

with the *j*th canonical basis vector $e_j \in \mathbb{R}^k$, since the derivative of the second order remainder at zero vanishes. Thus, $\partial_{p_j} u_0 \mid_{p=0} = R_j$, the *j*th column of R, $1 \leq j \leq k$. This shows that the tangent space has maximal dimension k, and $\mathcal{M}_s[\overline{u}]$ is really a *k*-dimensional manifold.

After time reversal, the proof for the unstable manifold is the same.

The stable manifold theorem tells us that essential properties of the dynamics near hyperbolic fixed points are shared by the full nonlinear system and its linearization. An even stronger result in this direction, which we state without proof (which can be found in e.g. [5]), is the *Hartman-Grobman theorem*. It says that close to an hyperbolic fixed point the dynamics of the nonlinear system and of the linearization are the same up to a diffeomorphism (differentiable and one-to-one with differentiable inverse):

Theorem 8. (Hartman-Grobman) Let $\overline{u} \in \mathbb{R}^n$ be an hyperbolic steady state of (3). Then there is a diffeomorphism $\varphi : U \to V$ between a neighborhood $U \subset \mathbb{R}^n$ of \overline{u} and a neighborhood $V \subset \mathbb{R}^n$ of the origin, such that $D\varphi(u)f(u) = \Lambda\varphi(u)$ (with Λ from (16)), i.e. if $u(t) \in U$ solves (3), then $v(t) := \varphi(u(t))$ solves

$$\dot{v} = D\varphi(u)\dot{u} = D\varphi(u)f(u) = \Lambda v$$
.

We shall be interested in dynamical systems containing *parameters*. For ordinary differential equations, this means to consider systems of the form

(23)
$$\dot{u} = f(u, r)$$
 with $r \in \mathbb{R}^l$

and to study the dependence of the long-time behavior on the parameters. In the following, we shall always assume smoothness of f, not only with respect to the state u, but also with respect to the parameters r:

(24)
$$f \in C^{\infty}(\mathbb{R}^n \times \mathbb{R}^l)^n$$

The parameter dependence motivates a second stability concept: A property of (23) is called <u>structurally stable</u> (or <u>generic</u>), if it is preserved under small parameter changes.

Theorem 9. The existence of an hyperbolic steady state is structurally stable: Let, for $r = r_0 \in \mathbb{R}^l$, (23) have an hyperbolic steady state $u_0 \in \mathbb{R}^n$. Then there exists a neighborhood $R \subset \mathbb{R}^l$ of r_0 , such that (23) has an hyperbolic steady state $\overline{u}(r)$ for every $r \in R$ with $\overline{u}(r_0) = u_0$ and $\overline{u} \in C^{\infty}(R)^n$. The dimensions of the stable and unstable manifolds through $\overline{u}(r)$ do not depend on $r \in R$.

Proof: (outline) Since the hyperbolicity of u_0 implies that $D_u f(u_0, r_0)$ is invertible, the existence and smoothness of \overline{u} are a consequence of the implicit function theorem. Since zeroes of polynomials depend continuously on parameters [1], the signs of the real parts of the eigenvalues of $D_u f(\overline{u}(r), r)$ do not change close to $r = r_0$.

5. Scalar ODEs – bifurcations

As we have seen in Section 3 one dimensional dynamics seems rather boring. Trajectories either converge to steady states or they take off towards $\pm \infty$. In this section we investigate how the long-time behavior might change with a varying parameter. Therefore we consider (23) with n = l = 1. Since hyperbolic steady states are structurally stable, qualitative changes in the dynamic behavior (called *bifurcations*) require the occurrence of a non-hyperbolic steady state. In the following we assume that for the critical parameter values r = 0 the origin u = 0 is a non-hyperbolic steady state, i.e.

(25)
$$f(0,0) = \partial_u f(0,0) = 0.$$

5.1. The fold. With the assumption (25), the Taylor expansion of f around u = r = 0 has the form

(26)
$$f(u,r) = a_{01}r + a_{20}u^2 + a_{11}ru + a_{02}r^2 + O(u^3 + r^3).$$

A simple example is

$$\dot{u} = r + u^2.$$

The bifurcation occurring at r = 0 can be described as follows: For r < 0 there are two hyperbolic steady states, the unstable point $u = \sqrt{-r}$ and the

asymptotically stable point $u = -\sqrt{-r}$. They merge at r = 0, and for positive r there is no stationary point.

In the literature this bifurcation is called the *fold* or the *saddle-node bifuraction*. It is not as special as it seems. Consider the general case (26) in the generic situation, where the first two coefficients a_{01} and a_{20} are different from zero. We claim that there is a transformation taking (23) with (26) to (27). Note the similarity to the Hartman-Grobman Theorem 8.

We shall not provide a full proof, but some formal arguments for this result. As a first step, we replace u by $\frac{u}{a_{20}}$ and r by $\frac{r}{a_{01}a_{20}}$. This transforms (26) to

$$\dot{u} = r + u^2 + a_{11}ru + a_{02}r^2 + O(u^3 + r^3),$$

after renaming coefficients. A heuristic argument is that the terms $a_{11}ru$, $a_{02}r^2$, and $O(r^3)$ are small compared to r and that $O(u^3)$ is small compared to u^2 , and that all these terms can therefore be neglected. We shall show how the two quadratic terms can be eliminated by the close-to-identity transformation

(28)
$$r = R + bR^2, \quad u = U + cU^2$$

It requires some computation to obtain the transformed equation

$$\dot{U} = R + U^2 + (a_{11} - 2c)RU + (a_{02} + b)R^2 + O(U^3 + R^3).$$

The choice $b = -a_{02}$, $c = a_{11}/2$ produces (27) up to a third order remainder. By replacing the quadratic polynomials on the right hand sides of (28) by complete Taylor expansions, the form (27) can be produced exactly (see, e.g., [2]).

This means that a fold occurs, whenever we have the form (26) with a_{01} , $a_{20} \neq 0$. Equation (27) is called the *normal form* of the fold. In the following we shall present the normal forms of other bifurcations without discussing the transformation to the normal form each time.

5.2. The transcritical bifurcation. The fold is *the* generic bifurcation in onedimensional dynamical systems. Other types of bifurcations occur in systems with special properties, which do not change with parameter variations. A typical property of this kind is an always existing special steady state, w.l.o.g. u = 0. In this case, the coefficients a_{01} and a_{02} in (26) vanish. Assuming apart from that a generic situation, means $a_{20}, a_{11} \neq 0$. A corresponding normal form is

$$\dot{u} = ru - u^2 \,.$$

This defines the *transcritical bifurcation* with the following properties: For all values of r we have the steady states u = 0 and u = r. For r < 0, u = 0 is asymptotically stable and u = r is unstable, and vice versa for r > 0. At the bifurcation an *exchange of stability* takes place.

5.3. The pitchfork bifurcation. Sometimes symmetries are present in dynamical systems and invariant under parameter changes. A simple example is a reflection symmetry, where the system does not change, when replacing u by -u.

This leads to the assumption that f is an odd function of u. The consequential normal form is

$$\dot{u} = ru - u^3 \,,$$

exhibiting the *pitchfork bifurcation*: For r < 0 there is one steady state u = 0, which is asymptotically stable. For positive r the stability of u = 0 is lost and transferred to the two new steady states $u = \pm \sqrt{r}$. It is a consequence of the symmetry that for each steady state its reflection also is a steady state with the same stability properties.

5.4. The spruce budworm – the cusp bifurcation. The spruce budworm is a north American tree pest, posing a recurrent threat for forests of conifer trees. Sometimes sudden dramatic increases in the budworm population are observed without big changes in the environment.

We shall describe a budworm population by a continuous dynamical system. Let $N(\tau)$ be a measure for the size of the population at time τ . The equation

$$\frac{dN}{d\tau} = RN\left(1 - \frac{N}{K}\right) - \frac{BN^2}{A^2 + N^2}$$

is a typical model of *population dynamics*. The factor R(1 - N/K) is the difference between the birth rate and the death rate. This is a standard model for competition. The second term on the right hand side describes the loss caused by natural enimies: birds in the case of the budworm, eating the budworm with a maximal rate B. The dependence on the population size N has the following interpretation: If the population is significantly smaller than the threshold A, it does not pay for the birds to look for the budworms, and they mainly look for other kinds of food. Above the critical size A the budworms become attractive as food and are eaten at a rate close to B.

We start by introducing the nondimensional variables

$$t := \frac{\tau}{A/B}$$
, $u(t) := \frac{N(tA/B)}{A}$

The equation for u reads

$$\dot{u} = ru\left(1 - \frac{u}{k}\right) - \frac{u^2}{1 + u^2}$$

with the dimensionless parameters r = RA/B und k = K/A. Note that we have reduced the number of parameters from four to two. This greatly simplifies the analysis of the qualitative properties of the model.

Apart from the trivial steady state u = 0 (always unstable, i.e. the budworms do not die out), there are other steady states satisfying

$$r\left(1-\frac{u}{k}\right) = \frac{u}{1+u^2}\,.$$

Depending on the values of r and k, this equation has 1–3 positive solutions. The regions with different numbers of steady states are separated by folds, occurring

when the derivatives of the left and right hand sides coincide. This requirement leads to the relations

$$r = \frac{2u^3}{(1+u^2)^2}, \qquad k = \frac{2u^3}{u^2 - 1} \qquad \text{with } u > 1,$$

which can be seen as parametrization (with parameter u) of a curve in the (r, k)plane. This curve has a cusp at the point $(r_0, k_0) = (3\sqrt{3}/8, 3\sqrt{3})$ (for $u_0 = \sqrt{3}$). Three positive steady states exist between the two branches of the curve, and one in the rest of the (r, k)-plane. In the latter case, the steady state is asymptotically stable, whereas in the former two of the steady states are asymptotically stable with an unstable steady state in between.

Now the following scenario is possible for the budworm population: Let r be fixed with a value between 1/2 and $3\sqrt{3}/8$, and let k increase slowly (e.g. by the growth of the trees). This gives a straight line in the r(,k)-plane, which twice intersects the bifurcation curve. Before the first intersection there exists a unique stable equilibrium with small values of the budworm population. At the first crossing of the bifurcation curve, a large stable and a middle sized unstable equilibrium are created, but the small equilibrium remains stable and the population remains at this low level. A dramatic change happens, however, at the second crossing of the bifurcation curve. Now the small stable and the medium sized unstable equilibrium disappear, and only the large stable equilibrium is left. A fast growth of the population has to be expected.

This qualitative behavior is already present in small neighborhoods of the cusp point (r_0, k_0) . A normal form of this so called *cusp bifurcation* is given by

$$\dot{u} = r + ku + u^3$$

where (r, k, u) now has to be interpreted as the deviation from (r_0, k_0, u_0) . The cusp bifurcation needs two parameters, whence it is called a *codimension 2 bifurcation*, in contrast to the bifurcations dealt with above, which are of codimension 1.

6. Scalar iterated maps – bifurcations and chaos

Instead of a general discussion, we only treat one (actually the most famous) example, the *logistic map*:

(29)
$$u_{k+1} = ru_k(1 - u_k),$$

with the parameter r > 0. This can be interpreted as a model for population dynamics, if we restrict to values of u_k between 0 and 1, such that u_{k+1} is nonnegative. In order to remain in the interval [0, 1], we also have to assume $r \leq 4$. Thus, for the rest of this section we assume

$$0 < r \le 4$$
, $0 \le u_0 \le 1$,

guaranteeing $0 \le u_k \le 1$ for all $k \ge 0$.

It turns out that the long-time behavior strongly depends on r. The situation is easy for r < 1: Since obviously $u_{k+1} \leq ru_k$, $u_k \leq r^k u_0$ follows by induction. All solutions converge to zero and the population dies out. In terms of the vocabulary of dynamical systems: Zero is the only steady state in the state space $\mathcal{M} = [0, 1]$, and it is globally asymptotically stable.

More generally, the stability of a steady state $\overline{u} = f(\overline{u})$ of the discrete dynamical system

$$(30) u_{k+1} = f(u_k)$$

can be examined by linearization: The recursion

$$v_{k+1} = f'(\overline{u})v_k \,,$$

the linearization of (30) at \overline{u} , can be expected to approximate small values of $v_k = u_k - \overline{u}$.

Theorem 10. Let $f : [a,b] \rightarrow [a,b]$ be twice continuously differentiable. The steady state \overline{u} of (30) is asymptotically stable, if $|f'(\overline{u})| < 1$. In the case $|f'(\overline{u})| > 1$ it is unstable.

Proof: By the Taylor formula, the exact equation for v_k can be written as

(31)
$$v_{k+1} = f(u_k) - f(\overline{u}) = f'(\overline{u})v_k + f''(\tilde{u}_k)v_k^2/2,$$

with $\tilde{u}_k \in [a, b]$, and therefore

$$|v_{k+1}| \le |v_k| \left(|f'(\overline{u})| + |v_k|M/2 \right) ,$$

with $|f''(\tilde{u}_k)| \leq M$. For $|f'(\overline{u})| < 1$ we choose $\delta := (1 - |f'(\overline{u})|)/M$, $r := (1 + |f'(\overline{u})|)/2 < 1$, and $|v_0| \leq \delta$. Induction implies $|v_k| \leq r^k \delta \to 0$, proving the first statement of the theorem.

For $|f'(\overline{u})| > 1$ we start again from (31) and deduce

$$|v_{k+1}| \ge |v_k|(|f'(\overline{u})| - |v_k|M/2)$$

For $|v_k| \leq \varepsilon := (|f'(\overline{u})| - 1)/M$ and $r := (1 + |f'(\overline{u})|)/2 > 1$ we then have

$$|v_{k+1}| \ge r|v_k|,$$

meaning that for arbitrarily small $|v_0|$ we reach $|v_k| > \varepsilon$ in finitely many steps, implying instability of \overline{u} .

Remark 9. In the critical case $|f'(\overline{u})| = 1$ every stability behavior is possible, as can be seen from the examples $u_{k+1} = u_k(1 \pm u_k)$ and $u_{k+1} = u_k$ with $\overline{u} = 0$.

Returning to (29), we see that for increasing values of r the steady state $\overline{u}_1 = 0$ loses its stability at the bifurcation point r = 1. For r > 1 there is a second steady state $\overline{u}_2 = 1 - 1/r$ which, by $f'(\overline{u}_2) = 2 - r$ is asymptotically stable for 1 < r < 3. This is the occurrence of a transcritical bifurcation in a discrete dynamical system.

At the second bifurcation point r = 3 also \overline{u}_2 loses its stability. The behavior of the dynamical system for r > 3 can be understood by analysing $z_k := u_{2k}$, $k \ge 0$, which solves the recursion

(32)
$$z_{k+1} = ru_{2k+1}(1 - u_{2k+1}) = r^2 z_k(1 - z_k)(1 - rz_k(1 - z_k))$$

Besides $\overline{u}_1 = 0$ and $\overline{u}_2 = 1 - 1/r$ this recursion has two more steady states for r > 3:

$$\overline{z}_{3,4} = \frac{1}{2r} \left(1 + r \pm \sqrt{(r+1)(r-3)} \right)$$

It is easily seen that $\overline{z}_3 = f(\overline{z}_4)$ and $\overline{z}_4 = f(\overline{z}_3)$, i.e. the points \overline{z}_3 and \overline{z}_4 constitute a *periodic orbit with period* 2 of the original recursion (29). Note that the periodic orbit is created at the steady state \overline{u}_2 :

$$\overline{u}_2 = \overline{z}_3 = \overline{z}_4 = \frac{2}{3}$$
 for $r = 3$.

Further results, described in the following, are not as easy to verify. It can be shown that \overline{z}_3 and \overline{z}_4 are asymptotically stable steady states of (32) for r > 3close to 3. This implies asymptotic stability of the periodic orbit of (29), where the meaning of this statement should be clear without precise definition. This stability gets lost at the further bifurcation point $r = r_4$. The bifurcation is similar to the one at r = 3: From each of the steady states \overline{z}_3 and \overline{z}_4 of $f \circ f$ bifurcate two new steady states of the four times iterated map $f \circ f \circ f \circ f$, which together form a periodic orbit of period 4 of (29). This is called a *period* doubling bifurcation. For increasing values of r there is a sequence of period doubling bifurcations at the bifurcation points $r_4 < r_8 < r_{16} < \ldots$ This sequence converges to the value $r_c < 4$. Typical trajectories of (29) with $r > r_c$ show appearantly completely irregular behavior. This sensational discovery (of the 1970s) has been termed deterministic chaos. As can be seen from the bifurcation diagram (Fig. 1), we actually have still not told the whole story.

We conclude by considering the special case r = 4, when the recursion can be solved explicitly with the ansatz $u_k = \sin^2 \varphi_k$, leading to $\varphi_{k+1} = 2\varphi_k$ and the explicit solution

$$u_k = \sin^2(2^k \varphi_0)$$
 with $\varphi_0 = \arcsin(\sqrt{u_0})$.

Note that there exist periodic trajectories with arbitrary period p (e.g. for $\varphi_0 = \pi/(2^p - 1)$), but for most initial values the behavior looks completely unpredictable, e.g. whenever $\varphi_0/\pi \notin \mathbb{Q}$.

7. INVARIANT REGIONS - LYAPUNOV FUNCTIONS

Definition 6. A set $M \subset \mathbb{R}^n$ is called positively invariant for (3), if every solution u of (3) with $u(0) \in M$ satisfies $u(t) \in M$ for all $t \ge 0$.

Lemma 4. Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with smooth boundary $\partial\Omega$, let $\nu(u)$, $u \in \partial\Omega$, denote the unit outward normal, and let $\nu(u) \cdot f(u) \leq 0$, $u \in \partial\Omega$. Then the closure $\overline{\Omega}$ is positively invariant for (3).



FIGURE 1. The bifurcation diagram of the logistic map

Proof: First we consider the stronger assumption $\nu(u) \cdot f(u) < 0$, $u \in \partial\Omega$. In this case every trajectory starting on $\partial\Omega$ enters Ω and, consequentially, cannot leave $\overline{\Omega}$. Since $\overline{\Omega}$ is bounded, this also implies existence of trajectories for all $t \geq 0$ by Theorem 2.

Now we return to the assumptions of the Theorem and define

$$f_{\varepsilon}(u) = f(u) - \varepsilon \nu(u), \qquad \varepsilon > 0,$$

which satisfies $\nu \cdot f_{\varepsilon} \leq -\varepsilon < 0$. Therefore, the solution u_{ε} of the initial value problem $\dot{u}_{\varepsilon} = f_{\varepsilon}(u_{\varepsilon}), u_{\varepsilon}(0) = u_0 \in \overline{\Omega}$, remains in $\overline{\Omega}$ for all times and, in particular, for an arbitrary $T > 0, u_{\varepsilon}(T) \in \overline{\Omega}$ holds. Thus $u_{\varepsilon} : [0,T] \to \mathbb{R}^n$ is bounded uniformly in ε as $\varepsilon \to 0$. By the differential equation, the same is true for \dot{u}_{ε} . As a consequence of the Arzela-Ascoli Theorem, there exists a sequence $\varepsilon_n \to 0$ such that $u_{\varepsilon_n} \to u$ uniformly on [0,T]. Therefore we can pass to the limit $\varepsilon \to 0$ in the integrated version

$$u_{\varepsilon}(t) = u_0 + \int_0^t (f(u_{\varepsilon}(s)) - \varepsilon \nu(u_{\varepsilon}(s))) ds$$

of the problem for u_{ε} with the result

$$u(t) = u_0 + \int_0^t f(u(s)) ds$$
.

Since this is equivalent to the problem $\dot{u} = f(u)$, $u(0) = u_0$, and the uniform convergence $u_{\varepsilon} \to u$ implies $u(T) \in \overline{\Omega}$, the proof is complete.

In the following we use the notation $B_r(u_0) = \{u \in \mathbb{R}^n : |u - u_0| < r\}$ for open balls in \mathbb{R}^n .

Definition 7. a) Let $u_0 \in \mathbb{R}^n$ be a steady state of (3) and let $V : \mathbb{R}^n \to \mathbb{R}$ satisfy $V(u_0) = 0$, V locally positive definite, i.e. $\exists r > 0$ such that V(u) > 0 for $u \in B_r(u_0) \setminus \{u_0\}$, and $\nabla V(u) \cdot f(u) \leq 0$ locally, i.e. for $u \in B_r(u_0)$. Then V is called a Lyapunov function for (f, u_0) .

b) For a Lyapunov function V we define for $\delta > 0$ the <u>sublevel set</u> S_{δ} as the connected component of $\{u : V(u) \leq \delta\}$ containing u_0 .

Lemma 5. Let V be a Lyapunov function for (f, u_0) . a) For every small enough r > 0 exists $\delta > 0$, such that $S_{\delta} \subset B_r(u_0)$. b) For every $\delta > 0$ exists r > 0, such that $B_r(u_0) \subset S_{\delta}$.

Proof: a) For given r > 0 choose $\delta > 0$ such that

$$\delta < \min_{\partial B_r(u_0)} V \,,$$

where the right hand side is positive for small enough r because of the local definiteness of V. This implies $S_{\delta} \cap \partial B_r(u_0) = \{\}$. Since also $u_0 \in S_{\delta}$, the connectedness of S_{δ} implies that it cannot contain any points outside of $B_r(u_0)$. b) For given $\delta > 0$ we define the closed level set $\Sigma_{\delta} := \{u \in \mathbb{R}^n : V(u) = \delta\}$. If it is empty, the result holds with arbitrary r > 0. Otherwise let

$$r:=\min_{\Sigma_{\delta}}|u-u_0|>0\,.$$

For $u \in B_r(u_0)$, $V(u) > \delta$ cannot hold since then, by the continuity of V and by $V(u_0) = 0$, V would have to take the value δ somewhere on the straight line segment between u_0 and u, in contradiction to the definition of r.

Lemma 6. Let V be a Lyapunov function for (f, u_0) . Then for small enough δ , sublevel sets S_{δ} are positively invariant for (3).

Proof: By Lemma 5 a), S_{δ} is bounded for δ small enough. For solutions u of (3), the Lyapunov function is non-increasing along the solution:

$$\frac{d}{dt}V(u(t)) = \nabla V(u(t) \cdot f(u(t)) \le 0,$$

which implies the result.

22

Theorem 11. Let V be a Lyapunov function for (f, u_0) .

a) Then u_0 is stable.

b) If furthermore $-\nabla V \cdot f$ is locally positive definite, then u_0 is asymptotically stable.

c) If V and $-\nabla V \cdot f$ are globally positive definite, and all sublevel sets are bounded, then u_0 is globally asymptotically stable.

Proof: a) Let $\varepsilon > 0$ and let δ be as in Lemma 5 a) with $r = \varepsilon$. For this δ let r be as in Lemma 5 b). Then for $u(0) \in B_r(u_0) \subset S_\delta$ we have $u(t) \in S_\delta \subset B_\varepsilon(u_0)$. b) Let $\delta > 0$ be small enough such that S_δ is bounded and positively invariant, and let u be a solution of (3) with $u(0) \in S_\delta$. Then by monotonicity there exists $\delta_* := \lim_{t\to\infty} V(u(t))$. Assume $\delta_* > 0$. Then every accumulation point u_* of u(t) satisfies $u_* \in \Sigma_{\delta_*}$ and therefore $u(t) \notin B_r(u_0)$ for some r > 0, $t \ge T$. This however implies

$$\limsup_{t \to \infty} \left(\frac{d}{dt} V(u(t)) \right) = \limsup_{t \to \infty} \nabla V(u(t)) \cdot f(u(t)) < 0,$$

a contradiction to the convergence of V(u(t)). Thus $\delta_* = 0$ with the consequence that u_0 is the only accumulation point of u(t).

c) Every u(0) lies in some sublevel set. The rest of the proof is as in b).

Example 1. a) $\ddot{u} + \sin u = 0$. $V(u) = 1 - \cos u + \dot{u}^2/2$. b) The equation $\ddot{u} + \sin u + k\dot{u} = 0$ for a pendulum with friction is equivalent to the first order system

$$\dot{u} = v$$
, $\dot{v} = -\sin u - kv$.

The origin is a steady state which can be shown to be asymptotically stable for k > 0 by linearization. A Lyapunov function is given by the total energy $V(u) = 1 - \cos u + v^2/2$. However, the decay

$$\dot{V} = -kv^2$$

is not negative definite.

c) Still for the damped pendulum, we try $V_{\varepsilon}(u) = 1 - \cos u + v^2/2 + \varepsilon uv, 0 < \varepsilon \ll 1$. Using the second order Taylor polynomial of the cosine and Young's inequality (see Appendix 2) with $p = q = 2, \gamma = 1$, we obtain

$$V_{\varepsilon}(u) \approx \frac{u^2 + v^2}{2} + \varepsilon uv \ge \frac{1 - \varepsilon}{2} (u^2 + v^2),$$

showing the local definiteness of V_{ε} for ε small enough. For the decay of V_{ε} we have

$$\begin{aligned} \dot{V}_{\varepsilon} &= -kv^2 + \varepsilon v^2 - \varepsilon u(\sin u + kv) \approx -\varepsilon u^2 - (k - \varepsilon)v^2 - \varepsilon kuv \\ &\leq -\varepsilon \left(1 - \frac{k\gamma}{2}\right)u^2 - \left(k - \varepsilon - \frac{\varepsilon k}{2\gamma}\right)v^2, \end{aligned}$$

where we have used Young's inequality again with p = q = 2, but now with general γ . Obviously the right hand side can be made negative definite by first choosing γ and then ε small enough. This shows the asymptotic stability of the origin by the Lyapunov function V_{ε} .

Example 2. Gradient flows: $f(u) = -\nabla V(u)$.

Example 3. Population dynamics, two populations:

$$\dot{u} = f(u, v)u, \qquad \dot{v} = g(u, v)v,$$

with

- a) cooperation: $\partial_v f, \partial_u g > 0$,
- b) competition: $\partial_v f, \partial_u g < 0$,
- c) predator-prey: $\partial_v f > 0, \ \partial_u g < 0$.

8. LIMIT CYCLES

Definition 8. A limit cycle of (3) is a periodic solution $u_{\infty}(t)$ with the additional property that there exists at least one other solution u(t), $t \ge 0$, and $\tau \in \mathbb{R}$ such that

$$\lim_{t \to \infty} (u(t) - u_{\infty}(\tau + t)) = 0.$$

This section is concerned with several ways to find limit cycles. This will be done for a prototypical example, the van der Pol oscillator:

(33)
$$\ddot{u} + u = r(1 - u^2)\dot{u}, \qquad r > 0.$$

8.1. Multiple scales. Here we will be concerned with small positive values of r, which we indicate by replacing the symbol r by ε :

(34)
$$\ddot{u} + u = \varepsilon (1 - u^2) \dot{u} \,.$$

As initial conditions we choose

(35)
$$u(0) = \overline{u} \in \mathbb{R}, \qquad \dot{u}(0) = 0$$

It is a natural idea to approximate the solution by a power series in ε and make the ansatz

$$u(t) = \sum_{k=0}^{N} \varepsilon^{k} u_{k}(t) + O(\varepsilon^{N+1}), \qquad N \in \mathbb{N}.$$

Conditions for the coefficients u_k can be found by substitution of the ansatz in (34), (35), by expanding the resulting expressions again in powers of ε , and by comparing coefficients. At the leading order, this leads to

$$\ddot{u}_0 + u_0 = 0$$
, $u_0(0) = \overline{u}$, $\dot{u}_0(0) = 0$,

with the solution $u_0(t) = \overline{u} \cos t$. At $O(\varepsilon)$ we obtain

$$\ddot{u}_1 + u_1 = (1 - u_0^2)\dot{u}_0 = \overline{u}\left(\frac{\overline{u}^2}{4} - 1\right)\sin t + \frac{\overline{u}^3}{4}\sin(3t), \qquad u_1(0) = \dot{u}_1(0) = 0,$$

where the differential equation is an inhomogeneous version of the equation for u_0 . The first term on the right hand side produces *resonance*. The solution is given by

$$u_1(t) = \frac{\overline{u}}{2} \left(\frac{\overline{u}^2}{4} - 1\right) \left(\sin t - t\cos t\right) + \frac{\overline{u}^3}{32} (3\sin t - \sin(3t)).$$

The resonance term $t \cos t$ makes our approach questionable, if we want to use it on long time intervals. For $t = O(1/\varepsilon)$ the correction $\varepsilon u_1(t)$ is not small compared to $u_0(t)$ any more.

In order to understand where the problem comes from, we analyze two simpler examples:

Example 4.

$$\ddot{u} + u = \varepsilon u$$
, $u(0) = \overline{u}$, $\dot{u}(0) = 0$,

the harmonic oscillator with perturbation of the frequency. By the modified frequency, the exact solution $u(t) = \overline{u} \cos(t\sqrt{1-\varepsilon})$ is far from the approximation $u_0(t) = \overline{u} \cos t$ for large t. It seems more natural to improve u_0 by an expansion of the frequency:

$$\sqrt{1-\varepsilon} = 1 - \frac{\varepsilon}{2} + O(\varepsilon^2) \,,$$

and to use the approximation

$$u(t) \approx \overline{u} \cos\left(t - \frac{\varepsilon t}{2}\right)$$
,

which looses its approximation quality for larger times than u_0 and, most importantly, reproduces the qualitative long time behavior of the exact solution in contrast to $u_0(t) + \varepsilon u_1(t)$.

Example 5.

$$\ddot{u} + u = -2\varepsilon \dot{u}, \qquad u(0) = \overline{u}, \quad \dot{u}(0) = 0$$

the harmonic oscillator with small friction. Now the exact solution is given by

$$u(t) = \overline{u}e^{-\varepsilon t}\cos\left(t\sqrt{1-\varepsilon^2}\right)$$

Obviously the decay to zero cannot be described by an expansion in powers of ε . There are two effects happening at the same time at two different time scales: oscillations at the scale t and exponential decay at the scale εt .

The method of multiple scales uses the rescaled times

$$T_j = \varepsilon^j t, \qquad j = 0, 1, \dots,$$

as independent variables. This means that formally the solution of the initial value problem (34), (35) for the van der Pol equation is written as

$$u(t) = U(T_0, T_1, \ldots) \, .$$

For the time derivatives we obtain

$$\frac{d}{dt} = \frac{\partial}{\partial T_0} + \varepsilon \frac{\partial}{\partial T_1} + O(\varepsilon^2), \qquad \frac{d^2}{dt^2} = \frac{\partial^2}{\partial T_0^2} + 2\varepsilon \frac{\partial^2}{\partial T_0 \partial T_1} + O(\varepsilon^2),$$

and therefore

(36)
$$\frac{\partial^2 U}{\partial T_0^2} + U = \varepsilon \left((1 - U^2) \frac{\partial U}{\partial T_0} - 2 \frac{\partial^2 U}{\partial T_0 \partial T_1} \right) + O(\varepsilon^2)$$

An asymptotic expansion $U = U_0(T_0, T_1) + \varepsilon U_1(T_0, T_1) + O(\varepsilon^2)$ of the new unknown leads to

$$\frac{\partial^2 U_0}{\partial T_0^2} + U_0 = 0 \,, \qquad U_0(0,0) = \overline{u} \,, \quad \frac{\partial U_0}{\partial T_0}(0,0) = 0 \,,$$

with the solution

$$U_0(T_0, T_1) = a(T_1)\cos(T_0 + b(T_1)),$$

where a, b satisfy the initial conditions

$$a(0) = \overline{u}, \qquad b(0) = 0$$

Otherwise, a and b are so far undetermined. The $O(\varepsilon)$ -terms in (36) give

$$\begin{aligned} \frac{\partial^2 U_1}{\partial T_0^2} + U_1 &= (1 - U_0^2) \frac{\partial U_0}{\partial T_0} - 2 \frac{\partial^2 U_0}{\partial T_0 \partial T_1} \\ &= a \left(\frac{a^2}{4} - 1 \right) \sin(T_0 + b) + \frac{a^3}{4} \sin(3(T_0 + b)) \\ &+ 2 \frac{\partial a}{\partial T_1} \sin(T_0 + b) + 2a \frac{\partial b}{\partial T_1} \cos(T_0 + b) \,. \end{aligned}$$

As in the naive approach at the beginning of this section, the terms with $\sin(T_0 + b)$ and with $\cos(T_0 + b)$ would produce resonance, i.e. an unbounded correction U_1 . Now the idea is to use the remaining freedom in choosing a and b to eliminate these terms:

$$\frac{\partial a}{\partial T_1} = \frac{a}{2} \left(1 - \frac{a^2}{4} \right) , \qquad \frac{\partial b}{\partial T_1} = 0 .$$

With the above initial conditions, a and b are determined uniquely. The equation for the amplitude a has three steady states, a = 0 (unstable) and $a = \pm 2$ (asymptotically stable). This predicts that all solutions with $\overline{u} \neq 0$ converge to a limit cycle with amplitude 2. The problem for a can also be solved explicitly, and we finally arrive at the formal approximation

$$u(t) = \frac{2\overline{u}\cos t}{\sqrt{4e^{-\varepsilon t} + \overline{u}^2(1 - e^{-\varepsilon t})}} + O(\varepsilon) \,.$$

To make this result rigorous would go beyond the aims of this course.

26

8.2. The Poincaré map. The Poincaré map is an alternative method for deriving the result of the previous section. Here we consider the ODE (34) only on finite time intervals. Therefore the expansion

$$u(t) = \overline{u}\cos t + \varepsilon \left(\frac{\overline{u}}{2}\left(\frac{\overline{u}^2}{4} - 1\right)\left(\sin t - t\cos t\right) + \frac{\overline{u}^3}{32}(3\sin t - \sin(3t))\right) + O(\varepsilon^2)$$

for the solution starting at $u(0) = \overline{u} > 0$, $\dot{u}(0) = 0$, can be used. In particular, we are interested in the point, where the trajectory in the (u, \dot{u}) -plane crosses the positive *u*-axis for the first time t = T > 0 after t = 0. By the expansion we expect

$$T = 2\pi + \varepsilon T_1 + O(\varepsilon^2) \,.$$

The $O(\varepsilon)$ -terms in the equation $\dot{u}(T) = 0$ give $T_1 = 0$. Therefore we have

$$u(T) = \overline{u} - \varepsilon \pi \overline{u} \left(\frac{\overline{u}^2}{4} - 1\right) + O(\varepsilon^2) =: f(\overline{u}).$$

The Poincaré map f maps the positive u-axis to itself. The long time behavior of u(t) can be understood by iterating f. The recursion

$$U_{n+1} = f(U_n)$$

has the fixed point U = 0 and another one close to U = 2. Fixed points of f correspond to periodic solutions of (34). Since

$$f'(0) = 1 + \varepsilon \pi + O(\varepsilon^2), \qquad f'(2) = 1 - 2\varepsilon \pi + O(\varepsilon^2),$$

U = 0 is unstable and the second steady state is asymptotically stable for small ε . As in the preceding section, we conclude the existence of a stable limit cycle approximated by $2 \cos t$. Again we do not make this formal analysis rigorous, although it is not very difficult in this case.

8.3. Relaxation oscillations. Now we consider large values of r in (33) and therefore set $r = 1/\varepsilon$. After rescaling time by $t \to t/\varepsilon$ the equation reads

(37)
$$\varepsilon^2 \ddot{u} + u = (1 - u^2) \dot{u} \,,$$

which we rewrite as the first order system

(38)
$$\varepsilon^2 \dot{u} = u - \frac{u^3}{3} - v, \qquad \dot{v} = u$$

Both the second order equation and the first order system are singularly perturbed in the sense that in the limit $\varepsilon \to 0$ the differentiation order is reduced. This means for example that in general the limiting equations cannot satisfy initial conditions.

Example 6. A simple example with a singularly perturbed ODE is the initial value problem

$$\varepsilon \dot{u} = -u + t, \quad u(0) = 1, \qquad 0 < \varepsilon \ll 1,$$

with the exact solution $u_{\varepsilon}(t) = t - \varepsilon + (1 + \varepsilon)e^{-t/\varepsilon}$. For every fixed t > 0, we have

$$\lim_{\varepsilon \to 0} u_{\varepsilon}(t) =: \overline{u}(t) = t$$

which can be obtained from the formal limit in the differential equation. However, the limit is not uniform with respect to t close to t = 0. In terms of the fast time scale $\tau = t/\varepsilon$ we have

$$\lim_{\varepsilon \to 0} u_{\varepsilon}(\varepsilon \tau) = e^{-\tau} \,,$$

and the sum of these two limits provides a uniformly valid approximation:

$$u_{\varepsilon}(t) = t + e^{-t/\varepsilon} + O(\varepsilon).$$

The second part can also be obtained by performing the rescaling in the differential equation,

$$\frac{du}{d\tau} = -u + \varepsilon \tau \,,$$

passing to the limit $\varepsilon \to 0$, and solving the resulting equation subject to the initial condition u(0) = 1.

For the van der Pol system (38) the limit $\varepsilon \to 0$ leads to the differentialalgebraic system

(39)
$$0 = \overline{u} - \frac{\overline{u}^3}{3} - \overline{v}, \qquad \dot{\overline{v}} = \overline{u}.$$

defining a flow along the N-shaped curve S given by the first equation. However, S should be split into its three monotone branches separated by the points (-1, -2/3) and (1, 2/3), along which \overline{u} can be expressed as a function of \overline{v} . Additionally, the middle branch is separated into two parts by the steady state at the origin. The flow along the two right branches, i.e. $\overline{u} > 0$, is towards (1, 2/3) and along the left branches towards (-1, -2/3). These points are reached in finite time. The dynamics away from S is described in terms of the fast time scale $\tau = t/\varepsilon^2$:

$$\frac{du}{d\tau} = u - \frac{u^3}{3} - v$$
, $\frac{dv}{d\tau} = \varepsilon^2 u$.

The limit $\varepsilon \to 0$ gives

$$\frac{d\hat{u}}{d\tau} = \hat{u} - \frac{\hat{u}^3}{3} - \hat{v} \,, \qquad \frac{d\hat{v}}{d\tau} = 0$$

This describes a flow along horizontal lines ($\hat{v} = \text{const}$), with steady states on the curve \mathcal{S} . For this flow, the points on the two outer branches of \mathcal{S} are stable, whereas the middle branch is unstable. The right branch attracts all points in the (u, v)-plane with v < -2/3, or with $-2/3 \le v < 2/3$, if they lie to the right of the middle branch. Let us start a trajectory in this region. Very fast, i.e. as a function of τ , the trajectory will go horizontally to the right branch of \mathcal{S} . There we switch to the slow time scale t and solve (39), until we reach (1, 2/3). The only way to continue from there is another fast horizontal move to the left branch

28

of S, which is met at the point (-2, 2/3). Another slow move along S takes us down to (-1, -2/3), from where we go fast and horizontally back to the right branch, which we meet at (2, -2/3), and which we follow slowly up to (1, 2/3), closing a periodic loop, which consists of two pieces of S and two horizontal line segments. In terms of the approximative dynamics described here, this periodic orbit attracts all other points and therefore it is a stable limit cycle.

Of course the results are again only formal. Rigorous proofs are available, but well beyond the scope of this course.

8.4. The Hopf bifurcation. In this section bifurcation theory is used for finding limit cycles. In (33) we rescale the unknown by $u \to u/\sqrt{r}$:

(40)
$$\ddot{u} + u = r\dot{u} - u^2\dot{u}.$$

The linearization

$$\ddot{z} + z = r\dot{z}$$

at the origin has a bifurcation at r = 0, where the eigenvalues

$$\lambda = \frac{r}{2} \pm i\sqrt{1 - \frac{r^2}{4}}$$

cross the imaginary axis, but without the occurrence of a zero eigenvalue as in the bifurcations considered in Section 5. Obviously this requires a system of at least second order.

We rewrite (40) as a first order system:

$$\dot{u} = v, \qquad \dot{v} = -u + rv - u^2 v,$$

Since the trajectories of the linearization at the bifucation point r = 0 are circles, it seems natural to introduce polar coordinates:

$$u = \rho \sin \varphi, \qquad v = \rho \cos \varphi,$$

which, after some computation, leads to

$$\dot{\varrho} = r \varrho \cos^2 \varphi - \varrho^3 \sin^2 \varphi \cos^2 \varphi, \dot{\varphi} = 1 - r \sin \varphi \cos \varphi + \varrho^2 \sin^3 \varphi \cos \varphi.$$

For small r and ρ , the first equation implies $\dot{\rho}$ is small compared to ρ . Therefore, on finite time intervals ρ can be approximated by a constant $\overline{\rho}$. The right hand side of the second equation can be approximated by 1. This leads to

$$\frac{d\varrho}{d\varphi} \approx r\overline{\varrho}\cos^2\varphi - \overline{\varrho}^3\sin^2\varphi\,\cos^2\varphi\,.$$

In order for having a periodic solution close to $\rho = \overline{\rho} > 0$, the integral of the right hand side with respect to φ over integrals of length 2π needs to vanish:

$$\pi r - A\overline{\varrho}^2 = 0$$

with

$$A = \int_0^{2\pi} \sin^2 \varphi \, \cos^2 \varphi \, d\varphi = \frac{\pi}{4} \, .$$

This gives $\overline{\rho} = 2\sqrt{r}$ in agreement with our results from Sections 8.1 and 8.2.

Note that the constant A above resulted from the nonlinearity. For systems with the same linearization but different nonlinearities the result differs only in the value of A. For the bifurcation to be called a Hopf bifurcation there are two requirements:

- (1) A pair of complex conjugate eigenvalues crosses the imaginary axis and
- (2) the constant A is different from zero.

The Hopf bifurcation comes in two different varieties:

- For A > 0 (as for the van der Pol oscillator) the bifurcation is called *supercritical*: For r < 0 there is a stable steady state, and for r > 0 an unstable steady state and a stable limit cycle, bifurcating from the steady state.
- For A < 0 the bifurcation is called *subcritical*: For r < 0 there is a stable steady state and an unstable limit cyle, and for r > 0 only an unstable steady state.

8.5. The Poincaré-Bendixson theorem.

Definition 9. For the ODE system (3) with n = 2 and a point $\overline{u} \in \mathbb{R}^2$, which is not a steady state, i.e. $f(\overline{u}) \neq 0$, a line segment

(41)
$$T(\overline{u}) := \{\overline{u} + sf(\overline{u})^{\perp} : s \in (-\delta, \delta)\}, \qquad \delta > 0,$$

where $f(\overline{u})^{\perp}$ is orthogonal to $f(\overline{u})$, and where

$$f(u) \cdot f(\overline{u}) \neq 0 \qquad \forall u \in T(\overline{u}),$$

is called a transversal line segment at \overline{u} .

By continuity of f, a transversal line segment is given by (41) for any small enough δ . The flow defined by (3) induces an orientation of $T(\overline{u})$: We shall say that it goes from the minus-side of $T(\overline{u})$ to the plus-side.

Lemma 7. Let $u_0, \overline{u} \in \mathbb{R}^2$, consider the forward trajectory $S_+(u_0) := \{S_t(u_0) : t \ge 0\}$ of (3) with n = 2. Then for the (empty, finite, or infinite) set

$$S_{+}(u_{0}) \cap T(\overline{u}) = \{S_{t_{j}}(u_{0}) = \overline{u} + s_{j}f(\overline{u})^{\perp} : t_{j} < t_{j+1}, j \in J\},\$$

the sequence $\{s_i\}_{i \in J}$ is monotone.

Proof: Only when $S_+(u_0) \cap T(\overline{u})$ has at least three elements $S_{t_j}(u_0)$, $S_{t_{j+1}}(u_0)$, $S_{t_{j+2}}(u_0)$, there is something to prove. W.l.o.g. we assume $s_j < s_{j+1}$ (otherwise change the orientation of $f(\overline{u})^{\perp}$) and construct a Jordan curve (closed, simple)

$$C := \{S_t(u_0) : t_j \le t \le t_{j+1}\} \cup \{\overline{u} + sf(\overline{u})^{\perp} : s_j \le s \le s_{j+1}\}.$$

By the Jordan curve theorem (see Appendix), $\mathbb{R}^2 \setminus C$ has two connected components, the bounded interior of C and the unbounded exterior of C. There are three possibilities: Either $S_{t_j}(u_0) = S_{t_{j+1}}(u_0)$, whence the trajectory is periodic and $S_+(u_0) \cap T(\overline{u})$ consists of only one point, or $f(S_{t_{j+1}}(u_0))$ points into the

30

interior of C, or it points into the exterior of C. In the first case there is nothing to prove. In the second case the interior of C is a positively invariant region, which means that $S_{t_{j+2}}(u_0)$ has to lie there, implying $s_{j+2} > s_{j+1}$. In the third case, an analogous argument with the exterior instead of the interior leads to the same conclusion, completing the proof.

Theorem 12. Consider the dynamical system generated by (3) with n = 2. Let $u_0 \in \mathbb{R}^2$ have a bounded forward trajectory $S_+(u_0)$, and assume that the omega limit $\omega(u_0)$ does not contain any steady states. Then $\omega(u_0)$ is a periodic trajectory.

Proof: By the boundedness of the forward trajectory, $\omega(u_0)$ is nonempty, positively invariant, connected, and closed (Theorem 4). Choose $\overline{u} \in \omega(u_0)$, $\tilde{t}_j \to \infty$ such that $S_{\tilde{t}_j}(u_0) \to \overline{u}$, and a transversal line segment $T(\overline{u})$. Then for $j \ge j_0$, the trajectory crosses $T(\overline{u})$ at $S_{t_j}(u_0)$ close to $S_{\tilde{t}_j}(u_0)$. Note that $\{t_j\}_{j\ge j_0}$ is a subsequence of the sequence with the same name from Lemma 7. The construction also implies $S_{t_j}(u_0) \to \overline{u}$. Lemma 7, i.e. the monotonicity of the crossing points along $T(\overline{u})$, implies

$$\omega(u_0) \cap T(\overline{u}) = \{\overline{u}\}.$$

By the positive invariance of $\omega(u_0)$ the forward trajectory through \overline{u} satisfies $S_+(\overline{u}) \subset \omega(u_0)$, implying

$$S_+(\overline{u}) \cap T(\overline{u}) = \{\overline{u}\}.$$

Therefore $S_+(\overline{u})$ is periodic.

Assume now that $\omega(u_0) \setminus S_+(\overline{u})$ is nonempty. The connectedness of $\omega(u_0)$ then implies that there are $u_1 \in S_+(\overline{u}), u_2 \in \omega(u_0) \setminus S_+(\overline{u})$ with $|u_1 - u_2|$ arbitrarily small. As a consequence, the trajectory through u_2 crosses a transversal line segment $T(u_1)$ in a point $S_{t_2}(u_2)$, and obviously $u_1 \in S_+(\overline{u}) \cap T(u_1)$. Since, however, $\omega(u_0) \cap T(u_1)$ contains at most one point, $S_{t_2}(u_2) = u_1$, implying the contradiction $u_2 \in S_+(\overline{u})$. We conclude that $\omega(u_0) = S_+(\overline{u})$.

Definition 10. A trajectory $\{S_t(u_0) : t \in \mathbb{R}\}$ of a continuous dynamical system is called a <u>heteroclinic orbit</u>, if the limits

$$u_{\pm} = \lim_{t \to \pm \infty} S_t(u_0)$$

exist and $u_{-} \neq u_{+}$. It is called a <u>homoclinic orbit</u>, if $u_{-} = u_{+}$.

Theorem 13. (Poincaré-Bendixson) Consider the dynamical system generated by (3) with n = 2. Let $u_0 \in \mathbb{R}^2$ have a bounded forward trajectory $S_+(u_0)$, and let $\omega(u_0)$ contain finitely many fixed points. Then one of the following holds: a) $\omega(u_0)$ consists of one steady state,

- b) $\omega(u_0)$ is a periodic trajectory,
- c) $\omega(u_0)$ is a union of steady states, homoclinic orbits, and heteroclinic orbits.

Example 7. Instead of a proof of the theorem, we discuss an example for case *c*):

$$\dot{u} = v + u^2 - \frac{u}{4}(v - 1 + 2u^2), \qquad \dot{v} = -2u(1 + v).$$

It has the steady states (0,0), (1,-1), (-1,-1), (2,-1). Furthermore the curves v = -1 and $2u^2 + v = 1$ are invariant. Together with the steady states (1,-1) and (-1,-1) they form the boundary of an invariant region D containing the unstable steady state (0,0). The function

$$H(u,v) = u^{2}(1+v) + \frac{v^{2}}{2}$$

vanishes at (0,0) and is positive elswhere in D. It takes the value 1/2 on ∂D and satisfies

$$\dot{H} > 0$$
 in D , $\dot{H} = 0$ on ∂D .

Since the steady states (1, -1), (-1, -1) on ∂D are saddles, it is clear that every trajectory starting in D except at the origin has ∂D as its omega limit.

For the last time, we return to the van der Pol oscillator system

(42)
$$\dot{u} = -v + f(u), \quad \dot{v} = u, \qquad f(u) = r(u - u^3/3), \quad r > 0.$$

Lemma 8. The dynamical system defined by (42) has a bounded positively invariant set D, and the origin (u, v) = (0, 0) lies in the interior of D.

Proof: Proseminar.

The computation

(43)
$$\frac{d}{dt}\frac{u^2 + v^2}{2} = uf(u) \ge 0 \quad \text{for } |u| \le \sqrt{3},$$

shows that the exterior of $B_r(0,0)$, $r \leq \sqrt{3}$, is also positively invariant. Since the origin is the only steady state, we conclude from the Poincaré-Bendixson theorem that the omega limits of all trajectories starting in $D \setminus \{(0,0)\}$ are periodic orbits lying in $D \setminus B_{\sqrt{3}}(0,0)$. We intend to prove that there is only one such periodic solution.

By the reflection symmetry $(u, v) \leftrightarrow (-u, -v)$ of (42), a periodic orbit crosses the *v*-axis at opposite points $(0, v_0)$ and $(0, -v_0)$. We can therefore consider a modified Poincaré map *P* by considering trajectories starting at u(0) = 0, $v(0) = -v_0 < 0$, define T > 0 as the smallest value, where u(T) = 0, v(T) > 0holds, and set $P(v_0) = v(T)$. Obviously, fixed points of *P* correspond to periodic solutions of (42). Using (43), this requires to find zeroes of

$$W(v_0) = \int_0^T u(t)f(u(t))dt \,.$$

The existence of such zeroes is already known form the Poincaré-Bendixson theorem. The uniqueness will follow from strict monotonicity of W. Since the trajectory (u(t), v(t)), 0 < t < T, lies in the right half plane, v(t) is strictly increasing, and u(t) has a unique maximum at the point where the trajectory crosses the curve v = f(u). Since f(u) > 0 for $0 < u < \sqrt{3}$, $W(v_0)$ is positive as long as this maximum is not bigger than $\sqrt{3}$. We therefore only consider values of v_0 large enough such that the maximal value of u(t) is larger than $\sqrt{3}$, i.e. the trajectory crosses v = f(u) at a point with v < 0. The trajectory can then be split into three parts, according to the sign of f(u), by the points $0 < t_1 < t_2 < T$ such that

$$u(t_1) = \sqrt{3}$$
, $v(t_1) < 0$, $u(t_2) = \sqrt{3}$, $v(t_2) > 0$.

Accordingly, $W(v_0)$ can be written as the sum of three contributions. In the first one we change to u as integration variable:

(44)
$$\int_0^{t_1} u(t)f(u(t))dt = \int_0^{\sqrt{3}} \frac{uf(u)du}{f(u) - v_1(u,v_0)}$$

The graph of v_1 is the trajectory between t = 0 and $t = t_1$. Obviously it is strictly decreasing as a function of v_0 . Since uf(u) > 0, (44) is strictly decreasing as a function of v_0 . For the third contribution, we proceed analogously:

(45)
$$\int_{t_2}^{T} u(t)f(u(t))dt = \int_{0}^{\sqrt{3}} \frac{uf(u)du}{v_3(u,v_0) - f(u)}$$

where the graph of v_3 is the trajectory between $t = t_2$ and t = T. It is strictly increasing as a function of v_0 , and therefore (45) is also strictly decreasing. For the middle contribution, we use integration with respect to v:

(46)
$$\int_{t_1}^{t_2} u(t) f(u(t)) dt = \int_{\underline{v}(v_0)}^{\overline{v}(v_0)} f(u_2(v, v_0)) dv.$$

As a consequence of the facts that f is negative and strictly decreasing in this region, that $\overline{v}(v_0)$ is strictly increasing, $\underline{v}(v_0)$ is strictly decreasing, and $u_2(v, v_0)$ is strictly increasing as function of v_0 , we conclude that also (46), and therefore W is strictly decreasing. This completes the proof of uniqueness of the periodic solution of the van der Pol equation.

9. The Lorenz equations

The meteorologist Edward N. Lorenz published in 1963 a model for atmospheric flow, together with his numerical observations of strange solution behavior. This was one of the starting points of chaos theory. The Lorenz equations are given by

(47)
$$\begin{aligned} \dot{u} &= \sigma(v-u), \\ \dot{v} &= ru - v - uw, \\ \dot{w} &= uv - bw, \end{aligned}$$

with the positive parameters σ, r, b . The system is invariant under the reflection $(u, v, w) \leftrightarrow (-u, -v, w)$. For r < 1, the origin $\overline{U}_1 = (0, 0, 0)$ is the only steady

state, and it is asymptotically stable, as can be seen from the linearization. Global asymptotic stability can be shown with the help of the Lyapunov function

$$L(u, v, w) = \frac{1}{2}(ru^{2} + \sigma v^{2} + \sigma w^{2}).$$

At r = 1, two more steady states

$$\overline{U}_{2,3} = (\pm \sqrt{b(r-1)}, \pm \sqrt{b(r-1)}, r-1)$$

appear in a pitchfork bifurcation. Linearization at $\overline{U}_{2,3}$ leads to the characteristic equation

$$\lambda^3 + \lambda^2(\sigma + 1 + b) + \lambda(\sigma + r)b + 2\sigma b(r - 1) = 0$$

For $0 < r - 1 \ll 1$, the critical eigenvalue, which is close to zero, can be approximated by $\lambda \approx \frac{2\sigma(1-r)}{\sigma+1} < 0$, showing a transfer of stability from \overline{U}_1 to \overline{U}_2 and to \overline{U}_3 at the bifurcation. These points can only lose their stability if purely imaginary eigenvalues occur for some value of r. Therefore we substitute $\lambda = i\omega$, $\omega \in \mathbb{R}$, in the above equation, giving

$$\omega^2 = b(\sigma + r) = \frac{2\sigma b(r-1)}{\sigma + 1 + b}.$$

Under the condition $\sigma > 1 + b$, the second equation holds for

$$r = r_c := \frac{\sigma(3 + \sigma + b)}{\sigma - 1 - b} > 1$$

At the bifurcation at $r = r_c$, \overline{U}_2 and \overline{U}_3 also lose their stability. The strange behavior mentioned above occurs for $\sigma > 1 + b$, $r > r_c$.

We shall prove that all solutions are attracted to a bounded domain by considering a modification of the above Lyapunov function:

$$L_1(u, v, w) = \frac{1}{2}(ru^2 + \sigma v^2 + \sigma (w - 2r)^2),$$

with the time derivative

$$\dot{L}_1 = -\sigma(ru^2 + v^2 + b(w - r)^2 - br^2),$$

which is negative except in the ellipsoid

$$\hat{E} := \{(u, v, w) : ru^2 + v^2 + b(w - r)^2 \le br^2\}.$$

Now we define

$$M := \max_{\hat{E}} L_1 \quad \text{and} \quad E := \{(u, v, w) : L_1(u, v, w) < M + 1\}.$$

Then $\dot{L}_1 < 0$ outside of E, since $\hat{E} \subset E$.

Lemma 9. The set E defined above is positively invariant for (47). Every trajectory reaches E in finite time.



FIGURE 2. The Lorenz attractor $% \left({{{\left[{{{\rm{T}}_{\rm{E}}} \right]}}} \right)$

Theorem 14. Let $B(t) := \{U(t) : \dot{U} = f(U), U(0) \in B(0)\}$ and $V(t) := \mu(B(t)) = \int_{B(t)} dU$. Then

$$\dot{V}(t) = \int_{B(t)} \nabla_U \cdot f(U) \, dU \, .$$

Proof: (not quite a proof, rather a heuristic argument) We split the boundary

$$\partial B(t) = \bigcup_{k=1}^{K} A_k$$

into K pieces small enough, such that they can be approximated by pieces of hyperplanes, and such that f(U) and the outward unit normal $\nu(U)$ can be approximated by constant vectors $f(U_k)$ and $\nu(U_k)$, $U_k \in A_k$, along A_k . Then after a small time step from t to $t + \Delta t$, the boundary can be approximated by

$$\partial B(t + \Delta t) \approx \bigcup_{k=1}^{K} (A_k + \Delta t f(U_k)).$$

The volume difference can then be approximated by the sum of the volumes of oblique cylinders:

$$V(t + \Delta t) - V(t) \approx \bigcup_{k=1}^{K} \Delta t f(U_k) \cdot \nu(U_k) \Delta A_k$$

where ΔA_k denotes the (n-1)-dimensional surface area of A_k . After division by Δt , the right hand side has the form of a Riemann sum for a surface integral. Therefore with $\Delta t \to 0$ and $N \to \infty$, we obtain

$$\dot{V}(t) = \int_{\partial B(t)} f(U) \cdot \nu(U) \, dA = \int_{B(t)} \nabla_U \cdot f(U) \, dU \, ,$$

where the second equality follows from the divergence theorem.

Application of this result to (47) gives

$$\dot{V} = -(\sigma + 1 + b)V,$$

and therefore $V(t) = V(0)e^{-(\sigma+1+b)t}$. This implies that the union of the omegalimits of all trajectories is the subset of a set with volume zero. However, numerical experiments show that it has a very complicated structure. It is an example for a *strange attractor* (see Fig. 2).

10. Hamiltonian mechanics

For conservative mechanical problems the Hamilton principle of the principle of stationary action can be used to derive the equations of motion. The formal procedure requires generalized coordinates $q = (q_1, \ldots, q_n)$ and a Lagrangian $L(v,q), (v,q) \in \mathbb{R}^{2n}$. For a time evolution q(t) of the generalized coordinates, we define the generalized velocities $v(t) = \dot{q}(t)$. The Hamilton principle states that the time evolution between $t = t_1$ and $t = t_2$ is a stationary point of the action integral

$$\mathcal{I}(q) := \int_{t_1}^{t_2} L(\dot{q}(t), q(t)) dt \,,$$

with given $q(t_1)$ and $q(t_2)$. This implies that Gateaux derivatives of \mathcal{I} in all directions have to vanish at q. A permissible direction for the Gateaux derivative is a function r(t) satisfying $r(t_1) = r(t_2) = 0$. The Gateaux derivative at q in the direction r is then given by

$$\frac{d}{d\varepsilon} \mathcal{I}(q+\varepsilon r) \Big|_{\varepsilon=0} = \int_{t_1}^{t_2} \left(\nabla_q L(\dot{q},q) \cdot r + \nabla_v L(\dot{q},q) \cdot \dot{r} \right) dt$$

$$= \int_{t_1}^{t_2} \left(\nabla_q L(\dot{q},q) - \frac{d}{dt} \nabla_v L(\dot{q},q) \right) \cdot r \, dt \,,$$

where the second equality is due to an integration by parts. The requirement that the right hand side vanishes for all permissible directions r leads to the *Euler-Lagrange equation*

(48)
$$\nabla_q L(\dot{q}, q) = \frac{d}{dt} \nabla_v L(\dot{q}, q) \,,$$

a system of second order ordinary differential equations.

Example 8. For the pendulum with length l and mass m swinging in the (x, y)-plane (with the positive x-axis pointing downwards) the angle with respect to the vertical (x-) direction can be used as generalized coordinate:

$$(x(t), y(t)) = l(\cos q(t), \sin q(t)).$$

The kinetic energy is then given as

$$E_{kin} = m \frac{\dot{x}^2 + \dot{y}^2}{2} = \frac{m l^2 \dot{q}^2}{2} = \frac{m l^2 v^2}{2}.$$

The potential energy due to gravity is

$$E_{pot} = -max = -mal\cos q \,,$$

which is a simplified model with constant acceleration a. The standard rule for obtaining the Lagrangian is

$$L = E_{kin} - E_{pot} = \frac{ml^2v^2}{2} + mal\cos q \,.$$

This gives the Euler-Lagrange equation

$$-mal\sin q = ml^2\ddot{q} \implies \qquad \ddot{q} + \omega^2\sin q = 0\,,$$

with the frequency $\omega = \sqrt{a/l}$.

Definition 11. For a strictly convex smooth function $f : \mathbb{R}^n \supset D(f) \to \mathbb{R}$, its Legendre transform f^* is defined as

$$f^*(p) := \sup_{v \in \mathbb{R}^n} \left(p \cdot v - f(v) \right),$$

for $p \in D(f^*) \subset \mathbb{R}^n$, where the right hand side is finite. For these p:

$$f^*(p) = p \cdot V(p) - f(V(p)), \quad with \ p = \nabla f(V(p)).$$

Lemma 10. a) The Legendre transform of a strictly convex function is strictly convex.
b) f^{**} = f.

Proof: The computation

$$\nabla f^*(p) = V(p) + p \cdot \nabla V(p) - \nabla f(V(p)) \cdot \nabla V(p) = V(p)$$

shows that ∇f and ∇f^* are inverse functions for each other, proving b). Therefore their Jacobians, i.e. the Hessians of f and f^* , are inverse matrices for each other. The positive definiteness of the former thus implies positive definiteness of the latter, proving a).

Typically, the Lagrangian is a strictly convex function of v, and $p = \nabla_v L(v, q)$ for fixed q defines a diffeomorphism between v and the momentum p. In this case the Legendre transform of L with respect to v is called the Hamiltonian

$$H(p,q) := L^*(p,q) = p \cdot V(p,q) - L(V(p,q),q), \qquad p = \nabla_v L(V(p,q),q).$$

Lemma 11. a) The Euler-Lagrange equations (48) are equivalent to the <u>Hamiltonian</u> dynamics

$$\dot{q} = \nabla_p H(p,q) \,, \qquad \dot{p} = -\nabla_q H(p,q) \,.$$

b) The Hamiltonian is constant along trajectories. By this property, it is called a first integral, a constant of motion, or a conserved quantity.

Proof: a) Since L is the Legendre transform of H, we have

$$\dot{q} = v = \nabla_p H \,.$$

On the other hand

$$\nabla_q H = p \cdot \nabla_q V - \nabla_v L \cdot \nabla_q V - \nabla_q L = -\nabla_q L \,,$$

and therefore, with $\nabla_v L = p$, (48) can be written as $\dot{p} = -\nabla_q H$. b)

$$\dot{H} = \nabla_p H \cdot \dot{p} + \nabla_q H \cdot \dot{q} = 0$$

Example 9. For the pendulum we get $p = ml^2 v$ and

$$H = \frac{p^2}{2ml^2} - mal\cos q = E_{kin} + E_{pot} \,,$$

with the Hamiltonian dynamics

$$\dot{q} = \frac{p}{ml^2}, \qquad \dot{p} = -mal\sin q.$$

Example 10. The Einstein equation

$$E_{kin} = m(v)c^2$$
, $m(v) = m_0\sqrt{1 + \frac{|v|^2}{c^2}}$, $v \in \mathbb{R}^3$,

gives the kinetic energy of a relativistic particle with the rest mass m_0 and the speed of light c. With a potential energy $E_{pot}(q)$, $q \in \mathbb{R}^3$, the equations of motion are

$$\frac{d}{dt}\left(\frac{m_0\dot{q}}{\sqrt{1+|\dot{q}|^2/c^2}}\right) = -E'_{pot}(q)\,.$$

The Hamiltonian is given by

$$H = -c^2 \sqrt{m_0^2 - \frac{|p|^2}{c^2}} + E_{pot}(q) \,.$$

Theorem 15. (Liouville) A Hamiltonian flow preserves volume in phase space.

Proof: Application of Theorem 14.

Theorem 16. (Poincaré recurrence theorem) Let $\mathcal{M} \subset \mathbb{R}^n$ be bounded, let S_t be a volume preserving dynamical system on \mathcal{M} , and let $U \subset \mathcal{M}$ be an open set. Then there exists a sequence $t_n \to \infty$, $t_n \in \mathcal{T}$, such that

$$(49) U \cap S_{t_n}(U) \neq \{\}.$$

Proof: Choose $t_0 > 0$ and consider the sequence $S_{2nt_0}(U) \subset \mathcal{M}$. Since all these sets have the same positive volume (the volume of U), they cannot be pairwise disjoint, since otherwise the volume of \mathcal{M} would be infinite. Therefore there exist $0 \leq m_0 < n_0$, such that

$$S_{2m_0t_0}(U) \cap S_{2n_0t_0}(U) \neq \{\}$$

Going back in time by $2m_0t_0$, we obtain

$$U \cap S_{t_1}(U) \neq \{\}, \quad \text{with } t_1 = 2(n_0 - m_0)t_0 \ge 2t_0.$$

Iterating this argument we construct a sequence $t_n \ge 2^n t_0$, satisfying (49).

Trying to solve the equations of motion, first integrals like the Hamiltonian are useful, of course. A way of finding additional first integrals is via *continuous symmetries*. Symmetries are defined via the actions of groups on the phase space. The action of a continuous group on the generalized coordinates q is written as $Q(s,q), s \in \mathbb{R}$, satisfying the group properties Q(0,q) = q and $Q(s + \sigma,q) = Q(s,Q(\sigma,q)), s, \sigma \in \mathbb{R}$. It will be assumed to be generated by the vector field f, i.e. by solving the initial value problem

$$rac{dQ}{ds} = f(Q) \,, \qquad Q(0,q) = q \,.$$

The action V(s, v, q) of the group on the generalized velocities is then defined consistently with the relation $\dot{q} = v$, i.e. we compute

$$\frac{d}{dt}Q(s,q(t)) = (\dot{q}(t) \cdot \nabla_q)Q(s,q(t)),$$

and set

$$V(s, v, q) = (v \cdot \nabla_q)Q(s, q).$$

Definition 12. The function L(v,q) has the symmetry generated by f, if

(50)
$$L(V(s, v, q), Q(s, q)) = L(v, q) \quad \forall s \in \mathbb{R}$$

The following is a famous result by one of the most important female mathematicians.

Theorem 17. (Noether) If the Lagrangian L has the symmetry generated by f, then

$$\mathcal{I}(v,q) = \nabla_v L(v,q) \cdot f(q) = p \cdot f(q)$$

is a first integral of the dynamics given by (48).

Proof: Differentiation of (50) with respect to s and evaluation at s = 0 gives

$$\nabla_v L \cdot ((v \cdot \nabla)f) + \nabla_q L \cdot f = 0.$$

Using this we compute

$$\dot{\mathcal{I}} = \nabla_v L \cdot \left((v \cdot \nabla) f \right) + \left(\frac{d}{dt} \nabla_v L \right) \cdot f = \left(-\nabla_q L + \frac{d}{dt} \nabla_v L \right) \cdot f = 0,$$

where the last equality is due to (48).

Example 11. The simplest example is symmetry with respect to a translation of the generalized coordinates. With $f(q) = e_j$, the *j*-th canonical basis vector in \mathbb{R}^n , we get $Q(s,q) = q + se_j$, V(s,v,q) = v, which means that the Lagrangian has the corresponding symmetry, if it is independent of q_j . In this case the *j*-th component of the momentum

$$p_j = \nabla_v L \cdot e_j$$

is a conserved quantity. Here we would not have needed the Noether theorem, since this is an obvious consequence of the equations of motion.

Example 12. Assume that $L(v,q) = \hat{L}(|v|, |q|)$. We consider the symmetry action generated by f(q) = Aq with a skew symmetric matrix A, i.e. $A^{tr} = -A$. Then we have Q(s,q) = R(s)q, V(s,v,q) = R(s)v, where the matrix R(s) satisfies

$$\frac{d}{ds}R = AR, \qquad R(0) = I_n,$$

and it is orthogonal, i.e. $R^{tr} = R^{-1}$, since

 $\frac{d}{ds}R^{tr}R = R^{tr}AR + (AR)^{tr}R = R^{tr}AR + R^{tr}A^{tr}R = 0 \quad and \quad R(0)^{tr}R(0) = I_n .$ As a consequence

$$|Q(s,q)|^{2} = (Rq) \cdot (Rq) = q^{tr} R^{tr} Rq = |q|^{2}, \qquad |V(s,v,q)|^{2} = |v|^{2},$$

showing the L has the corresponding symmetry, i.e. it is rotationally symmetric. As a consequence, $p^{tr}Aq$ is conserved for any skew symmetric matrix A. This

40

gives the $\frac{n(n-1)}{2}$ (the dimension of the space of skew symmetric matrices, or the number of pairs (i, j) with i < j) independent conserved quantities

$$\mathcal{I}_{ij} := p_i q_j - p_j q_i , \qquad 1 \le i < j \le n .$$

For n = 3 this can be written as conservation of the <u>angular momentum</u> $p \times q$, where \times denotes the vector product.

Example 13. Point particles with binary interactions: Consider K particles with masses $m^1, \ldots, \overline{m^K}$ in three-dimensional space with positions $q^1, \ldots, q^K \in \mathbb{R}^3$ and velocities $v^1, \ldots, v^K \in \mathbb{R}^3$. We collect them in the vectors $q = (q^1, \ldots, q^K)$, $v = (v^1, \ldots, v^K) \in \mathbb{R}^n$, n = 3K. The Lagrangian is given by

$$L(v,q) = \sum_{k=1}^{K} \frac{m^k |v^k|^2}{2} - \sum_{1 \le k < l \le K} E_{kl}(|q^k - q^l|),$$

with the potential energy $E_{kl}(|q^k - q^l|)$ of a binary interaction assumed to only depend on the distance between the particles. The Lagrangian has translation symmetries generated by

$$f_a(q) = (a, \ldots, a), \qquad a \in \mathbb{R}^3$$

with the group action $Q_a(s,q) = (q^1 + sa, \ldots, q^K + sa), V_a(s,v,q) = v$, and rotation symmetries generated by

$$f_A(q) = (Aq^1, \dots, Aq^K), \qquad A \in \mathbb{R}^{3 \times 3}, \quad A^{tr} = -A,$$

with the group action $Q_A(s,q) = (R(s)q^1, \ldots, R(s)q^K), V_A(s,v,q) = (R(s)v^1, \ldots, R(s)v^K),$ where R is as in the previous example. By the Noether theorem these lead to conservation of total momentum and of total angular momentum, i.e. of

$$p_0 = \sum_{k=1}^{K} p^k = \sum_{k=1}^{K} m^k v^k$$
 and of $l_0 = \sum_{k=1}^{K} p^k \times q^k$.

Of course we also have conservation of the total energy

77

$$E = H(p,q) = \sum_{k=1}^{K} \frac{|p^k|^2}{2m^k} + \sum_{1 \le k < l \le K} E_{kl}(|q^k - q^l|)$$

Example 14. A special case of the previous example is (Newton's) gravitational interaction with

$$E_{kl}(|q^k - q^l|) = -\frac{m^k m^l G}{|q^k - q^l|},$$

where G > 0 denotes the gravitational constant. The equations of motion are given by

$$\dot{q^k} = v^k$$
, $\dot{v^k} = \sum_{l \neq q} \frac{m^l G(q^l - q^k)}{|q^l - q^k|^3}$, $k = 1, \dots, K$.

42

These are invariant under <u>Galilei transformations</u> of the form $v^k \rightarrow v^k - v_0$, $q^k \rightarrow q^k - tv_0$, i.e. change to a new frame of reference moving with constant velocity v_0 . The choice

$$v_0 = \frac{\sum_{k=1}^K m^k v^k}{\sum_{k=1}^K m^k} \,,$$

i.e the time independent average velocity, makes the total momentum vanish after the transformation: $p_0 = 0$. This can be seen as making the center of mass

$$q_0 = \frac{\sum_{k=1}^K m^k q^k}{\sum_{k=1}^K m^k} \,,$$

motionless, and it will be assumed at the origin in the following: $q_0 = 0$.

We shall consider the <u>Kepler problem</u> or <u>two-body problem</u> with K = 2. With a frame of reference as described above, where the center of mass is fixed at the origin, we have

(51)
$$m^1 q^1 + m^2 q^2 = 0, \qquad m^1 v^1 + m^2 v^2 = 0.$$

We also have the conservation of energy and of angular momentum,

$$E = m^{1} \frac{|v^{1}|^{2}}{2} + m^{2} \frac{|v^{2}|^{2}}{2} - \frac{m^{1} m^{2} G}{|q^{1} - q^{2}|}, \qquad l_{0} = m^{1} v^{1} \times q^{1} + m^{2} v^{2} \times q^{2},$$

and the equations of motion

$$\dot{q^1} = v^1$$
, $\dot{q^2} = v^2$, $m^1 \dot{v_1} = -m^2 \dot{v_2} = \frac{m^1 m^2 G(q^2 - q^1)}{|q^2 - q^1|^3}$.

These imply that q^1 and q^2 are orthogonal to l_0 . We introduce an orthonormal basis $\{b^1, b^2\}$ of the orthogonal complement of l_0 such that $b^1 \times b^2 = l_0/|l_0|$, and we introduce polar coordinates by

$$q^1 - q^2 = r \cos \varphi \, b^1 + r \sin \varphi \, b^2 \, .$$

With (51), q^1, q^2, v^1, v^2 can be computed in terms of r and φ . The conservation of angular momentum can then be written as

$$r^2 \dot{\varphi} = -\frac{|l_0|}{m^*}, \qquad with \quad m^* = \frac{m^1 m^2}{m^1 + m^2},$$

and the conservation of energy as

$$E = \frac{m^*}{2}(\dot{r}^2 + r^2\dot{\varphi}^2) - \frac{m^1m^2G}{r}$$

The former can be used to eliminate $\dot{\varphi}$ from the latter, producing a first order differential equation for r(t). Its solution is facilitated by writing r as a function of φ via

$$\dot{r} = -\frac{dr}{d\varphi} \, \frac{|l_0|}{m^* r^2} \,,$$

giving

$$E = \frac{|l_0|^2}{2m^*} \left(\frac{1}{r^4} \left(\frac{dr}{d\varphi} \right)^2 + \frac{1}{r^2} \right) - \frac{m^1 m^2 G}{r} \,.$$

This will be simplified by a series of transformations. First, $\rho = 1/r$, leading to

$$\left(\frac{d\varrho}{d\varphi}\right)^2 + (\varrho - A)^2 = B^2, \quad \text{with} \quad A = \frac{m^* m^1 m^2 G}{|l_0|^2}, \quad B^2 = A^2 + \frac{2m^* E}{|l_0|^2}.$$

This suggests $\rho = A + B \cos \psi$, with the result

$$\left(\frac{d\psi}{d\varphi}\right)^2 = 1\,.$$

With the solution $\psi = \varphi - \varphi_0$, we finally obtain

$$r = \frac{A^{-1}}{1 + \varepsilon \cos(\varphi - \varphi_0)}, \quad with \quad \varepsilon = \sqrt{1 + \frac{2E|l_0|^2}{m^*(m^1m^2)^2G^2}}.$$

Introducing the cartesian coordinates $(x, y) = r(\cos(\varphi - \varphi_0), \sin(\varphi - \varphi_0))$, this can be written as

$$(1-\varepsilon^2)x^2 + y^2 = \frac{1}{A^2} + \frac{2\varepsilon x}{A},$$

showing that the trajectories of $q^1 - q^2$, and therefore also of q^1 and q^2 are conic sections. In particular, for negative energy E, i.e. $\varepsilon < 1$, they are ellipses, and for positive energies they are hyperbolas, with parabolas in the intermediate case E = 0.

Appendix 1 – second order Taylor remainders

Let $g:\,\mathbb{R}^n\to\mathbb{R}^n$ be smooth in a neighborhood of 0. Then the second order remainder

$$r(v) = g(v) - g(0) - Dg(0)v$$

satisfies

$$r(v_1) - r(v_2) = [Dg(\hat{v}) - Dg(0)](v_1 - v_2),$$

with \hat{v} between v_1 and v_2 . Assume $v_1, v_2 \in B_r$ and let L denote the Lipschitz constant of Dg in B_r . Then also $\hat{v} \in B_r$, and we have

(52)
$$|r(v_1) - r(v_2)| \le rL|v_1 - v_2|,$$

i.e. the Lipschitz constant of the second order remainder is locally small.

Appendix 2 – Young's inequality

Lemma 12. Let $a, b, \gamma > 0, p \ge 1, 1/p + 1/q = 1$. Then

$$ab \leq \frac{\gamma a^p}{p} + \frac{b^q}{\gamma^{q-1}q} \, .$$

Proof: With $\alpha = \gamma^{1/p} a$, $\beta = \gamma^{-1/p} b$, it suffices to prove the inequality with $\gamma = 1$. With t = 1/p, 1 - t = 1/q, the right hand side is a convex combination of α^p and β^q . Therefore the concavity of the logarithm implies

 $\log\left(t\alpha^p + (1-t)\beta^q\right) \ge t\log(\alpha^p) + (1-t)\log(\beta^q) = \log\alpha + \log\beta = \log(\alpha\beta).$

By the monotonicity of the logarithm the proof is complete.

References

- D. Brink, Hölder continuity of roots of complex and p-adic polynomials, Comm. in Algebra 38 (2010), pp. 1658–1662.
- J. Guckenheimer und P. Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, Springer, New York, 1983.
- [3] E.N. Lorenz, Deterministic nonperiodic flow, J. of the Atmospheric Sciences 20 (1963), pp. 130–141.
- [4] S.H. Strogatz, Nonlinear Dynamics and Chaos (with Applications to Physics, Biology, Chemistry, and Engineering), Addison-Wesley Publ. Comp., Reading, 1994.
- [5] G. Teschl, Ordinary Differential Equations and Dynamical Systems, AMS.