Dynamical Systems Lecture Notes

by Leo Butler

E-mail address: l.butler@ed.ac.uk
URL: http://www.maths.ed.ac.uk/~lbutler/

School of Mathematics, Room 6214 James Clerk Maxwell Building, The King's Buildings, University of Edinburgh, Edinburgh EH9 3JZ, Documented generated at 12:41, Friday 5th November, 2010

Contents

Preface iii

Chapter 1. Continuous and discrete dynamical systems 1
1.1. What is a dynamical system? 1
1.2. Examples of dynamical systems 1
1.3. Orbits, fixed points and periodic orbits 4
1.4. Counting periodic points of a map 4
1.5. Exercises 5
1.6. Poincaré maps 5
1.7. Exercises 7

Chapter 2. Linear autonomous systems 9
2.1. Discrete systems 9
2.2. Classification of fixed points in two dimensions 11
2.3. Comparison with continuous systems 11
2.4. Exercises 12

Chapter 3. Fixed points in nonlinear systems 13
3.1. Stability 13
3.2. Stable and unstable manifolds 15
3.3. The Hartman-Grobman Theorem 18
3.4. Eigenvalues with modulus 1 21
3.5. Stability of periodic orbits 25

Chapter 4. One-dimensional dynamics 27
4.1. Introduction 27
4.2. Bifurcation theory 27
4.3. Logistic map: period-doubling cascade and chaos 32
4.4. Chaotic systems and symbolic dynamics 37

Chapter 5. Two-dimensional dynamics 43
5.1. Introduction 43
5.2. Stable and unstable manifolds 43
5.3. Bifurcation theory 45
5.4. Attractors and basins of attraction: the Hénon map 47

Chapter 6. Fractals and their dimensions 49

Preface

Dynamical systems originates in Newtonian mechanics and the study of differential equations. It starts where the study of differential equations and their solutions ends; namely, the study of dynamical systems attempts to cope with two problems:

(i) most systems of differential equations are either impossible to solve in closed form, or their solutions are so complicated as to be uninformative; and

(ii) explicit solutions generally offer quantitative, not qualitative, information.

Let us illustrate this contention with an example. The solutions to the non-linear ('true') frictionless pendulum

$$\ddot{x} = -\frac{mg}{l}\sin(x) \tag{0.0.1}$$

can be expressed, using conservation of energy, in the implicit form as

$$t = \int_{x_0}^{x} \frac{dy}{\sqrt{2E + \frac{mg}{l}\cos(y)}}. \tag{0.0.2}$$

This integral tells us surprisingly little about the solution $x = x(t)$ – without some knowledge of elliptic integrals.

On the other hand, we will develop tools to study differential equations like (0.0.1), and these tools will lead to insights where we have know explicit solutions, like when the pendulum is 'jiggled'. This leads to the fascinating world of chaos...

A note. These notes are marked as a *Draft*. This is because they will be revised as the course progresses. The cover page contains information on when the present version was made.

A final, non-draft, version will be made available in time for the May 2011 exam diet.

CHAPTER 1

Continuous and discrete dynamical systems

1.1. What is a dynamical system?

A dynamical system (DS) consists of:

(i) A state space M, a point $\mathbf{x}$ of which defines the state of a system. M can be $\mathbb{R}^d$, a region of $\mathbb{R}^d$ (e.g., the interval $[0,1]$, the sphere S^2), $\mathbb{C}^d$, the set of all sequences of coin tosses, etc.

(ii) An independent variable (time t) which can be
- continuous, $t \in \mathbb{R}$;
- discrete, $t = n \in \mathbb{Z}$ (or $n \in \mathbb{N}$).

(iii) A rule describing the evolution of the system in time.
- For continuous-time DS, it is given by a system of differential equations (DE):

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t), \quad \text{where} \quad \mathbf{x} \in M \subset \mathbb{R}^d,\, \mathbf{f} : M \times \mathbb{R} \to \mathbb{R}^d,$$

  with the associated flow mapping $\varphi_t(\mathbf{x}) = \mathbf{x}(t)$ where $\mathbf{x}(t)$ solves the DE with the initial condition $\mathbf{x}(0) = \mathbf{x}$.
- For discrete-time DS, it is given by a difference equation (also called a recurrence relation or iterated map):

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n, n), \quad \text{where} \quad \mathbf{x} \in M,\, \mathbf{F} : M \times \mathbb{Z} \to M.$$

Remarks:

- Higher-order difference equations can be turned into sets of first-order ones.
- Non-autonomous difference equations (i.e., with $\mathbf{F}$ depending explicitly on n) can be turned into autonomous ones by introducing an additional dependent variable, m say, obeying $m_{n+1} = m_n + 1$.
- Similarly, higher-order differential equations can be transformed into sets of first-order differential equations, and non-autonomous differential equations can be transformed into autonomous ones.

1.2. Examples of dynamical systems

1.2.1. Particle dynamics. The position $\mathbf{x} \in \mathbb{R}^3$ of a particle of mass m moving under the action of a force $\mathbf{G}(\mathbf{x})$ is governed by Newton's law $m\ddot{\mathbf{x}} = \mathbf{G}$. The DS description $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ is obtained by taking

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ m\dot{\mathbf{x}} \end{pmatrix} \qquad \text{and} \qquad \mathbf{f} = \begin{pmatrix} m^{-1}\mathbf{y}_2 \\ \mathbf{G} \end{pmatrix}.$$

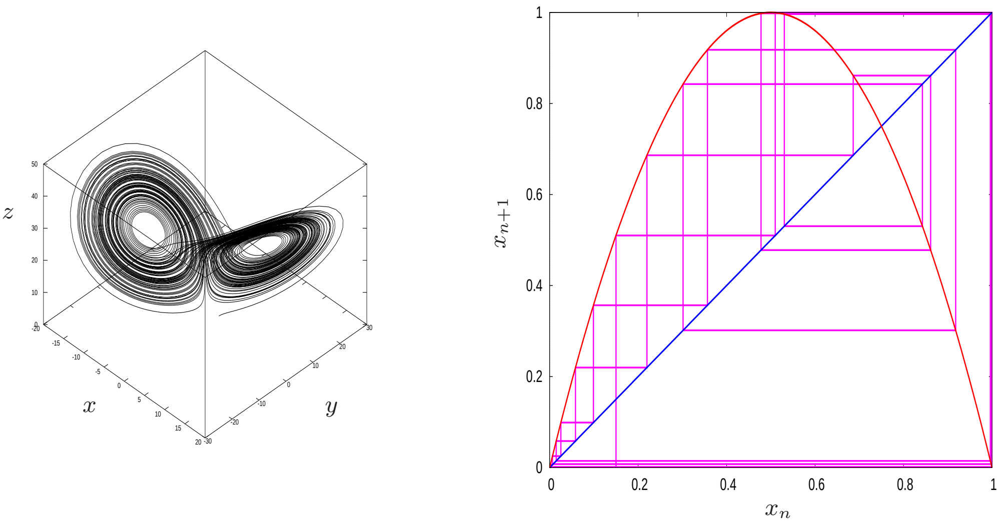
1.2.2. The Lorenz equations. An idealised model of convection in the atmosphere is given by the three ordinary differential equations

$$\dot{x} = \sigma(y - x), \quad \dot{y} = \rho x - y - xz, \quad \dot{z} = xy - \beta z, \tag{1.2.1}$$

where σ, ρ and β are positive constants [see Strogatz, chapter 9 for a nice exposition].

1.2.3. Population dynamics. Discrete dynamical systems arise in models of the evolution of the population of animal species: in these models, the dependent variables $\mathbf{x}_n$ represent the number of individuals of each species at generation n. For a single species, a simple model is given by the map

$$x_{n+1} = x_n r(x_n),$$



(a) The Lorenz attractor for the Lorenz equations (1.2.1) with $\sigma = 10$, $r = 28$ and $\beta = \frac{8}{3}$.

(b) A 'cobweb' diagram of the logistic map (1.2.2) with $\mu = 4$.

FIGURE 1.2.1. The phase portraits of two dynamical systems.

where $r(x)$ is the reproductive rate (difference between birth and death rates), a decreasing function of x. A standard choice for $r(x)$ is $\mu(1-x)$, where μ is a positive constant; this gives the *logistic map*

$$x_{n+1} = \mu x_n(1 - x_n). \tag{1.2.2}$$

Note that the state space M can be taken to be $[0, 1]$ when $0 \le \mu \le 4$; attention is often restricted to $[0, 1]$ even when $\mu > 4$. The dynamics of this system, which we will study in some detail, depends on the value of μ and can be very complicated, as numerical simulations demonstrate. Another model takes $r(x) = \mu \exp(-x)$; it has a dynamics similar to that of the logistic map. [See Strogatz, sections 10.0-5 or Alligood, Sauer and York, sections 1.2 and 1.5] `Worksheet 1`

1.2.4. The bouncing ball. A physical system governed by a discrete mapping is provided by a model of an elastic ball bouncing on a table which is vibrating vertically. Let us derive this mapping. Let $U(t_n)$ be the downward velocity of the ball immediately before the nth impact with the table at time t_n, and let $V(t_n)$ be the upward velocity of the ball immediately after the impact. If the velocity of the table (positive upward) is given by $W(t)$, the impact is modelled by

$$V(t_n) - W(t_n) = \alpha[U(t_n) + W(t_n)],$$

where $0 < \alpha < 1$ is a constant called the restitution coefficient. Between impacts, the ball is uniformly decelerated by gravity; assuming the vertical displacement of the table is negligible, this implies that the time between successive impacts is[1]

$$t_{n+1} - t_n = \frac{U(t_{n+1}) - (\,-V(t_n)\,)}{g},$$

where g is the acceleration of gravity, and the relationship between velocities

$$U(t_{n+1}) = V(t_n).$$

Combining these three equations to eliminate $U(t_n)$ gives the system

$$\begin{aligned} V(t_{n+1}) &= \alpha V(t_n) + (1+\alpha)W(t_{n+1}), \\ t_{n+1} &= t_n + 2V(t_n)/g. \end{aligned}$$

Suppose one takes the table velocity as $W(t) = -\beta\cos(\omega t)$ with two constants β and ω. Replacing $V(t_n)$ and t_n by

$$v_n = 2\omega V(t_n)/g \quad \text{and} \quad \phi_n = \omega t_n$$

[1]Recall: acceleration $= \frac{\text{change in velocity}}{\text{change in time}}$

leads to the standard form

$$\begin{aligned} \phi_{n+1} &= \phi_n + v_n, \\ v_{n+1} &= \alpha v_n - \gamma \cos(\phi_n + v_n), \end{aligned} \tag{1.2.3}$$

where $\gamma = 2\omega\beta(1+\alpha)/g$. When $\alpha = 1$, the system preserves area; the map in this case is called the *standard* or *Chirikov* map and it has been studied quite intensively. This system can be simulated for various values of the parameters α and γ. Note that one can consider ϕ_n modulo 2π since it appears inside a cosine. [See Guckenheimer and Holmes, section 2.4] `Worksheet 1`

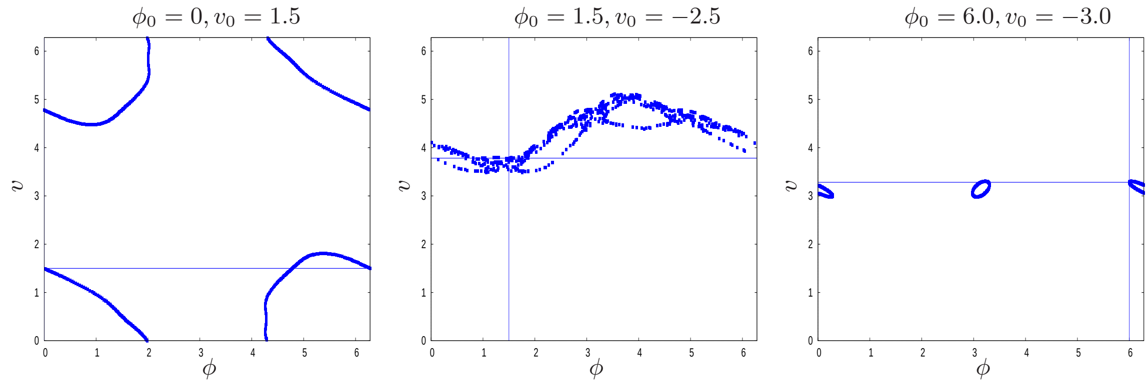
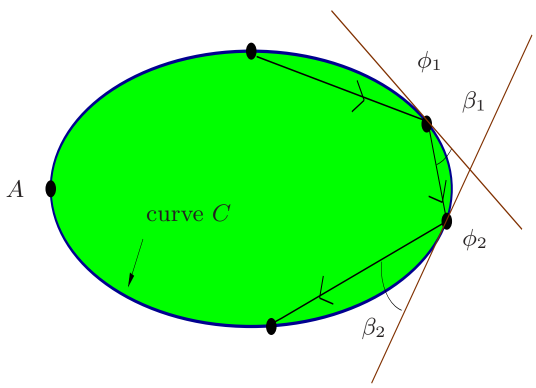
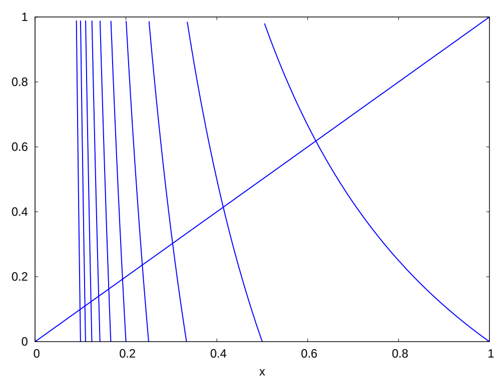


FIGURE 1.2.2. The first 610 points on orbits of the bouncing ball map (1.2.3) with $\alpha = \gamma = 1$. Both ϕ and v are taken $\bmod 2\pi$; the left figure depicts a closed curve, while the right figure depicts two closed curves (apparently).

1.2.5. The Billiard map. Let $C \subset \mathbb{R}^2$ be a smooth curve. A particle ('a billiard') moves freely and frictionally in the region bounded by C at a unit speed. At times $t_n > 0$ the particle collides with the boundary C and reflects off the boundary. The trajectory of the particle is then describable in terms of the data at each collision: an arc length ϕ_n and an angle of incidence/reflection β_n.



(a) The billiard map $(\phi_1, \beta_1) \mapsto (\phi_2, \beta_2)$; ϕ is the arc-length from A in a counter-clockwise direction and β is the counterclockwise angle from the chord to tangent line.



(b) The graph of the Gauss map.

FIGURE 1.2.3.

1.2.6. The Gauss map. Let $u \geq 0$ and let $[u]$ be the greatest integer that is no larger than u and let $\{u\} = u - [u]$. Colloquially, $[u]$ is the integer part of u and $\{u\}$ is its fractional or decimal part. Define

$$g(x) = \begin{cases} \{1/x\} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases} \tag{1.2.4}$$

which is called the Gauss map. The DS $x_{n+1} = g(x_n)$ can be understood from the perspective of continued fractions. `Worksheet 1`

1.2.7. Finite-difference schemes. Difference equations arise when finite-difference approximations of differential equations are made, e.g., in order to solve the differential equations numerically. Using the (forward) Euler scheme, for instance, the differential equation $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ is approximated by the difference equation

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{f}(\mathbf{x}_n),$$

with $t = nh$.

1.3. Orbits, fixed points and periodic orbits

The *orbit* of a discrete dynamical system is the sequence of its successive iterates $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots)$, with $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$. Note that

$$\mathbf{x}_n = \mathbf{F}(\mathbf{F}(\cdots \mathbf{F}(\mathbf{x}_0) \cdots)) = \mathbf{F}^n(\mathbf{x}_0),$$

where $\mathbf{F}^n$ is the nth iterate of the map, i.e., the composition of the map with itself n times. The study of discrete dynamical systems can thus be regarded as the study of multiple composition of functions.

Definition 1.3.1.

A *fixed point* $\mathbf{x}^*$ of a map satisfies

$$\mathbf{x}^* = \mathbf{F}(\mathbf{x}^*).$$

A *periodic point* $\mathbf{x}^*$ of period m is a fixed point of $\mathbf{F}^m$.

The orbit of a fixed point $\mathbf{x}_0 = \mathbf{x}^*$ is the the trivial orbit $(\mathbf{x}_0, \mathbf{x}_0, \cdots)$. An orbit $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \cdots)$ is periodic with period m iff $\mathbf{x}_m = \mathbf{x}_0$; the iterates thus repeat after m iterations, so that the orbit has the form

$$(\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{m-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{m-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_{m-1}, \mathbf{x}_0, \mathbf{x}_1, \cdots,).$$

Clearly, each point on this *periodic orbit* satisfies $\mathbf{F}^m(\mathbf{x}_n) = \mathbf{x}_n$, $n = 0, 1, \cdots, m-1$; i.e., each point is a periodic point of period m, too. One says that m is a period of the orbit; the smallest positive period of an orbit is called its *prime period* m. The term m-cycle is also employed to designate period-m orbits.

The concepts just defined for discrete DS have the following counterparts for continuous dynamical systems: A *trajectory* is a function $\mathbf{x}(t)$ satisfying the ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. A *fixed point* $\mathbf{x}^*$ is a point satisfying

$$\mathbf{f}(\mathbf{x}^*) = \mathbf{0}\,;$$

fixed points represent *equilibrium solutions* and are such that if $\mathbf{x} = \mathbf{x}^*$ initially then $\mathbf{x}(t) = \mathbf{x}^*, \forall t$. A trajectory is *periodic* if $\exists T > 0$ such that $\mathbf{x}(t+T) = \mathbf{x}(t)$.

1.4. Counting periodic points of a map

Let $\mathbf{F} : X \to X$ be a map and let $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ be the natural dynamical system. For $m \geq 1$, let P_m (resp. p_m) be the number of periodic points (resp. prime periodic points) of period m.

A periodic point of period m has a unique prime period $d \leq m$. Moreover, d must divide m. Therefore:

$$P_m = \sum_{d\,:\,d|m} p_d, \tag{1.4.1}$$

that is to say: the number of periodic orbits of period m is the sum of the number of prime periodic orbits of period d as d ranges over all divisors of m (including 1 and m itself).

1.4.1. Objectives. When studying DS, the main interest lies in qualitative features: fixed points, periodic orbits, stability, etc. The long-time behaviour of the solutions, i.e., the behaviour for large n or t, is particularly interesting. Our attention is generally on properties that are independent of the choice of a coordinate system; this leads to idea of equivalence: all DS that can be transformed into one another by a smooth change of variables are regarded as equivalent.

1.5. Exercises

(i) Prove that if x is a periodic point of period m, then its prime period d divides m. [Hint: use *reductio ad absurdum*.]
(ii) Determine the fixed points of the logistic map, equation (1.2.2). Determine its period-2 periodic points. How many prime period-2 points does it have? [The answer depends on μ.]
(iii) Determine the fixed points of the standard map, equation (1.2.3). What about its period-2 periodic points?
(iv) Let $C \subset \mathbb{R}^2$ be a closed, convex, smooth curve. Let L be a chord of C whose length is maximal amongst all chords. Let the two points of $C \cap L$ have arc lengths $\{\phi_0, \phi_1\}$. Show that the billiard map has a period-2 orbit $\{(\phi_0, \pi/2), (\phi_1, \pi/2)\}$.

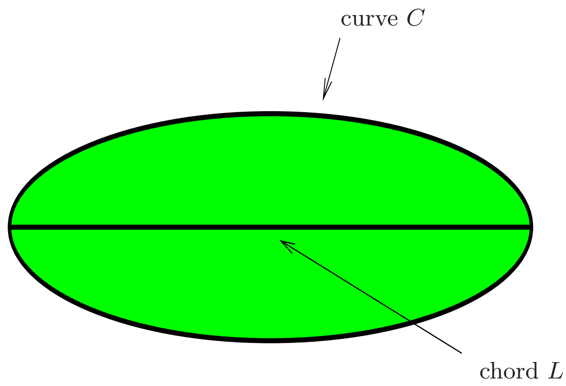


FIGURE 1.5.4. A period-2 point of a billiard map.

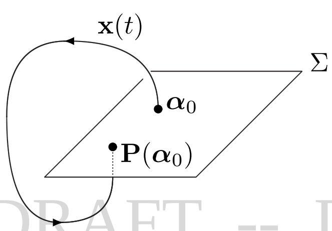
(v) Show that if (ϕ, β) is a period-2 point of a billiard map, then $\beta = \pi/2$.
(vi) If the curve C is an ellipse, can you find a prime period-4 periodic point? Can you generalise this construction?
(vii) Determine the fixed points of the Gauss map, equation (1.2.4).
(viii) The map $f : S^1 \to S^1$ defined by $f(z) = z^2$ has how many periodic points of period n? How many prime period-n periodic points does it have?
(ix) Apply the Euler scheme to integrate the ODE

$$\dot{x} = 50x(1 - x) \tag{1.5.1}$$

with a step size of h. What are the fixed points of (1.5.1)? Does it have any non-trivial periodic points? What about the Euler scheme?

1.6. Poincaré maps

The finite-differences schemes give a first connection between continuous and discrete DS. A second is provided by the concept of the *Poincaré map* (or first return map). The idea behind Poincaré maps is the conversion of a continuous DS into a discrete one in a space of reduced dimensionality. We first discuss Poincaré maps for autonomous DS before considering the particular case of systems with periodic time dependence.



1.6.1. Autonomous systems. Consider the continuous system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in a state space M of dimension d and let $\Sigma \subset M$ denote a surface of dimension $d - 1$. To construct the Poincaré map the trajectories $\mathbf{x}(t)$ that start on Σ are followed. Let $\boldsymbol{\alpha}_0 \in \Sigma$ be a starting point, i.e., $\boldsymbol{\alpha}_0 = \mathbf{x}(0)$, and let $\boldsymbol{\alpha}_1$ be the point where the trajectory $\mathbf{x}(t)$, $t > 0$ crosses Σ for the first time in the same sense. The Poincaré map $\mathbf{P} : \Sigma \to \Sigma$ is then defined by $\boldsymbol{\alpha}_1 = \mathbf{P}(\boldsymbol{\alpha}_0)$. The Poincaré map can be employed to generate the sequence $\{\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n, \cdots\}$ of the successive intersections of a trajectory with Σ. By construction, $\boldsymbol{\alpha}_n = \mathbf{P}(\boldsymbol{\alpha}_{n-1}) = \mathbf{P}^n(\boldsymbol{\alpha}_0)$. (Note that the Poincaré map is not always well defined: orbits starting on Σ may not come back to Σ.)

The properties of a Poincaré map reflects the properties of the continuous-time DS from which it is constructed. In particular:

- Fixed points of $\mathbf{P}$, i.e., points $\boldsymbol{\alpha}^*$ such that $\mathbf{P}(\boldsymbol{\alpha}^*) = \boldsymbol{\alpha}^*$ are associated with periodic solutions with $\mathbf{x}(t) = \mathbf{x}(t+T) = \boldsymbol{\alpha}^*$ that cross Σ in a given sense once per period.
- More generally, periodic points of $\mathbf{P}$ are associated with periodic solutions for $\mathbf{x}(t)$.

Examples

(i) Consider the equation $\ddot{x} + 2\dot{x} + 2x = 0$, which we treat as a 2-dimensional continuous system by writing $y = \dot{x}$ in the usual way; i.e.,

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ -2y - 2x \end{pmatrix}.$$

The ODE has general solution $x = e^{-t}(A\cos t + B\sin t)$. To define the Poincaré map, take Σ to be the positive part of the x-axis. Then for $\alpha > 0$ we consider the solution with $x(0) = \alpha$, $y = \dot{x}(0) = 0$:

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \alpha e^{-t}\,(\cos t + \sin t\,) \\ -2\alpha e^{-t}\sin t \end{pmatrix};$$

the next return to the positive x-axis is when $x = e^{-2\pi}\alpha$. Hence $P(\alpha) = e^{-2\pi}\alpha$.

(ii) For nonlinear systems the Poincaré map can generally only be obtained numerically: the differential equations are solved numerically and the successive intersections of a trajectory with Σ are plotted, providing a Poincaré section. A Maple routine constructs Poincaré sections for particular continuous dynamical systems called *Hamiltonian systems*. These have an even number of dynamical variables, $\mathbf{x} = (p_i, q_i)$, $i = 1, 2, \cdots, d$ and their governing ODEs are given by

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i}, \quad i = 1, \cdots, d, \tag{1.6.1}$$

where $H = H(p_i, q_i)$ is a (scalar) function. `Worksheet 2`

1.6.2. Time-periodic systems. A DS of dimension d with explicit time dependence $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$ can be regarded as autonomous systems in a state space of dimension $d+1$ by introducing $s = t$ as an additional dependent variable and writing the differential equations

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, s) \\ \dot{s} = 1 \end{cases}$$

For time periodic systems of period T, since $\mathbf{f}(\mathbf{x}, s) = \mathbf{f}(\mathbf{x}, s+T) = \mathbf{f}(\mathbf{x}, s+2T) = \cdots$, we take $s = t \bmod T$. A natural d-dimensional section surface Σ in the extended $d+1$-dimensional state space is then a slice $s = \text{const}$. At each return on that slice, t has increased by T. Thus the Poincaré map transforms $\mathbf{x}(t)$ into $\mathbf{x}(t+T)$; explicitly we have $\mathbf{x}(t+T) = \mathbf{P}(\mathbf{x}(t))$ and in general $\mathbf{x}(t+nT) = \mathbf{P}^n(\mathbf{x}(t))$.

Examples

(i) Consider the equation $\ddot{x} - x = \cos t$, which we can write as the continuous system:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ x + \cos t \end{pmatrix}.$$

The general solution is $x = Ae^{t} + Be^{-t} - \frac{1}{2}\cos t$. A calculation shows that the Poincaré map (i.e., the map $\mathbf{P}$ such that $\mathbf{P}\,(\mathbf{x}(0)) = \mathbf{x}(2\pi)$) is given by

$$\mathbf{P}(\boldsymbol{\alpha}) = D\boldsymbol{\alpha} + \mathbf{c}$$

where the matrix D and vector $\mathbf{c}$ are given by $D = \begin{pmatrix} \cosh(2\pi) & \sinh(2\pi) \\ \sinh(2\pi) & \cosh(2\pi) \end{pmatrix}$, $\mathbf{c} = \frac{1}{2}\begin{pmatrix} \cosh(2\pi)-1 \\ \sinh(2\pi) \end{pmatrix}$.
$\mathbf{P}$ has a unique fixed point $\boldsymbol{\alpha} = \begin{pmatrix} -1/2 \\ 0 \end{pmatrix}$ corresponding to the periodic solution $x = -\frac{1}{2}\cos t$.

(ii) For nonlinear systems, a numerical solution is necessary to obtain the successive iterates of the Poincaré map. As an example, one can consider Poincaré sections for the forced pendulum, with equation $\ddot{x} + \sin(x) = k\sin(\omega t)$, where k and ω are two constant parameters. `Worksheet 2`

1.7. Exercises

(i) Verify the claims made in example i of 1.6.1.
(ii) Show that solutions to the hamiltonian ODE for H (equation 1.6.1) preserve H: $\dot{H} \equiv 0$.
(iii) Let $H = \frac{1}{2}(p_1^2 + q_1^2) + \frac{1}{6}(p_2^2 + q_2^2)$. Show that $\Sigma = \{p_1 > 0, p_2 > 0, q_1 = 0\}$ is a Poincaré section for the ODE in equation (1.6.1). Find the Poincaré map for this section.
(iv) Write the system in example ii of 1.6.2 as an autonomous DS in three variables (x, v, s), where $v = \dot{x}$ and s is time-like. Since one can view s as a variable mod $\frac{2\pi}{\omega}$, show that

$$\Sigma = \{(x, v, 0) \ : x, v \in \mathbb{R}\}$$

is a Poincaré section. Find the fixed points of the Poincaré map.

CHAPTER 2

Linear autonomous systems

2.1. Discrete systems

Let us adopt the state space $M = \mathbb{R}^d$. *Linear* autonomous discrete systems are governed by the equation

$$\mathbf{x}_{n+1} = A\,\mathbf{x}_n, \tag{2.1.1}$$

where $\mathbf{x} \in \mathbb{R}^d$ and A is a $d \times d$ real matrix. Such systems often appear when nonlinear systems are analysed locally, in the vicinity of a fixed point for example. Note that $\mathbf{x} = \mathbf{0}$ is a fixed point for (2.1.1); it is unique if 1 is not an eigenvalue of A.

The general solution to (2.1.1) takes the form

$$\mathbf{x}_n = A^n \mathbf{x}_0.$$

The nth power of A is computed efficiently using the Jordan form of A. We suppose here that A has eigenvalues λ_i $(i = 1, 2, \cdots, d)$ and is diagonalisable, so that

$$S^{-1}AS = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d) \equiv \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix},$$

where S is the matrix of the eigenvectors of A. Then we have

$$A^n = S\,\Lambda^n\,S^{-1} = S\,\mathrm{diag}(\lambda_1^n, \lambda_2^n, \ldots, \lambda_d^n)\,S^{-1}.$$

The use of the matrix S can be regarded as the definition of a variable transformation from $\mathbf{x}_n$ to a new vector $\mathbf{u}_n$, with $\mathbf{x}_n = S\mathbf{u}_n$. Since the columns of S are the (generalised) eigenvectors of A, this transformation can be interpreted as the use of the eigenvectors as new basis vectors. The evolution of $\mathbf{u}$ is governed by

$$\mathbf{u}_{n+1} = S^{-1}AS\mathbf{u}_n = \Lambda\mathbf{u}_n,$$

so that

$$\mathbf{u}_n = \Lambda^n \mathbf{u}_0 \qquad \text{with} \qquad \mathbf{u}_0 = S^{-1}\mathbf{x}_0.$$

The behaviour of the orbits $\mathbf{x}_n$ in a linear autonomous system is straightforward: this is particularly clear from the fact that they can be described by an analytical expression. Orbits generally either converge to or diverge from the origin; this can be analysed using the transformed variable $\mathbf{u}_n$ and clearly depends on the eigenvalues λ_i:

- If $|\lambda_i| < 1$, $\forall i$, then the origin is attracting; i.e., all orbits converge to the origin as $n \to +\infty$.
- If $|\lambda_i| > 1$, $\forall i$, then the origin is repelling; i.e., all orbits diverge from the origin as $n \to +\infty$. An alternative definition states that the origin is attracting for $n \to -\infty$, i.e., for *backward orbits*.
- If there are eigenvalues both with modulus greater than 1 and less than 1, the origin is a *saddle*; most orbits diverge from the origin both forward $(n \to \infty)$ and backward $(n \to -\infty)$ in time.

When the origin is a saddle, some particular orbits converge to the origin as $n \to \pm\infty$. It is useful to find which ones. Consider the case of a diagonalisable matrix A and assume that the origin is *hyperbolic*, i.e., that no eigenvalue has modulus equal to 1. Its Jordan form Λ can be written as

$$\Lambda = \begin{pmatrix} \mu_1 & & & & & \\ & \ddots & & & & \\ & & \mu_r & & & \\ & & & \sigma_1 & & \\ & & & & \ddots & \\ & & & & & \sigma_{d-r} \end{pmatrix}, \quad |\mu_i| < 1, \quad |\sigma_i| > 1.$$

Let us write $\mathbf{u}_n = \begin{pmatrix} \mathbf{u}_n^+ \\ \mathbf{u}_n^- \end{pmatrix}$ where $\mathbf{u}_n^+ \in \mathbb{R}^r$ and $\mathbf{u}_n^- \in \mathbb{R}^{d-r}$. It is clear that orbits with $\mathbf{u}_0^- = \mathbf{0}$ converge to $\mathbf{0}$ for $n \to \infty$ while those with $\mathbf{u}_0^+ = \mathbf{0}$ converge to $\mathbf{0}$ for $n \to -\infty$. These conditions can be formulated in terms of the original variable $\mathbf{x}_n = S\mathbf{u}_n$ by noting that the columns of S are the eigenvectors of A. Thus, orbits $\mathbf{x}_n$ converge to $\mathbf{0}$ forward in time provided that $\mathbf{x}_0$ is a linear combination of the r eigenvectors associated with the eigenvalues μ_i with modulus less than 1. Similarly, orbits $\mathbf{x}_n$ converge to $\mathbf{0}$ backward in time provided that $\mathbf{x}_0$ is a linear combination of the $d-r$ eigenvectors associated with the eigenvalues σ_i with modulus greater than 1. Non-zero imaginary parts for the eigenvalues indicate that the orbits are rotating around the origin. Let us define

Definition 2.1.2.

The stable subspace of a linear discrete-time DS on $\mathbb{R}^d$ is the set of $\mathbf{x} = \mathbf{x}_0$ such that $\mathbf{x}_n \to \mathbf{0}$ as $n \to \infty$.

The unstable subspace of a linear discrete-time DS on $\mathbb{R}^d$ is the set of $\mathbf{x} = \mathbf{x}_0$ such that $\mathbf{x}_n \to \mathbf{0}$ as $n \to -\infty$.

It is a consequence of the discussion above that one can describe the stable and unstable subspaces entirely in terms of the spectrum of the linear DS:

Theorem 2.1.

The stable subspace $E^+ \subset \mathbb{R}^d$ as the subspace spanned by the eigenvectors associated with eigenvalues μ_i with $|\mu_i| < 1$.

The unstable subspace $E^- \subset \mathbb{R}^d$ as the subspace spanned by the eigenvectors associated with eigenvalues σ_i with $|\sigma_i| > 1$.

The subspaces E^+ and E^- are sometimes denoted by E^s and E^u; they are also called forward contracting subspace and backward contracting subspace, respectively. The hyperbolicity requirement ensures that $E^+ \oplus E^- = \mathbb{R}^d$. Note that E^+ and E^- are *invariant* subspaces; i.e., if $\mathbf{x}_0 \in E^+$ then $\mathbf{x}_n \in E^+ \ \forall n$ and, similarly, if $\mathbf{x}_0 \in E^-$ then $\mathbf{x}_n \in E^- \ \forall n$ (can you prove it?).

Example. Consider the two-dimensional system

$$\mathbf{x}_{n+1} = A\mathbf{x}_n = \begin{pmatrix} 3/2 & 1 \\ 1 & 0 \end{pmatrix} \mathbf{x}_n,$$

where $\mathbf{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$. The eigenvalues of A are found to be $-1/2$ and 2, so $\mathbf{0}$ is a hyperbolic saddle; corresponding eigenvectors are

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$

respectively. It follows that

$$S = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}, \quad S^{-1} = \frac{1}{5} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

Introduce new variables $\mathbf{u}_n$ decomposed in u_n^+ and u_n^- according to

$$\mathbf{u}_n = \begin{pmatrix} u_n^+ \\ u_n^- \end{pmatrix} = S^{-1} \begin{pmatrix} x_n \\ y_n \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} = \frac{1}{5} \begin{pmatrix} x_n - 2y_n \\ 2x_n + y_n \end{pmatrix}.$$

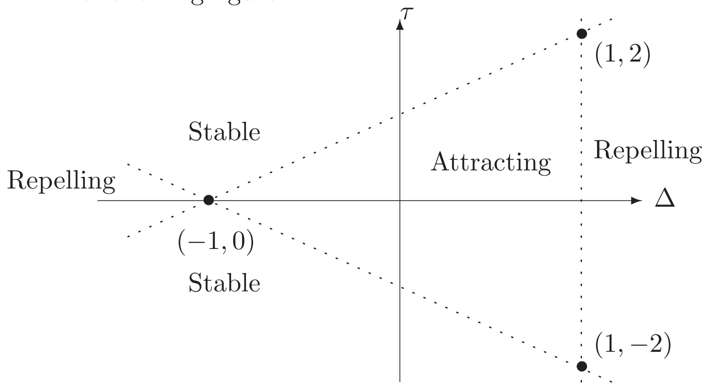
The stable subspace E^+, spanned by $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ corresponds to $u^- = 0$ and so has equation $2x + y = 0$.
The unstable subspace E^-, spanned by $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, corresponds to $u^+ = 0$ and so has equation $x - 2y = 0$.

2.2. Classification of fixed points in two dimensions

In two dimensions, the properties of the origin are easily classified. We consider the generic possibilities for which the two eigenvalues λ_i satisfy $|\lambda_i| \neq 1$. These are:

- $|\lambda_1| > 1$ and $|\lambda_2| > 1$: the origin is a repelling node;
- $|\lambda_1| < 1$ and $|\lambda_2| < 1$: the origin is an attracting node;
- $|\lambda_1| < 1$ and $|\lambda_2| > 1$ (or $|\lambda_1| > 1$ and $|\lambda_2| < 1$) : the origin is a saddle;
- $\lambda_1 = \overline{\lambda}_2 \in \mathbb{C}$: the origin is a *centre*, which is repelling if $|\lambda_1| > 1$ and attracting if $|\lambda_1| < 1$.

It is sometimes convenient to classify the properties of the origin just described using the trace τ and determinant Δ of the matrix A. The eigenvalues are related to them through the characteristic equation $\lambda^2 - \tau\lambda + \Delta = 0$. The regions of the (Δ, τ)-plane for which the origin is attracting, repelling, or a saddle are shown in the following figure [1].



The distinction between centres and nodes can also be worked out.

2.3. Comparison with continuous systems

The solution of the linear autonomous system $\dot{\mathbf{x}} = A\mathbf{x}$ with $\mathbf{x}(0) = \mathbf{x}_0$ is written in terms of the exponential of the matrix A as[2]

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0.$$

To compute the exponential, one can use the diagonalisation of A, which can again be regarded as a variable transformation from $\mathbf{x}$ to a new vector $\mathbf{u}$, with $\mathbf{x} = S\mathbf{u}$. The evolution of $\mathbf{u}$ is governed by $\dot{\mathbf{u}} = S^{-1}AS\mathbf{u} = \Lambda\mathbf{u}$, so that $\mathbf{u} = \exp(\Lambda t)\mathbf{u}_0$, with $\mathbf{u}_0 = S^{-1}\mathbf{x}_0$. The behaviour of the origin is then seen to depend on the real part of the eigenvalues:

- If $\mathrm{Re}\,\lambda_i < 0$, $\forall i$, then the origin is attracting; i.e., all trajectories converge to the origin as $t \to +\infty$.
- If $\mathrm{Re}\,\lambda_i > 0$, $\forall i$, then the origin is repelling; i.e., all trajectories diverge from the origin as $t \to +\infty$. An alternative definition states that the origin is attracting for $t \to -\infty$, i.e. for *backward trajectories*.
- If there are eigenvalues with positive and negative real parts, the origin is a *saddle*; most trajectories diverge from the origin both forward ($t \to \infty$) and backward ($t \to -\infty$).

In analogy with the definition for discrete systems, we define:

---

[1] See Hand-in Assignment 1, Question 2.

[2] Recall: The exponential of a $n \times n$ matrix Q is defined as $e^Q = \sum_{j=0}^{\infty} \frac{1}{j!} Q^j = I + Q + \frac{1}{2}Q^2 + \cdots$. If $Q = \mathrm{diag}\,(\alpha_1, \alpha_2, \ldots, \alpha_n)$ then $e^Q = \mathrm{diag}\,(e^{\alpha_1}, e^{\alpha_2}, \ldots, e^{\alpha_n})$. If $U = VWV^{-1}$ then $e^U = Ve^WV^{-1}$ where U, V and W are $n \times n$ matrices.

Definition 2.3.3.

The stable subspace of a linear continuous-time DS on $\mathbb{R}^d$ is the set of $\mathbf{x} = \mathbf{x}_0$ such that $\mathbf{x}(t) \to \mathbf{0}$ as $t \to \infty$.

The unstable subspace of a linear continuous-time DS on $\mathbb{R}^d$ is the set of $\mathbf{x} = \mathbf{x}_0$ such that $\mathbf{x}(t) \to \mathbf{0}$ as $t \to -\infty$.

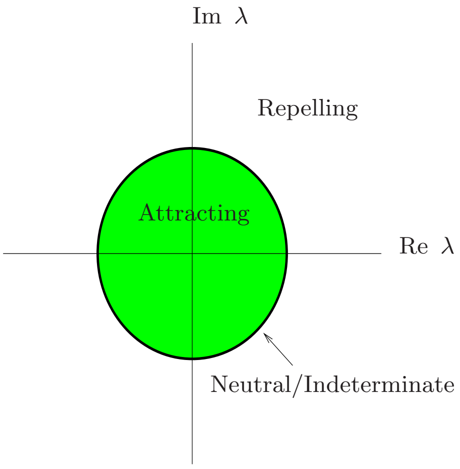
One can describe the stable and unstable subspaces entirely in terms of the spectrum of the linear DS:

Theorem 2.2.

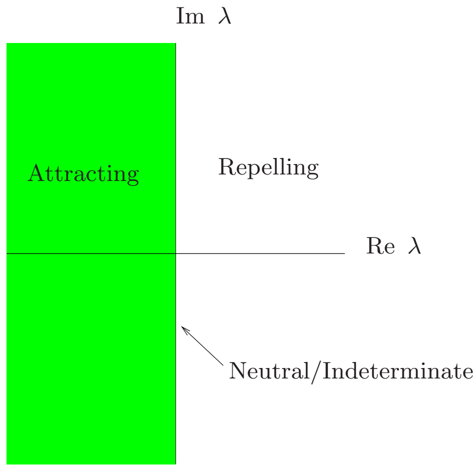
The stable subspace $E^+ \subset \mathbb{R}^d$ is the subspace spanned by the eigenvectors associated with eigenvalues with negative real parts.

The unstable subspace $E^- \subset \mathbb{R}^d$ is the subspace spanned by the eigenvectors associated with eigenvalues with positive real parts.

The comparison between the behaviour associated with a single eigenvalue λ in discrete and continuous systems is summarised by the figure 2.3.1 below.



(a) Discrete time



(b) Continuous time

FIGURE 2.3.1. Stability in discrete and continuous time.

2.4. Exercises

(i) Consider the continous-time linear DS

$$\dot{\mathbf{x}} = A\mathbf{x}, \qquad \mathbf{x} \in \mathbb{R}^3. \tag{2.4.1}$$

If the trace of A is zero and the determinant of A is 1, show that the origin is a hyperbolic saddle.

(ii) Sketch the phase portrait of

$$\mathbf{x}_{n+1} = A\mathbf{x}_n, \qquad A = \begin{bmatrix} 3/2 & 1 \\ 1 & 0 \end{bmatrix}. \tag{2.4.2}$$

(iii) Sketch the phase portrait of the linear system

$$\mathbf{x}_{n+1} = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 4 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \mathbf{x}_n. \tag{2.4.3}$$

(iv) Sketch the phase portrait of the linear system

$$\dot{\mathbf{x}} = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 4 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \mathbf{x}. \tag{2.4.4}$$

CHAPTER 3

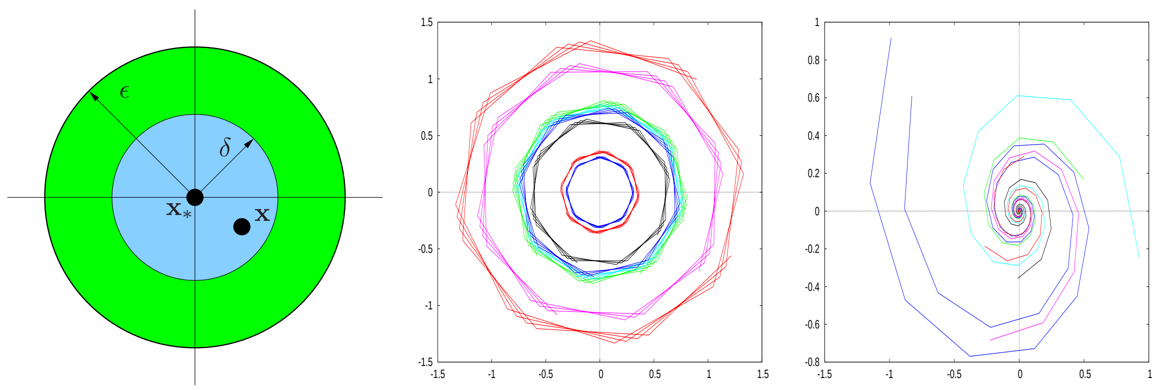
Fixed points in nonlinear systems

3.1. Stability

The results of the previous section can be applied to study the dynamics of nonlinear systems in the vicinity of fixed points of mappings. Consider the nonlinear system $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ where $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{F} : \mathbb{R}^d \to \mathbb{R}^d$. Of particular interest is the *stability* of fixed points. A fixed point $\mathbf{x}_* = \mathbf{F}(\mathbf{x}_*)$ is stable provided that orbits that start close enough to $\mathbf{x}_*$ stay close to $\mathbf{x}_*$ for all time. Formally,

Definition 3.1.4.

$\mathbf{x}_*$ is *stable* if $\forall \epsilon > 0, \exists \delta > 0$ such that $|\mathbf{x}_0 - \mathbf{x}_*| < \delta \Rightarrow |\mathbf{x}_n - \mathbf{x}_*| < \epsilon, \forall n > 0$.
$\mathbf{x}_*$ is *asymptotically stable* if there is a neighbourhood U of $\mathbf{x}_*$ such that $\mathbf{x}_0 \in U$ implies that $\mathbf{x}_n \to \mathbf{x}_*$ as $n \to \infty$.



(a) Definition of Stability (b) An Example of Stability (c) Asymptotic Stability: $\mathbf{x}_n \to \mathbf{x}_*$ as $n \to \infty$

FIGURE 3.1.1. (a) Given $\epsilon > 0$, there must be a $\delta > 0$ such that the orbit of any point $\mathbf{x}$ within δ of $\mathbf{x}_*$ is contained within a distance ϵ of $\mathbf{x}_*$. (b-c) Asymptotic stability implies stability but not vice versa.

To study stability, we consider the system locally, in the vicinity of $\mathbf{x}_*$. We introduce the new vector variable $\boldsymbol{\eta}_n$, with $\mathbf{x}_n = \mathbf{x}_* + \boldsymbol{\eta}_n$, where $|\boldsymbol{\eta}_n|$ is assumed to be small. We have

$$\mathbf{x}_{n+1} = \mathbf{x}_* + \boldsymbol{\eta}_{n+1} = \mathbf{F}(\mathbf{x}_* + \boldsymbol{\eta}_n).$$

Assuming that $\mathbf{F}$ is twice continuously differentiable, we expand it as a Taylor series:[1]

$$\mathbf{F}(\mathbf{x}_* + \boldsymbol{\eta}_n) = \mathbf{F}(\mathbf{x}_*) + A\,\boldsymbol{\eta}_n + O(|\boldsymbol{\eta}_n|^2)$$

where $A = D\mathbf{F}|_{\mathbf{x}_*}$ is the matrix of the derivatives of $\mathbf{F}$ (i.e., the Jacobian matrix) evaluated at $\mathbf{x}_*$. Thereby, we see that

$$\boldsymbol{\eta}_{n+1} = A\,\boldsymbol{\eta}_n + O(|\boldsymbol{\eta}_n|^2), \qquad \text{where } A = D\mathbf{F}|_{\mathbf{x}_*}. \tag{3.1.1}$$

[1] A function $f(x) = o(x)$ if $f(x)/x \to 0$ as $x \to 0$. We say that "f is *little-o of x*." A function $f(x) = O(x^2)$ if $f(x)/x^2 < Constant$ for all x sufficiently small and we say that "f is *big-O of* x^2." For example $x^{3/2} = o(x)$ but not $O(x^2)$, while $x^2 = o(x)$ and $x^2 = O(x^2)$. If $\mathbf{F}$ is continuously differentiable, then $\mathbf{F}(\mathbf{x}_* + \boldsymbol{\eta}) = \mathbf{x}_* + A\boldsymbol{\eta} + o(\boldsymbol{\eta})$, but if $\mathbf{F}$ is twice continuously differentiable, then the $o(|\boldsymbol{\eta}|)$ term is $O(|\boldsymbol{\eta}|^2)$. The following discussion is true for the continuously differentiable case, but is simplified if $\mathbf{F}$ is assumed to be twice-continuously differentiable.

That is, for $\boldsymbol{\eta}_n$ sufficiently small the behaviour of the nonlinear system approximates to that of the corresponding linear system when terms of $O(|\boldsymbol{\eta}_n|^2)$ are neglected. We conclude that

- If $\mathbf{x}_*$ is attracting (i.e., if all the eigenvalues of A have a modulus less than 1), then $\mathbf{x}_*$ is a stable fixed point for the nonlinear system.
- If $\mathbf{x}_*$ is repelling or is a saddle (i.e., there is at least one eigenvalue of A with modulus greater than 1), then $\mathbf{x}_*$ is an unstable fixed point for the nonlinear system.

Remarks:

- If the eigenvalues of A have a modulus equal to 1, then one cannot conclude about stability from the linearised system (3.1.1). Higher-order terms in the Taylor series must be taken into account (see §3.3 below).
- Saddle points are unstable because some (in fact, most) orbits leave any neighbourhood of the fixed point as $n \to \infty$. However, as in the case for linear systems, not all orbits leave the fixed point for $n \to \infty$; similarly not all backward orbits leave the fixed point as $n \to -\infty$. The next section deals with the study of such orbits.

Example. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(\mathbf{x}) = \begin{bmatrix} x^2\, y - \dfrac{7\, y}{4} + \dfrac{x}{4} \\ y^2 + \dfrac{3\, y}{4} + x^2 - \dfrac{x}{2} \end{bmatrix}$$

where $\mathbf{x} = [x; y]$. Since there are no constant terms, $\mathbf{x}_{*,0} = [0; 0]$ is a fixed point. Some additional calculations show that there is a single additional real fixed point at approximately $\mathbf{x}_{*,1} \cong [.37449; .05816]$. One computes that

$$DF = \begin{bmatrix} \dfrac{1}{4} + 2\, x\, y & x^2 - \dfrac{7}{4} \\ 2\, x - \dfrac{1}{2} & \dfrac{3}{4} + 2\, y \end{bmatrix}$$

whence

$$DF_{[0;0]} = \begin{bmatrix} \dfrac{1}{4} & -\dfrac{7}{4} \\ -\dfrac{1}{2} & \dfrac{3}{4} \end{bmatrix} \qquad\qquad DF_{\mathbf{x}_{*,1}} \cong \begin{bmatrix} .29356 & -1.60975 \\ .24898 & .86631 \end{bmatrix}$$

which have eigenvalues

$$\frac{2 \pm \sqrt{15}}{4} \cong -.46825, 1.4682 \qquad\qquad .57994 \pm .56461i.$$

One sees that the origin is a hyperbolic fixed point with both stable and unstable directions, so it is not a stable fixed point. On the other hand, the eigenvalues at $\mathbf{x}_{*,1}$ have modulus ~ 0.80939, so this fixed point is stable and indeed asymptotically stable.

Exercise 3.1.1.

(i) Let $\mathbf{F} : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

(3.1.2)
$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} \dfrac{1}{2}x - \dfrac{3}{4}y + xy^2 \\ \dfrac{3}{4}x - \dfrac{1}{2}y + x^2 + y^2 \end{bmatrix}$$

where $\mathbf{x} = [x; y]$. Is $\mathbf{x}_* = [0; 0]$ a stable fixed point? Asymptotically stable? **Solution.** Yes, to both. The eigenvalues of $d\mathbf{F}_{\mathbf{x}_*}$ are $\pm\frac{sqrt{-5}}{4}$ which have modulus < 1.

(ii) Determine all the real fixed points of the map $\mathbf{F}$. Determine their stability properties.

(iii) Identify the complex numbers with $\mathbb{R}^2$ via $z \mapsto [x; y]$ where $z = x + iy$. Define the map f by

(3.1.3)
$$f(z) = \exp(z).$$

Find the fixed points of f and determine if they are stable.

3.2. Stable and unstable manifolds

We consider a (hyperbolic) saddle at $\mathbf{0}$. (Any fixed point $\mathbf{x}_*$ can be translated to $\mathbf{0}$ by a change of coordinates, so there is no loss of generality involved in assuming that the fixed point lies at the origin of our coordinate system). As in §2.1, $A = D\mathbf{F}|_{\mathbf{0}}$ is assumed to have r eigenvalues μ_i with $|\mu_i| < 1$ and $d - r$ eigenvalues σ_i with $|\sigma_i| > 1$. For simplicity we assume that all the eigenvalues are real and distinct.

The dynamics in the vicinity of the origin are governed by the equation

$$\mathbf{x}_{n+1} = A\mathbf{x}_n + O(|\mathbf{x}_n|^2), \qquad \text{where } A = D\mathbf{F}|_{\mathbf{0}}$$

obtained by Taylor expansion. Using the variable $\mathbf{u}_n = S^{-1}\mathbf{x}_n$, i.e., using the eigenvectors of A as a basis, the matrix can take its Jordan form. Since $\mathbf{0}$ is a hyperbolic saddle, we can use the decomposition $\mathbf{u}_n = \begin{pmatrix} \mathbf{u}_n^+ \\ \mathbf{u}_n^- \end{pmatrix}$, where $\mathbf{u}_n^+ \in \mathbb{R}^r$ is associated with the eigenvalues with absolute values less than 1 and $\mathbf{u}_n^- \in \mathbb{R}^{d-r}$ is associated with the eigenvalues with absolute values greater than 1. If we write the diagonal matrix $\Lambda = S^{-1}AS$ in block diagonal form (see p. 8 of notes)

$$\Lambda = \begin{pmatrix} \mu & 0 \\ 0 & \sigma \end{pmatrix},$$

where $\mu = \text{diag}(\mu_1, \ldots, \mu_r)$ and $\lambda = \text{diag}(\lambda_1, \ldots, \lambda_{d-r})$, then the equation $\mathbf{u}_{n+1} = \Lambda\mathbf{u}_n + O(|\mathbf{u}_n^2|)$ becomes

$$\begin{aligned} \mathbf{u}_{n+1}^+ &= \mu\mathbf{u}_n^+ + O(|\mathbf{u}_n|^2), \\ \mathbf{u}_{n+1}^- &= \sigma\mathbf{u}_n^- + O(|\mathbf{u}_n|^2). \end{aligned}$$

For the linearised system, the stable and unstable subspaces E^+ and E^-, which are defined by $\mathbf{u}_n^- = \mathbf{0}$ and $\mathbf{u}_n^+ = \mathbf{0}$, respectively, contain the orbits converging to $\mathbf{0}$ as $n \to \pm\infty$. How does this change when the (small) nonlinear terms are taken into account?

The *stable manifold theorem* asserts the existence of a local stable manifold W_{loc}^+ in some neighbourhood of $\mathbf{0}$ such that orbits starting on W_{loc}^+ approach $\mathbf{0}$ for $n \to +\infty$. The local stable manifold W_{loc}^+ is a graph tangent to E^+ at $\mathbf{0}$; i.e., it is given by

$$\mathbf{u}^- = \mathbf{g}(\mathbf{u}^+) = O(|\mathbf{u}^+|^2),$$

where $\mathbf{g}$ is a map from E^+ to E^-. Furthermore, W_{loc}^+ is *invariant*: this means that $\mathbf{F}(W_{loc}^+) \subseteq W_{loc}^+$ or, equivalently, if $\mathbf{x}_0 \in W_{loc}^+$, then $\mathbf{x}_n \in W_{loc}^+$ for all $n \geq 0$.

The global stable manifold W^+ is the set of all $\mathbf{x}_0$ such that $\mathbf{x}_n \to \mathbf{0}$ as $n \to \infty$. It is a consequence of the stable manifold theorem that $W^+ = \cup_{n\geq 0}\mathbf{F}^{-n}(W_{loc}^+)$. The global stable manifold is often a very complicated object.

Similarly, there exists a local *unstable manifold* W_{loc}^- in some neighbourhood of $\mathbf{0}$ such that orbits starting on W_{loc}^- approach $\mathbf{0}$ for $n \to -\infty$. The unstable manifold W_{loc}^- is a graph tangent to E^- at $\mathbf{0}$, i.e., it is given by

$$\mathbf{u}^+ = \mathbf{h}(\mathbf{u}^-) = O(|\mathbf{u}^-|^2),$$

where $\mathbf{h}$ is a map from E^- to E^+. The global unstable manifold W^- is also invariant.

The local stable and unstable manifolds can be computed perturbatively in the vicinity of $\mathbf{0}$, by expanding $\mathbf{F}$ and the functions $\mathbf{g}$ and $\mathbf{h}$ in power series. As an example, consider the two-dimensional case. The stable manifold is given by $u^- = g(u^+)$; it is found by expanding g according to

$$u^- = g(u^+) = a_2(u^+)^2 + a_3(u^+)^3 + a_4(u^+)^4 + \cdots, \tag{3.2.1}$$

where a_2, a_3, a_4, …, are constant coefficients to be determined. From (3.2.1), u_{n+1}^- is written as

$$u_{n+1}^- = a_2(u_{n+1}^+)^2 + a_3(u_{n+1}^+)^3 + a_4(u_{n+1}^+)^4 + \cdots.$$

The iterates $n+1$ on each side can be expressed in terms of (u_n^+, u_n^-) using the mapping $\mathbf{F}$; furthermore u_n^- can be eliminated in favour of u_n^+ using the stable manifold equation (3.2.1). This leads to a equation between two polynomials of u_n^+. Identifying the coefficients of the various powers of u_n^+ leads to equations for $a_2, a_3, a_4, \ldots$, and thus to the determination of an approximation of the stable manifold. Carrying out this calculation by hand to orders beyond the second or third order in u_n^+ becomes tedious; it is however relatively straightforward to implement the procedure in Maple. `Worksheet 3`

Example. Consider the $\mathbb{R}^2$ system

$$\left.\begin{aligned} x_{n+1} &= 2x_n \qquad - y_n^2 \\ y_{n+1} &= \qquad -\tfrac{1}{2}y_n + x_n^2 \end{aligned}\right\} \qquad (a)$$

which has a fixed point at the origin. Since the fixed point is at the origin, the corresponding linear system is given by dropping the non-linear terms

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix},$$

which is already in diagonal form so $u^- = x$ and $u^+ = y$. [That is, the stable subspace E^+ is given by the line $x = 0$ while the unstable subspace E^- is given by the line $y = 0$.] The stable manifold W^+ in the neighbourhood of the origin is given by

$$x = g(y) = a_2y^2 + a_3y^3 + \cdots . \qquad (b)$$

To compute the coefficients, we exploit the invariance of W^+. Assume that $(x_n, y_n) \in W^+$, so $x_n = g(y_n)$. On the one hand, since the stable manifold is invariant $(x_{n+1}, y_{n+1}) \in W^+$, i.e. $x_{n+1} = g(y_{n+1})$. This yields

$$\begin{aligned} x_{n+1} &\overset{(b)}{=} a_2y_{n+1}^2 + a_3y_{n+1}^3 + \cdots \\ &\overset{(a)}{=} a_2(-\frac{1}{2}y_n + x_n^2)^2 + a_3(-\frac{1}{2}y_n + x_n^2)^3 + \cdots \\ &\overset{(b)}{=} a_2\left[-\frac{1}{2}y_n + (a_2y_n^2 + a_3y_n^3 + \cdots)^2\right]^2 + a_3\left[-\frac{1}{2}y_n + (a_2y_n^2 + a_3y_n^3 + \cdots)^2\right]^3 + \cdots \\ &= \frac{1}{4}a_2y_n^2 - \frac{1}{8}a_3y_n^3 + \cdots . \end{aligned}$$

On the other hand, invariance of W^+ means that $x_{n+1} = 2g(y_n) - y_n^2$ by (a), so

$$x_{n+1} = 2(a_2y_n^2 + a_3y_n^3 + \cdots) - y_n^2 = (2a_2 - 1)y_n^2 + 2a_3y_n^3 + \cdots .$$

Comparing the coefficients of powers of y_n in the previous two lines, we find

$$\frac{1}{4}a_2 = 2a_2 - 1 \qquad \text{i.e.,} \quad a_2 = \frac{4}{7},$$
$$-\frac{1}{8}a_3 = 2a_3 \qquad \text{i.e.,} \quad a_3 = 0.$$

Hence, we find that the stable manifold W^+ is approximately given by the parabola $x = \frac{4}{7}y^2$ (correct to third-order).

Similarly, the unstable manifold W^- in the neighbourhood of the origin is given by

$$y = h(x) = b_2x^2 + b_3x^3 + \cdots .$$

It is left as an exercise for the reader to show that the unstable manifold W^- is approximately given by the parabola $y = \frac{2}{9}x^2$ (correct to third-order). `Worksheet 3`

3.2.1. Why do these calculations work? Let us consider the problem of determining the function $\mathbf{u}^- = \mathbf{g}(\mathbf{u}^+)$ whose graph is the stable manifold. We will focus on the two-dimensional case, where the stable manifold is a real-valued function of one variable. The higher-dimensional case is not conceptually more difficult, but its does require some additional book-keeping.

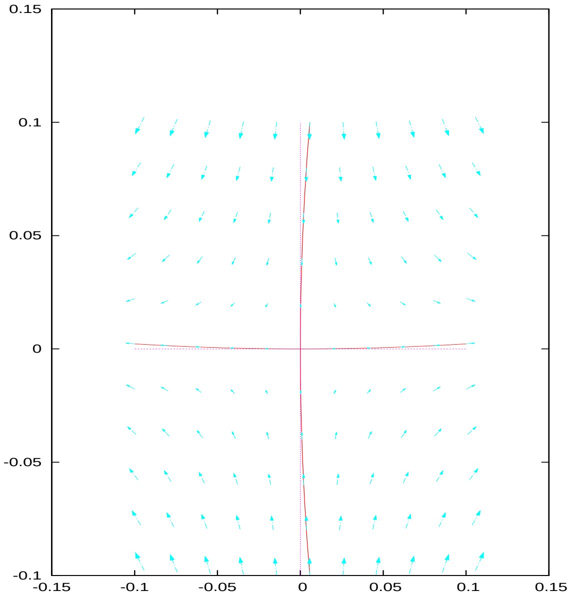
To simplify notation, let $x = u^+$ (resp. $y = u^-$) be the coordinate on the stable (resp. unstable) subspace. Let the dynamical system be $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ where $\mathbf{x} = [x,\ y]$ and $\mathbf{F}$ is defined by

$$x_{n+1} = \mu x_n + a(x_n, y_n),$$
$$y_{n+1} = \sigma y_n + b(x_n, y_n),$$

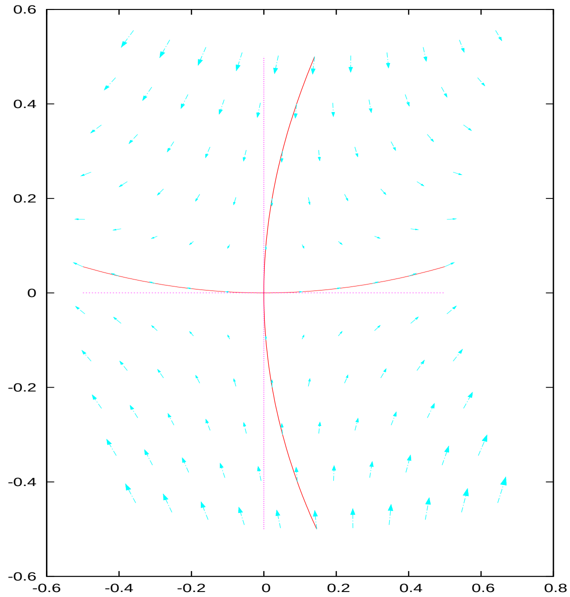
where $|\mu| < 1$, $|\sigma| > 1$ and a, b are analytic functions which vanish to second order at the origin. Assume that the stable manifold W^+ is the graph $y = g(x)$. The invariance of the stable manifold means that if $(x_n, y_n) \in W^+$, then $(x_{n+1}, y_{n+1}) \in W^+$. We can compute y_{n+1} in two ways:

first: $\mathbf{x}_n, \mathbf{x}_{n+1} \in W^+$ implies

$$y_{n+1} = g(x_{n+1}) = g(\mu x_n + a(x_n, g(x_n))), \qquad (*)$$



(a) In the small



(b) In the large

FIGURE 3.2.2. A plot of the stable ($x = \frac{4}{7}y^2 + \cdots$) and unstable ($y = \frac{2}{9}x^2 + \cdots$) manifolds. The arrows indicate the direction that a point is mapped; they should be tangent to the manifolds.

second: $\mathbf{x}_n \in W^+$ and $\mathbf{x}_{n+1} = \mathbf{F}(x_n)$ implies

$$y_{n+1} = \sigma y_n + b(x_n, y_n) = \sigma g(x_n) + b(x_n, g(x_n)). \qquad (**)$$

From these two equations, we get

$$g(\mu x + a(x, g(x))) = \sigma g(x) + b(x, g(x)) \qquad (***)$$

for all x in a sufficiently small neighbourhood of $x = 0$. Let us write each of a, b and g as a Maclaurin polynomial of degree N, and group the homogeneous bits of each function together, so we can write

$$g(x) = \sum_{k=2}^{N} g_k x^k + O(x^{N+1}), \qquad a(x,y) = \sum_{k=2}^{N} a_k(x,y) + O(|x,y|^{N+1}),$$
$$b(x,y) = \sum_{k=2}^{N} b_k(x,y) + O(|x,y|^{N+1}).$$

The composition of a homogeneous polynomial of degree d with one of degree d' yields a homogeneous polynomial of degree dd'. This is an elementary observation, which is used below repeatedly.

Clearly, if (***) is true, then the degree k part of each side must be equal for $k = 0, \ldots, N$. Conversely, if the degree k part of each side of (***) are equal for $k = 0, \ldots, N$, then the equation is satisfied up to a remainder of order x^{N+1}. Since a, b and g vanish to second order at 0, the first non-trivial homogeneous bit of (***) is the degree 2 bit:

$$g_2 \cdot (\mu x)^2 = \sigma g_2 \cdot x^2 + b_2(x, 0) \quad \Longrightarrow \quad g_2 \cdot (\mu^2 - \sigma) = p_2, \tag{$***_2$}$$

where $p_2 x^2 = b_2(x, 0)$. Equation (***$_2$) has a solution for g_2 since $|\mu|^2 < 1$, $|\sigma| > 1$.

In general, for the degree N bit of (***), we get an equation similar to (***$_2$), namely

$$g_N \cdot (\mu^N - \sigma) = p_N \tag{$***_N$}$$

where p_N depends on a_k, b_k for $k \le N$ and g_k for $k \le N-1$. If we assume that (***$_k$) has a solution for g_k for $k = 2, \ldots, N-1$, then it is clear that (***$_N$) has a solution.

By induction, there is a formal power series for g that formally solves (***). Since a, b are assumed to be analytic, the stable manifold theorem says that this formal power series actually converges on a sufficiently small interval around 0.

Exercise 3.2.2.

(i) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the map

(3.2.2)
$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -\frac{1}{2}x + xy^2 \\ 3y + x^2 \end{bmatrix}.$$

Compute the Taylor series expansions for the stable and unstable manifolds of the hyperbolic fixed point $(x, y) = (0, 0)$ up to degree 3.

(ii) Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be the map

(3.2.3)
$$g(\mathbf{x}) = \begin{bmatrix} y^2 - \dfrac{10\,y}{3} + 7\,x \\ -y^2 - 2\,x\,y - x^2 + 10\,x - \dfrac{14\,y}{3} \end{bmatrix} \qquad \text{where } \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Compute the Taylor series expansions for the stable and unstable manifolds of the hyperbolic fixed point $(x, y) = (0, 0)$ up to degree 3.

3.3. The Hartman-Grobman Theorem

Consider the dynamical system

$$\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{G}(\mathbf{x}_n) = \mathbf{F}(\mathbf{x}_n), \tag{DS}$$

where $\mathbf{0}$ is a hyperbolic fixed point and $\mathbf{G}$ vanishes to second order at the origin. When is (DS) equivalent to the linear system

$$\mathbf{y}_{n+1} = \mathbf{A}\mathbf{y}_n, \tag{LDS}$$

at least on some neighbourhood of the origin? That is, when is there a coordinate transformation $\mathbf{y} = \mathbf{Q}(\mathbf{x})$ that transforms (DS) into (LDS)?

Theorem 3.3.

[Hartman-Grobman] If $\mathbf{0}$ is a hyperbolic fixed point of (DS), then there is a neighourhood U of $\mathbf{0}$ and a homeomorphism $\mathbf{Q}$ that transforms (DS) into (LDS).

Recall that a homeomorphism is a 1-1, continuous map with a continuous inverse. Unlike the stable manifold, the homeomorphism may not be as smooth as the original dynamical system.

Unfortunately, the Hartman-Grobman theorem is not constructive, so actually finding a formula for $\mathbf{Q}$ is not straightforward. However, there is the tried-and-true method of Taylor series, as in the the stable manifold theorem. Let us try this method. Beforehand, let us note that solving for $\mathbf{Q}$ is equivalent to solving for its inverse, which we will call $\mathbf{H}$. It transpires that it is somewhat easier to find $\mathbf{H}$, so this is what we will do.

Example. Let our dynamical system $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ be defined by

$$\begin{aligned} x_{n+1} &= \frac{1}{2}x_n \qquad + x_n y_n + \alpha x_n^2 y_n, \\ y_{n+1} &= \qquad 2y_n + \beta x_n y_n^2. \end{aligned}$$

Clearly the origin is a hyperbolic fixed point, so let us try to construct the transformation $\mathbf{x} = \mathbf{H}(\mathbf{y})$ that transforms this system into its linear part.

First, let us note that we can solve for $\mathbf{H}$ one degree at a time. Let's solve for the quadratic part:

$$\mathbf{H}(x,y) = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{2,0}x^2 + a_{1,1}xy + a_{0,2}y^2 \\ b_{2,0}x^2 + b_{1,1}xy + b_{0,2}y^2 \end{bmatrix} + \cdots .$$

From this, we compute that

$$\mathbf{F}(\mathbf{H}(x,y)) = \begin{bmatrix} \frac{1}{2}x \\ 2y \end{bmatrix} + \begin{bmatrix} \frac{1}{2}a_{2,0}x^2 + (\frac{1}{2}a_{1,1}+1)xy + \frac{1}{2}a_{0,2}y^2 \\ 2b_{2,0}x^2 + 2b_{1,1}xy + 2b_{0,2}y^2 \end{bmatrix} + \cdots .$$

$$\mathbf{H}(\mathbf{A}(x,y)) = \begin{bmatrix} \frac{1}{2}x \\ 2y \end{bmatrix} + \begin{bmatrix} \frac{1}{4}a_{2,0}x^2 + a_{1,1}xy + 2a_{0,2}y^2 \\ \frac{1}{4}b_{2,0}x^2 + b_{1,1}xy + 4b_{0,2}y^2 \end{bmatrix} + \cdots .$$

We see that the linear terms are already equal (as they should be), while the quadratic terms are equal iff

$$\frac{1}{2}a_{1,1} + 1 = a_{1,1} \qquad \Longrightarrow \qquad a_{1,1} = 2,$$

and all other coefficients are zero. That is,

$$\mathbf{H}(x,y) = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2xy \\ 0 \end{bmatrix} + \begin{bmatrix} a_{3,0}x^3 + a_{2,1}x^2y + a_{1,2}xy^2 + a_{0,3}y^3 \\ b_{3,0}x^3 + b_{2,1}x^2y + b_{1,2}xy^2 + b_{0,3}y^3 \end{bmatrix} + \cdots .$$

From this, we compute that up to terms of degree four and higher,

$$\mathbf{F}(\mathbf{H}(x,y)) = \begin{bmatrix} \frac{1}{2}x + 2xy \\ 2y \end{bmatrix} + \begin{bmatrix} \frac{1}{2}a_{3,0}x^3 + (\frac{1}{2}a_{2,1}+\alpha)x^2y + (\frac{1}{2}a_{1,2}+2)xy^2 + \frac{1}{2}a_{0,3}y^3 \\ 2b_{3,0}x^3 + 2b_{2,1}x^2y + (2b_{1,2}+\beta)xy^2 + 2b_{0,3}y^3 \end{bmatrix}$$

$$\mathbf{H}(\mathbf{A}(x,y)) = \begin{bmatrix} \frac{1}{2}x \\ 2y \end{bmatrix} + \begin{bmatrix} \frac{1}{8}a_{3,0}x^3 + \frac{1}{2}a_{2,1}x^2y + 2a_{1,2}xy^2 + 8a_{0,3}y^3 \\ \frac{1}{8}b_{3,0}x^3 + \frac{1}{2}b_{2,1}x^2y + 2b_{1,2}xy^2 + 8b_{0,3}y^3 \end{bmatrix}.$$

These are equal iff

$$\frac{1}{2}a_{2,1} + \alpha = \frac{1}{2}a_{2,1}, \quad \frac{1}{2}a_{1,2} + 2 = 2a_{1,2}, \quad 2b_{1,2} + \beta = 2b_{1,2},$$

and all other coefficients are zero. We see that $a_{1,2} = 4/3$ solves the middle equation. However, the resonant equations have a solution if and only if $\alpha = \beta = 0$ – in which case $a_{2,1}$ and $b_{1,2}$ can assume any value! `Worksheet 3.3`

As a final note, this proves that when α or β is non-zero the homeomorphism $\mathbf{H}$ does not have all third-order derivatives at $\mathbf{0}$: if it did, then Taylor's theorem implies there would be a solution to the equations in degree 3.

3.3.1. An explanation. Let us conceptualize our example. Let $\mathbb{P}_d$ be the vector space of formal power series of maps from $\mathbb{R}^d$ to itself, and let $\mathbb{P}_{k,d}$ be the subspace of homogeneous maps of degree k. For $f \in \mathbb{P}_d$, we will write $f = \sum_k f_k$ where $f_k \in \mathbb{P}_{k,d}$ for all k.

Note that finding an invertible map that transforms (DS) into (LDS) is equivalent to finding an invertible map that transforms (LDS) into (DS). The latter will prove to be "easier," so let us do that. Assume that $\mathbf{x} = \mathbf{H}(\mathbf{y})$ does this. Then

$$\mathbf{H}(\mathbf{A}\mathbf{y}) = \mathbf{F}(\mathbf{H}(\mathbf{y})) \qquad \Longrightarrow \qquad \mathbf{H}_k(\mathbf{A}\mathbf{y}) = [\mathbf{F}(\mathbf{H}(\mathbf{y}))]_k.$$

Now, the degree k bit of $\mathbf{F}(\mathbf{H}(\mathbf{y}))$ starts with $\mathbf{A}.\mathbf{H}_k(\mathbf{y})$ plus additional terms that arise from the bits of $\mathbf{F}$ and $\mathbf{H}$ in lower degrees. Thus, our equation is of the form

$$\mathbf{H}_k(\mathbf{A}\mathbf{y}) = \mathbf{A}.\mathbf{H}_k(\mathbf{y}) + \mathbf{M}_k(\mathbf{y}), \tag{\%}$$

where $\mathbf{M}_k$ is a degree k homogeneous polynomial determined by the lower-degree terms of $\mathbf{F}$ and $\mathbf{H}$. Let us note that $\mathbf{H}_k$ is unknown and we would like to solve this equation. We note that it is linear: indeed, let $\mathcal{L}_k : \mathbb{P}_{k,d} \to \mathbb{P}_{k,d}$ be the transformation

$$\mathcal{L}_k(h) = h \circ \mathbf{A} - \mathbf{A} \circ h.$$

It is an (easy) exercise to show that $\mathcal{L}_k$ is linear and that equation (%) is the equation

$$\mathcal{L}_k(\mathbf{H}_k) = \mathbf{M}_k,$$

which has a solution iff $\mathbf{M}_k$ lies in the image of $\mathcal{L}_k$. Let us see when this is the case.

Assume that $\mathbf{A} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$. There is a convenient basis of $\mathbb{P}_{k,d}$ that diagonalizes $\mathcal{L}_k$. For each vector of natural numbers $\alpha = [\alpha_1, \ldots, \alpha_d]$ and each standard basis vector of $\mathbb{R}^d$, $\mathbf{e}_j$, let $\mathbf{f}_{\alpha,j}(\mathbf{y}) = \mathbf{y}_1^{\alpha_1} \cdots \mathbf{y}_d^{\alpha_d}\, \mathbf{e}_j$. The collection of $\mathbf{f}_{\alpha,j}$ for $\alpha_1 + \cdots + \alpha_d = k$ forms a basis of $\mathbb{P}_{k,d}$. One notes

$$\mathcal{L}_k(\mathbf{f}_{\alpha,j})(\mathbf{y}) = (\lambda_1 \mathbf{y}_1)^{\alpha_1} \cdots (\lambda_d \mathbf{y}_d)^{\alpha_d}\, \mathbf{e}_j - \lambda_j \mathbf{y}_1^{\alpha_1} \cdots \mathbf{y}_d^{\alpha_d}\, \mathbf{e}_j = (\lambda_1^{\alpha_1} \cdots \lambda_d^{\alpha_d} - \lambda_j)\mathbf{f}_{\alpha,k}(\mathbf{y}).$$

Therefore, the basis $\mathbf{f}_{\alpha,j}$ diagonalizes $\mathcal{L}_k$. From this we also see that

$$\ker \mathcal{L}_k = \operatorname{span}\{\mathbf{f}_{\alpha,j} \;:\; \lambda_1^{\alpha_1} \cdots \lambda_d^{\alpha_d} = \lambda_j\}, \qquad \operatorname{im} \mathcal{L}_k = \operatorname{span}\{\mathbf{f}_{\alpha,j} \;:\; \lambda_1^{\alpha_1} \cdots \lambda_d^{\alpha_d} \neq \lambda_j\}.$$

Definition 3.3.5.

A resonance of degree k is an α, j such that $\lambda_1^{\alpha_1} \cdots \lambda_d^{\alpha_d} = \lambda_j$.

Resonances are obstacles to solving equation (%). They explain why the homeomorphism may not be smooth. However, most d-tuples of eigenvalues have no resonances, in which case we can solve for $\mathbf{H}$ as a formal power series. If we can solve (%) in degrees $1, \ldots, N$, then (DS) and (LDS) are equivalent up to terms of order $N + 1$.

In the above example, since $\lambda_1\lambda_2 = 1$, we have $\lambda_1^2\lambda_2 = \lambda_1$ and $\lambda_1\lambda_2^2 = \lambda_2$. Each are degree 3 resonances with $\alpha = (2, 1), j = 1$ and $\alpha = (1, 2), j = 2$ respectively. Our discussion above shows that we could find $\mathbf{H}(\mathbf{y})$ to kill off the quadratic part, but that we will not be able to remove those two cubic terms.

These observations lead us to record the following useful addition to the Hartman-Grobman theorem.

Theorem 3.4.

[Sternberg Linearisation] If $\mathbf{0}$ is a hyperbolic fixed point of (DS), and the eigenvalues of $\mathbf{A}$ are resonance-free, and $\mathbf{F}$ is infinitely differentiable (C^∞), then the homeomorphism of the Hartman-Grobman theorem can be chosen to be C^∞ with a C^∞ inverse.

Exercise 3.3.3.

(i) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the map defined in equation 3.2.2. Compute, to third order, the Taylor expansion of a homeomorphism h defined on a neighbourhood of $\mathbf{0}$ that conjugates f to its linear part: $f \circ h = h \circ A$, $A = df_{\mathbf{0}}$.

(ii) Give an explicit formula for $\mathbf{H}_k$ in terms of $\mathbf{M}_k$, by writing the latter in terms of the basis $\mathbf{f}_{\alpha,j}$.

(iii) Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x/2 + y^2 \\ 2y + (x + 2y^2)^2/4 \end{bmatrix}. \tag{3.3.1}$$

Is there a homeomorphism that linearises g at the origin and is also differentiable?

3.4. Eigenvalues with modulus 1

The previous section shows how near hyperbolic fixed points the behaviour of the nonlinear system is close to that of the linearised system. It is not so in the (non-generic) case where some eigenvalues have modulus equal to 1. Here we treat the two-dimensional case, when the two eigenvalues of $A = D\mathbf{F}|_{\mathbf{x}_*}$ have modulus 1 and are complex valued. This case is particularly important since the stability of the fixed point is left undecided by the linear analysis.

Consider the 2-dimensional system $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ with a fixed point $\mathbf{x}_*$ at which the linearised system has the complex pair of eigenvalues $\lambda_1 = \overline{\lambda}_2 = \lambda$ with $|\lambda| = 1$. The fixed point is then a centre for the linearised system, but for the nonlinear system the higher-order terms will typically convert it into either an attracting or repelling focus. Below is described a method for determining which of these two alternatives occurs; it involves making a series of coordinate changes to reduce the system to a *normal form* which is easier to study. To derive this normal form, we make the further assumptions $\lambda^3 \neq 1, \lambda^4 \neq 1$. This means that $\lambda \neq \pm 1, \dfrac{-1 \pm \sqrt{-3}}{2}, \pm\sqrt{-1}$.

(i) First, by making a linear change of coordinates, we can arrange that $\mathbf{x}_* = \mathbf{0}$. We then introduce the complex-valued variable $z = \mathbf{t} \cdot \mathbf{x}$, where $\mathbf{t}$ is an eigenvector of A^{T}, the transpose of A, associated with the eigenvalue λ ; i.e.,

$$A^{\mathrm{T}}\mathbf{t} = \lambda \mathbf{t}\,.$$

Thus, the linearised system $\mathbf{x}_{n+1} = A\mathbf{x}_n$ may be expressed as $z_{n+1} = \lambda z_n$ (since $\mathbf{t} \cdot \mathbf{x}_{n+1} = \mathbf{t} \cdot (A\mathbf{x}_n) = (A^{\mathrm{T}}\mathbf{t}) \cdot \mathbf{x}_n = \lambda \mathbf{t} \cdot \mathbf{x}_n$) and the general nonlinear system $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n) = A\mathbf{x}_n + O(|\mathbf{x}_n|^2)$ may be expressed as

$$z_{n+1} = \lambda z_n + a z_n^2 + b z_n \overline{z}_n + c\overline{z}_n^2 + O(|z_n|^3). \tag{3.4.1}$$

The complex-valued constant coefficients a, b and c in (3.4.1) can be determined through expressing z_n and $\overline{z}_n$ as functions of x_n and y_n in the Taylor expansion of $\mathbf{F}$.

(ii) The next step is to make another change of variables to eliminate the quadratic terms in our dynamical system representation. We try

$$w_n = z_n + \alpha z_n^2 + \beta z_n \overline{z}_n + \gamma \overline{z_n}^2. \tag{3.4.2}$$

Then we use (3.4.1) to expand w_{n+1}, keeping only terms up to second-order (and noting, for example that the difference between z_n^2 and w_n^2 is third order, so z_n^2 can be replaced by w_n^2, etc). This gives

$$w_{n+1} = \lambda z_n + a z_n^2 + b z \overline{z}_n + c\overline{z}_n^2 + \alpha\lambda^2 z_n^2 + \beta\lambda\overline{\lambda} z_n \overline{z}_n + \gamma\overline{\lambda}^2 \overline{z}_n^2 + O(|z_n|^3). \tag{3.4.3}$$

But from (3.4.2) we have

$$\lambda z_n = \lambda w_n - \left(\alpha\lambda z_n^2 + \beta\lambda z_n \overline{z}_n + \gamma\lambda\overline{z}_n^2 \right)$$

which we substitute into (3.4.3) to give

$$w_{n+1} = \lambda w_n + z_n^2 \left(a + \alpha\lambda^2 - \lambda\alpha \right) + z_n\overline{z}_n \left(b + \beta\lambda\overline{\lambda} - \lambda\beta \right) + \overline{z}_n^2 \left(c + \gamma\overline{\lambda}^2 - \lambda\gamma \right) + O(|z_n|^3).$$

Thus, the system reduces to

$$w_{n+1} = \lambda w_n + O(|w_n|^3)$$

provided we take

$$\alpha = \frac{a}{\lambda(1-\lambda)}, \quad \beta = \frac{b}{\lambda - 1}, \quad \gamma = \frac{c}{\lambda - \overline{\lambda}^2} = \frac{c}{\lambda(1-\overline{\lambda}^3)}. \tag{3.4.4}$$

Provided that $\lambda^3 \neq 1$, these coefficients are all well-defined. Thereby, we write

$$w_{n+1} = \lambda w_n + p w_n^3 + q w_n^2 \overline{w}_n + r w_n \overline{w}_n^2 + s\overline{w}_n^3 + O(|w_n|^4) \tag{3.4.5}$$

where p, q, r, s are complex-valued constants.

(iii) The final stage is to eliminate as many of the third-order terms as possible, by a similar procedure to that used in (ii). We make the change of variable

$$\zeta_n = w_n + d w_n^3 + e w_n^2 \overline{w}_n + f w_n \overline{w}_n^2 + g\overline{w}_n^3 \tag{3.4.6}$$

and try to choose d, e, f, g so as to eliminate the third-order terms. Following the argumentation of (ii), we combine (3.4.5) with (3.4.6) to give

$$\begin{aligned}\zeta_{n+1} &= \lambda w_n + p w_n^3 + q w_n^2 \overline{w}_n + r w_n \overline{w}_n^2 + s \overline{w}_n^3 \\ &\quad + d\lambda^3 w_n^3 + e\lambda^2 \overline{\lambda} w_n^2 \overline{w}_n + f\lambda \overline{\lambda}^2 w_n \overline{w}_n^2 + g \overline{\lambda}^3 \overline{w}_n^3 + O(|w_n|^4);\end{aligned} \tag{3.4.7}$$

but from (3.4.6) we have

$$\lambda w_n = \lambda \zeta_n - \lambda \left(d w_n^3 + e w_n^2 \overline{w}_n + f w_n \overline{w}_n^2 + g \overline{w}_n^3 \right)$$

which we substitute into (3.4.7) to give

$$\begin{aligned}\zeta_{n+1} &= \lambda \zeta_n + w_n^3 \left(p + d\lambda^3 - \lambda \mathrm{d}\right) + w_n^2 \overline{w}_n \left(q + e\lambda^2 \overline{\lambda} - \lambda e\right) \\ &\quad + w_n \overline{w}_n^2 \left(r + f\lambda \overline{\lambda}^2 - \lambda f\right) + \overline{w}_n^3 \left(s + g\overline{\lambda}^3 - \lambda g\right) + O(|w_n|^4).\end{aligned}$$

In order to eliminate cubic terms we require

$$d = \frac{p}{\lambda(1-\lambda^2)}, \quad e = \frac{q}{\lambda(1-\lambda\overline{\lambda})}, \quad f = \frac{r}{\lambda(1-\overline{\lambda}^2)}, \quad g = \frac{s}{\lambda - \overline{\lambda}^3} = \frac{s}{\lambda(1-\overline{\lambda}^4)}.$$

There is a problem with the term in $w_n^2 \overline{w}_n$: its elimination requires $e\lambda(1-|\lambda|^2) = q$ which is problematic since $|\lambda| = 1$. There is a second potential problem in the denominators of d, f, g: if $\lambda^4 = 1$, then at least one denominator is zero. So we choose d, f, g according to the three equations above but take $e = 0$, *provided that* $\lambda^4 \neq 1$. Then the system reduces to

$$\zeta_{n+1} = \lambda \zeta_n + q \zeta_n^2 \overline{\zeta}_n + O(|\zeta|^4)$$

This is the desired normal form.

Analysis of the normal form

The behaviour of this normal form is analysed as follows: Multiplying by $\overline{\zeta}_{n+1}$ gives

$$|\zeta_{n+1}|^2 = |\zeta_n|^2 + 2h|\zeta_n|^4 + O(|\zeta_n|^5),$$

where

$$h = \frac{1}{2}\left(\frac{q}{\lambda} + \frac{\overline{q}}{\overline{\lambda}}\right) = \mathrm{Re}\left(\frac{q}{\lambda}\right).$$

Clearly $|\zeta_n| \to 0$ for $n \to +\infty$ if $h < 0$ and $|\zeta_n| \to \infty$ if $h > 0$. Therefore, we conclude that $\mathbf{x}_* = \mathbf{0}$ is attracting (stable) provided that $h < 0$ and repelling (unstable) provided that $h > 0$.

Determination of h

We next describe how to calculate the quantity h, whose sign determines the nature of the focus at $\mathbf{x}_* = \mathbf{0}$. We need the term in $z_n^2\overline{z}_n$, so we consider

$$z_{n+1} = \lambda z_n + az_n^2 + bz_n\overline{z}_n + c\overline{z}_n^2 + mz_n^2\overline{z}_n + \cdots$$

where the dots denote other cubic terms and $O(|z_n|^4)$ terms. We have to repeat stage (ii), but keep track of the $z_n^2\overline{z}_n$ terms. Repeating the argumentation of (ii), with the variable w_n having the form (3.4.2), we find

$$\begin{aligned} w_{n+1} &= \lambda z_n + az_n^2 + bz\overline{z}_n + c\overline{z}_n^2 + mz_n^2\overline{z}_n + \alpha\left(\lambda^2 z_n^2 + 2\lambda b z_n\overline{z}_n\right) \\ &\quad + \beta\left(\lambda\overline{\lambda}z_n\overline{z}_n + a\overline{\lambda}z_n^2\overline{z}_n + \overline{b}\lambda z_n^2\overline{z}_n\right) + \gamma\left(\overline{\lambda}^2\overline{z}_n^2 + 2\overline{c}\overline{\lambda}z_n^2\overline{z}\right) + \cdots \end{aligned}$$

where we have ignored cubic terms other than $z_n^2\overline{z}_n$. As in (ii), we eliminate the quadratic terms by choosing α, β and γ according to (3.4.4). Thereby, we obtain

$$w_{n+1} = \lambda w_n + \left(m + \frac{(2\lambda - 1)ab}{\lambda(1-\lambda)} + \frac{\lambda|b|^2}{\lambda - 1} + \frac{2\lambda|c|^2}{\lambda^3 - 1}\right) w_n^2\overline{w}_n$$

where again we have ignored cubic terms other than $w_n^2\overline{w}_n$, and terms of order higher than 3. Finally, we conclude that

$$h = \mathrm{Re}\left[\frac{m}{\lambda} + \frac{(2\lambda-1)ab}{\lambda^2(1-\lambda)} + \frac{|b|^2}{\lambda - 1} + \frac{2|c|^2}{\lambda^3 - 1}\right].$$

Example. The dynamical system

$$\begin{aligned} x_{n+1} &= y_n \\ y_{n+1} &= -x_n + y_n + x_n y_n \end{aligned}$$

has the origin as a fixed point, so the linearized system is obtained by dropping the nonlinear terms. The resulting matrix

$$D\mathbf{F}|_{\mathbf{0}} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$$

has eigenvalues $\lambda = \lambda_1 = \frac{1+i\sqrt{3}}{2} = \overline{\lambda}_2$ (note that $|\lambda| = 1$ and $\lambda^3 = -1$). We consider the eigenvector $\mathbf{t}$ satisfying

$$\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}^{\mathrm{T}} \mathbf{t} = \lambda\mathbf{t},$$

which we calculate to be $\mathbf{t} = \begin{pmatrix} \frac{1}{2}\left(-1 + i\sqrt{3}\right) \\ 1 \end{pmatrix}$. Now, let us express the recurrence relation governing our system in terms of the the complex variable

$$z = \mathbf{t}\cdot\mathbf{x} = \frac{1}{2}\left(-1 + i\sqrt{3}\right)x + y;$$

i.e.,

$$\begin{aligned} z_{n+1} &= \frac{1}{2}\left(-1 + i\sqrt{3}\right)y_n + y_n - x_n + x_n y_n \\ &= \frac{1}{2}\left(1 + i\sqrt{3}\right)\left[\frac{1}{2}\left(-1 + i\sqrt{3}\right)x_n + y_n\right] + x_n y_n \\ &= \frac{1}{2}\left(1 + i\sqrt{3}\right)z_n + x_n y_n. \end{aligned}$$

By considering the combinations $z_n - \overline{z}_n$ and $z_n + \overline{z}_n$, we see that

$$x_n = \frac{1}{i\sqrt{3}}(z_n - \overline{z}_n) \qquad \text{and} \qquad y_n = \frac{1}{2}\left[z_n + \overline{z}_n + \frac{1}{i\sqrt{3}}(z_n - \overline{z}_n)\right];$$

hence,

$$z_{n+1} \;=\; \frac{1}{2}\left(1 + i\sqrt{3}\right) z_n + \frac{1}{i2\sqrt{3}}(z_n - \overline{z}_n)\left[z_n + \overline{z}_n + \frac{1}{i\sqrt{3}}(z_n - \overline{z}_n)\right]$$

which may be expressed in the form

$$z_{n+1} = \lambda z_n + a z_n^2 + b z_n \overline{z}_n + c\overline{z}_n^2 + m z_n^2 \overline{z}_n$$

with $a = -\frac{1}{6} - \frac{i}{2\sqrt{3}}$, $b = \frac{1}{3}$, $c = -\frac{1}{6} + \frac{i}{2\sqrt{3}}$ and $m = 0$. Thus, we find that $h = -\frac{1}{6}$ and we therefore conclude that the origin is attracting (stable).

Exercise 3.4.4.

(i) Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be the map

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -2y^2 - 6xy + y - 4x^2 + 2x \\ 4y^2 + 12xy - y + 8x^2 - 3x. \end{bmatrix} \tag{3.4.8}$$

Show that the dynamical system $\mathbf{x}_{n+1} = g(\mathbf{x}_n)$ has a fixed point at the origin which is asymptotically stable and the origin is only stable for the linearised dynamical system.

3.4.1. A deeper analysis. Let our dynamical system (3.4.1) be written as $z_{n+1} = f(z_n)$ where $f(z) = \lambda z + az^2 + bz\bar{z} + c\bar{z}^2 + \cdots$. We want to find a transformation $w = q(z)$ (equation (3.4.2) which transforms (3.4.1) into the linear equation $w_{n+1} = \lambda w_n$. As in the discussion of the Hartman-Grobman theorem, it is more convenient to try to find the inverse h of q. In this case, we want to solve

$$f(h(z)) = h(\lambda z). \tag{$*$}$$

To solve (*), we tried to solve it in successive degrees. Why could we solve the equation in degree 2, but we ran into problems in degree 3? The first answer is: the degree 3 part actually contains dynamical information (stability information). A second answer lies in analyzing the algebraic structure of the problem. To see this, let $\mathbb{P}_k$ be the vector space of degree k homogeneous polynomials in z and $\bar{z}$ and let $\mathbb{P}$ be the vector space of formal power series in z and $\bar{z}$. For $h \in \mathbb{P}$ we write $h = \sum_k h_k$ where $h_k \in \mathbb{P}_k$. Define a map $\mathcal{L}_k : \mathbb{P}_k \to \mathbb{P}_k$ by

$$\mathcal{L}_k(h) = h \circ \lambda - \lambda \circ h,$$

where we view λ as a function that acts by scalar multiplication. The degree k part of (*) is an equation of the form

$$\mathcal{L}_k(h_k) = m_k,$$

where m_k is a polynomial that depends on $f_1, \ldots, f_k$ and $h_1, \ldots, h_{k-1}$.

What does this operator look like? For $a + b = k$, let $r_{a,b}(z) = z^a\,\bar{z}^b$. One sees that

$$\mathcal{L}_k(r_{a,b}) = (\lambda^{a-b} - \lambda) r_{a,b},$$

since $\lambda\bar{\lambda} = 1$. Since $\mathcal{L}_k$ is diagonal in this basis, it is easy to see that:

$$\ker \mathcal{L}_k = \{r_{a,b} \;:\; a + b = k, a - b = 1 \bmod n\}, \qquad \operatorname{im} \mathcal{L}_k = \{r_{a,b} \;:\; a + b = k, a - b \neq 1 \bmod n\},$$

where n is the smallest positive integer such that $\lambda^n = 1$ (if no such n exists, set it equal to 0).

It is straightforward to see that

(i) $\mathcal{L}_2$ is invertible, unless $\lambda^3 = 1$ ($a = 0, b = 2$ and $a - b = 1 \bmod 3$);
(ii) $\mathcal{L}_3$ has kernel spanned by $z^2\bar{z}$ ($a = 2, b = 1$), unless $\lambda^4 = 1$, when $\bar{z}^3$ is also in the kernel ($a = 0, b = 3$ and $a - b = 1 \bmod 4$).

As an exercise, can you determine the image and kernel of $\mathcal{L}_4$?

3.5. Stability of periodic orbits

In discrete systems, periodic orbits can be regarded as fixed points of the iterated map. Thus the stability of a periodic orbit $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m\}$ of period m may be determined by the eigenvalues of the matrix

$$D\mathbf{F}^m|_{\mathbf{x}_i}$$

where $\mathbf{x}_i$ can be any point of the periodic orbit (since they are all fixed points of $\mathbf{F}^m$). The eigenvalues of the matrix are in fact independent of $\mathbf{x}_i$ as is found using the chain rule:

$$\begin{aligned} D\mathbf{F}^m|_{\mathbf{x}_1} &= D\left[\mathbf{F}(\mathbf{F}\cdots\mathbf{F}(\mathbf{x}))\right]|_{\mathbf{x}_1} \\ &= D\mathbf{F}|_{\mathbf{F}^{m-1}(\mathbf{x}_1)}\, D\mathbf{F}|_{\mathbf{F}^{m-2}(\mathbf{x}_1)} \cdots D\mathbf{F}|_{\mathbf{x}_1} \\ &= D\mathbf{F}|_{\mathbf{x}_m}\, D\mathbf{F}|_{\mathbf{x}_{m-1}} \cdots D\mathbf{F}|_{\mathbf{x}_1}. \end{aligned}$$

CHAPTER 4

One-dimensional dynamics

4.1. Introduction

We now study in detail the nonlinear dynamics of $\mathbb{R}^1$ discrete systems:

$$x_{n+1} = F(x_n).$$

The successive iterates x_n of the map can be obtained graphically on a staircase diagram, as illustrated in figure 4.1.1. `Worksheet 4`

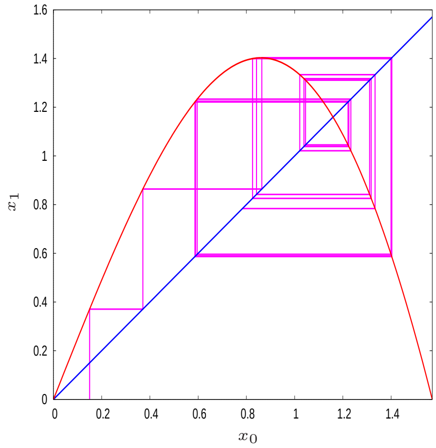


FIGURE 4.1.1. Staircase diagram for the mapping $x_{n+1} = 2.8x_n \cos x_n$.

Fixed points x^* are seen immediately as intersections between the graph $y = F(x)$ and $y = x$. Their stability depends on the derivative of the map $F'(x^*)$ evaluated at x^*:

- If $|F'(x^*)| < 1$, then x^* is stable and it can be shown to be attracting in its neighbourhood; i.e., $\exists \epsilon > 0$ such that all x with $|x - x^*| < \epsilon$ have $|F^k(x)| \to x^*$ for $k \to \infty$.
- Conversely, if $|F'(x^*)| > 1$, then x^* is unstable and repelling in its neighbourhood.

Periodic points can be determined graphically by considering the iterated map $F^k(x) = F(F(\cdots(F(x))\cdots))$. See `Worksheet 4` for examples of the numerical determination of fixed and periodic points using Maple.

4.2. Bifurcation theory

Many dynamical systems depend on a parameter (e.g., μ for the logistic map $x_{n+1} = \mu x_n(1 - x_n)$), and one is interested in the behaviour of the system for different values of this parameter. When the parameter value is changed continuously, the qualitative properties of the system (existence and stability of fixed points, of periodic orbits, etc.) remain the same — except when the parameter passes through a critical value, in which case the qualitative properties change suddenly and the system is said to undergo a *bifurcation*.

We consider some important bifurcations which arise when the number of fixed points or their stability change.

A highly important result about the bifurcation of fixed points is the following 'non-bifurcation' result. We consider the map $x_{n+1} = F_\mu(x_n)$ which depends on a parameter μ and suppose that for $\mu = \mu^*$ a fixed point x^* exists:

$$F_{\mu^*}(x^*) = x^*.$$

Then, provided that $F'_{\mu^*}(x^*) \neq 1$, it can be shown that for any μ sufficiently close to μ^* (but not equal to μ^*), the system has a fixed point, i.e., $\exists x_\mu$ such that $F_\mu(x_\mu) = x_\mu$. To prove this, consider the function $G(x, \mu) = F_\mu(x) - x$. Clearly,

$$G(x^*, \mu^*) = 0 \qquad \text{and} \qquad \left.\frac{\partial G}{\partial x}\right|_{(x^*,\,\mu^*)} \neq 0,$$

so that, by the *implicit function theorem*, the equation $G(x, \mu) = 0$ (which determines the fixed points) has a solution $x = x_\mu$ for x_μ near x^* and μ near μ^*.

Thus, a fixed point persists in the neighbourhood of x^* and μ^* provided $F'_{\mu^*}(x^*) \neq 1$. However, a change in the stability of the fixed point is possible only if $|F'_\mu(x)|$ passes through 1. We consider the cases for $F'_\mu(x)$ passing through $+1$ and -1, respectively, in the following 2 sections. Note that, in the following sections, fixed points will not generally be labelled with the supercript * where it is clear from the context that the point in question is a fixed point.

4.2.1. Saddle–node bifurcation. We consider the *saddle–node* bifurcation[1] which arises at the critical point x_c for the critical parameter value μ_c. This type of bifurcation is associated with a change in the number of fixed points; it corresponds to the passage of $F'_\mu(x)$ through $+1$.

For $\mu < \mu_c$, the system has no fixed points in the vicinity of x_c; for $\mu > \mu_c$, the system has two fixed points, one of which is unstable (a saddle in general) and the other one is stable (a node). More formally, the one-parameter family $F_\mu(x)$ of one-dimensional mappings undergoes a saddle–node bifurcation at the parameter value μ_c if there is an open interval I and an $\varepsilon > 0$ such that:

(i) For $\mu \in (\mu_c - \varepsilon, \mu_c)$, F_μ has no fixed points in the interval I.
(ii) At $\mu = \mu_c$, F_μ has one fixed point in I at x_c and $F'_{\mu_c}(x_c) = 1$.
(iii) For $\mu \in (\mu_c, \mu_c + \varepsilon)$, F_μ has two fixed points in I, one attracting and one repelling.

A saddle–node bifurcation also occurs if the direction of the bifurcation is reversed. That is, a saddle–node bifurcation occurs at the parameter value μ_c if there is an open interval I and an $\varepsilon > 0$ such that:

(ir) For $\mu \in (\mu_c, \mu_c + \varepsilon)$, F_μ has no fixed points in the interval I.
(iir) At $\mu = \mu_c$, F_μ has one fixed point in I at x_c and $F'_{\mu_c}(x_c) = 1$.
(iiir) For $\mu \in (\mu_c - \varepsilon, \mu_c)$, F_μ has two fixed points in I, one attracting and one repelling.

An example is provided by the system $x_{n+1} = x_n + \mu - x_n^2$ with $\mu \in (-1, 1)$. For $\mu < \mu_c = 0$, there are no fixed points; when $\mu = \mu_c = 0$ there is a fixed point at $x = x_c = 0$; when $\mu > \mu_c$, there are two fixed points, $x = \sqrt{\mu}$ which is stable, and $x = -\sqrt{\mu}$ which is unstable (see figure 4.2.2).

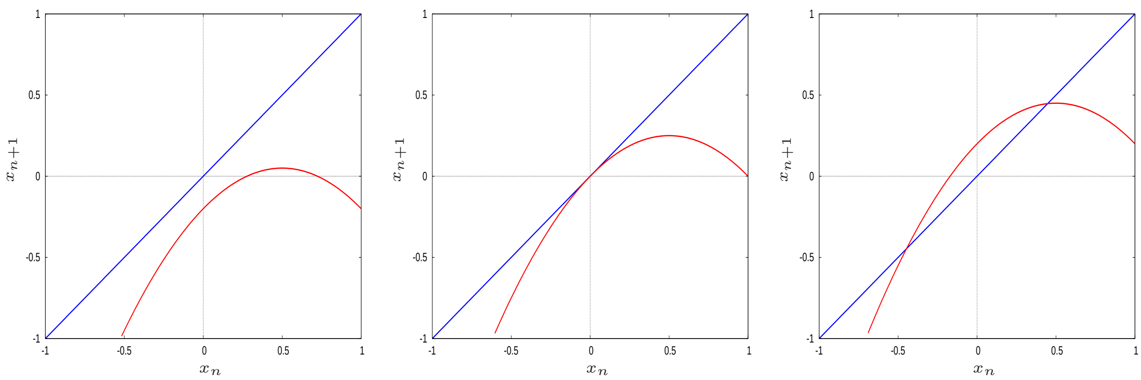
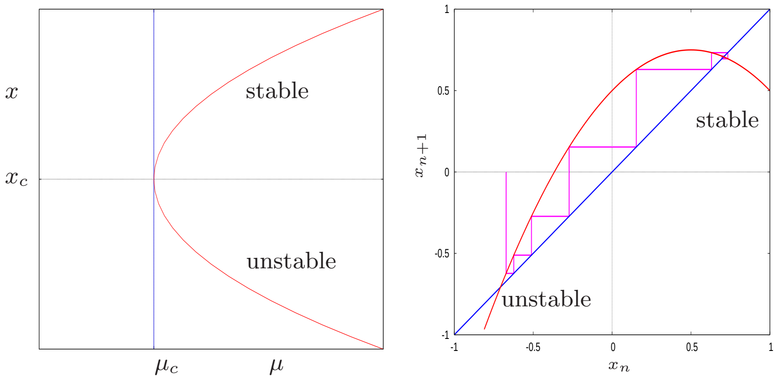


FIGURE 4.2.2. The saddle-node bifurcation for the map $f_\mu(x) = x + \mu - x^2$.

In one dimension, one can represent the saddle–node bifurcation in the space (μ, x) as shown in figure 4.2.3.

[1]The term *saddle–node* derives from a description of this bifurcation in higher dimensions.



(a) Fixed points as a function of μ. (b) Cobweb diagram for $\mu = 0.5$.

FIGURE 4.2.3. The fixed-points in a saddle-node bifurcation.

4.2.2. Flip bifurcation. We next consider the situation which arises when $F'_\mu(x)$ passes through -1. In this case the fixed point persists, but becomes unstable, and the bifurcation is called a *flip bifurcation* (or period-doubling bifurcation). For this bifurcation the change in the stability of the fixed point is accompanied by the appearance of a period-two orbit. More formally, the one-parameter family $F_\mu(x)$ of one-dimensional mappings undergoes a flip bifurcation at the parameter value μ_c if there is an open interval I and an $\varepsilon > 0$ such that:

(i) For each $\mu \in [\mu_c - \varepsilon, \mu_c + \varepsilon]$, there is a unique fixed point x_μ for F_μ in the interval I.
(ii) For $\mu \in (\mu_c - \varepsilon, \mu_c]$, F_μ has no orbits of period 2 in the interval I and x_μ is attracting (resp. repelling).
(iii) For each $\mu \in (\mu_c, \mu_c + \varepsilon)$, there is a unique period–2 orbit $\{x_\mu^{(1)}, x_\mu^{(2)}\}$ in the interval I with $F_\mu(x_\mu^{(1)}) = x_\mu^{(2)}$. This period–2 orbit is attracting (resp. repelling); meanwhile the fixed point x_μ is repelling (resp. attracting).
(iv) As $\mu \to \mu_c$, we have $x_\mu^{(1)} \to x_\mu$ and $x_\mu^{(2)} \to x_\mu$.

Thus, there are two typical cases for a flip bifurcation: (a) as μ increases, a fixed point may change its nature from attracting to repelling and, at the same time, give birth to an attracting period-2 cycle; and (b) as μ increases, a fixed point may change its nature from repelling to attracting and, at the same time, give birth to a repelling period-2 cycle.

As in the saddle–node case, the direction in which the bifurcation occurs may be reversed. That is, the one-parameter family $F_\mu(x)$ of one-dimensional mappings undergoes a flip bifurcation at the parameter value μ_c if there is an open interval I and an $\varepsilon > 0$ such that:

(ir) For each $\mu \in [\mu_c - \varepsilon, \mu_c + \varepsilon]$, there is a unique fixed point x_μ for F_μ in the interval I.
(iir) For $\mu \in (\mu_c, \mu_c + \varepsilon)$, F_μ has no orbits of period 2 in the interval I and x_μ is attracting (resp. repelling).
(iiir) For each $\mu \in (\mu_c - \varepsilon, \mu_c]$, there is a unique period–2 orbit $\{x_\mu^{(1)}, x_\mu^{(2)}\}$ in the interval I with $F_\mu(x_\mu^{(1)}) = x_\mu^{(2)}$. This period–2 orbit is attracting (resp. repelling); meanwhile the fixed point x_μ is repelling (resp. attracting).
(ivr) As $\mu \to \mu_c$, we have $x_\mu^{(1)} \to x_\mu$ and $x_\mu^{(2)} \to x_\mu$.

To examine this bifurcation, we consider the one-parameter family F_μ of one-dimensional mappings, and suppose F_μ has a fixed point x_μ depending smoothly on μ. By making a change of coordinates (depending on μ) we can arrange that $x_\mu = 0$. Then we can write

$$F_\mu(x) = ax + bx^2 + cx^3 + \cdots$$

where the coefficients a, b, c depend on μ. We are interested in the change that occurs as a passes through -1 (since $F'_{\mu_c}(0) = a$) and so we suppose $a = -1$ when $\mu = \mu_c$.

To eliminate the x^2 term in F_μ we try the change of variable: $y_n = x_n + \alpha x_n^2$. We find that

$$\begin{aligned} y_{n+1} &= x_{n+1} + \alpha x_{n+1}^2 \\ &= a x_n + b x_n^2 + c x_n^3 + \alpha\left(a^2 x_n^2 + 2ab x_n^3\right) + O(x_n^4) \\ &= a\left(x_n + \frac{b+\alpha a^2}{a} x_n^2\right) + (c + 2\alpha ab)\, x_n^3 + O(x_n^4) \\ &= a y_n + \beta y_n^3 + O(y_n^4) \end{aligned}$$

provided we take

$$\alpha = \frac{b+\alpha a^2}{a}, \qquad \text{i.e.,} \qquad \alpha = \frac{b}{a(1-a)}$$

and

$$\beta = c + 2\alpha ab = c + \frac{2b^2}{1-a}.$$

Thereby, the system $x_{n+1} = F_\mu(x_n)$ is transformed to $y_{n+1} = G_\mu(y_n)$ wherein

$$G_\mu(y) = ay + \beta y^3 + O(y^4).$$

For a period-2 orbit we require

$$\begin{aligned} y &= G_\mu(G_\mu(y)), \\ \text{i.e.,} \qquad y &= a^2 y + a\beta(1+a^2)y^3 + O(y^4) \end{aligned}$$

which has the possible solutions $y = 0$ (corresponding to the fixed point of F_μ) and

(4.2.1) $$y \approx \pm\sqrt{\frac{1-a^2}{a\beta(1+a^2)}}.$$

4.2.3. Supercritical case: $\beta(\mu_c) > 0$**.** In this case $\beta(\mu) > 0$ for μ near μ_c. Consider a sufficiently small $\delta > 0$. Then we have the following scenarios

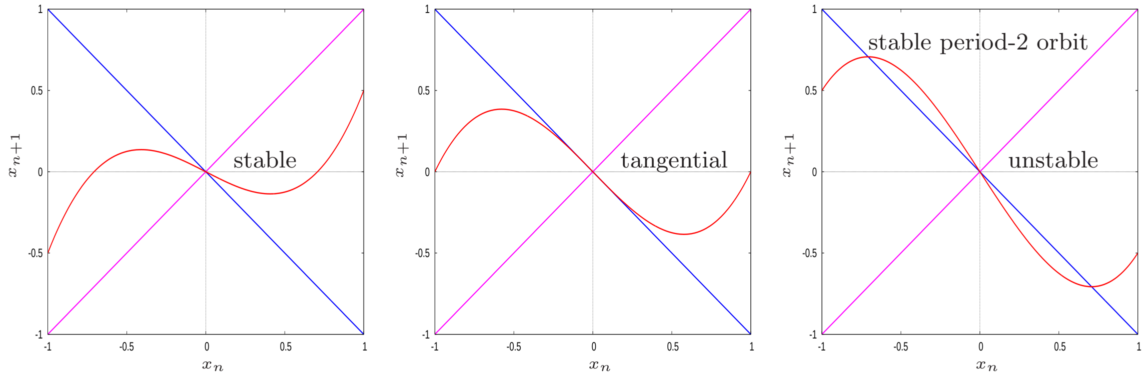
- If $-1-\delta < a(\mu) < -1$ then:
  (a) $|a(\mu)| > 1$ so the fixed point of F_μ is repelling;
  (b) 2 solutions (4.2.1) exist (y_1 and y_2, say) and at these points we have:
  $$\left|\frac{d}{dy}\left(a^2 y + a\beta(1+a^2)y^3\right)\right|_{y=y_{1,2}} < 1;$$
  i.e., the 2 points y_1 and y_2 comprise an attracting period-2 orbit.
- If $-1 < a(\mu) < -1+\delta$ then:
  (a) $|a(\mu)| < 1$ so the fixed point of F_μ is attracting;
  (b) no real-valued solutions (4.2.1) exist.

Thus the bifurcation is associated with a change in the *attractor* (i.e., the set of points attracting nearby orbits), from a fixed point to a period-2 orbit[2]. Typical orbits for $a > -1$ and $a < -1$ are shown in figure 4.2.4.

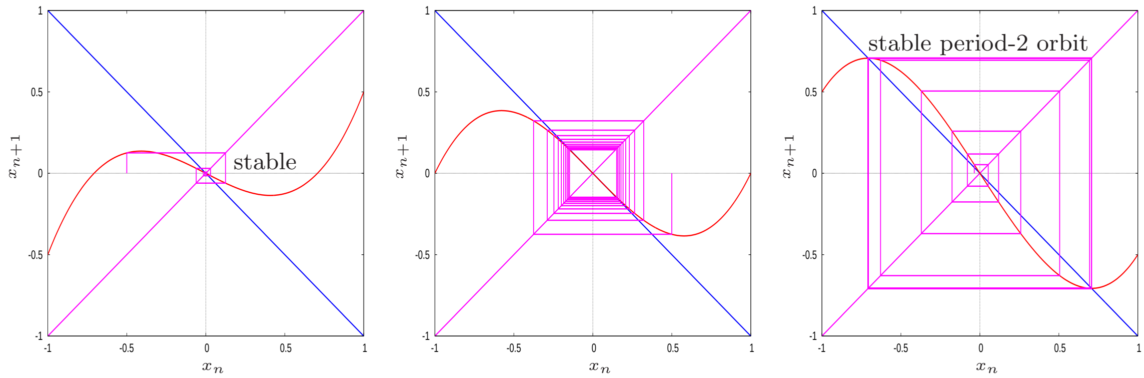
4.2.4. Subcritical case: $\beta(\mu_c) < 0$**.** In this case $\beta(\mu) < 0$ for μ near μ_c. Consider a sufficiently small $\delta > 0$. Then we have the following scenarios

- If $-1-\delta < a(\mu) < -1$ then:
  (a) $|a(\mu)| > 1$ so the fixed point of F_μ is repelling;
  (b) no real-valued solutions (4.2.1) exist.
- If $-1 < a(\mu) < -1+\delta$ then:
  (a) $|a(\mu)| < 1$ so the fixed point of F_μ is attracting;

[2]More precisely, an *attractor* may be defined as a closed set A with the properties:
(i) *A is an invariant set*: $\mathbf{x}_0 \in A \Rightarrow \mathbf{x}_n \in A\ \forall n$.
(ii) *A attracts an open set of initial conditions*: $\exists$ open set $U \supset A$ such that if $\mathbf{x}_0 \in U$ then the 'distance' from $\mathbf{x}_n$ to A tends to zero as $n \to \infty$. The largest such U is called the *basin of attraction* of A.
(iii) *A is minimal*: there is no proper subset of A that satisfies conditions (i) and (ii).



(a) $-1 < \mu < 0, -1 < a < 0$, $x = 0$ is stable.
(b) $\mu = 0, a = -1$, graph of f_0 crosses $y = -x$ tangentially.
(c) $\mu > 0, a < -1$, $x = 0$ is unstable, a stable period-2 orbit appears.



(d) The orbit quickly converges to $x = 0$.
(e) The fixed point preserves non-linear stability, despite being linearly indeterminate.
(f) The orbit is repelled from $x = 0$ and attracted to the period-2 orbit.

FIGURE 4.2.4. The supercritical flip bifurcation for the map $f_\mu(x) = -(1 + \mu)x + x^3$.

(b) 2 solutions (4.2.1) exist (y_1 and y_2, say) and at these points we have:

$$\left|\frac{d}{dy}\left(a^2 y + a\beta(1 + a^2)y^3\right)\right|_{y=y_{1,2}} > 1;$$

i.e., the 2 points y_1 and y_2 comprise a repelling period-2 orbit.

The nature of the flip bifurcation can also be related to the *Schwarzian derivative* of the map F_μ, denoted by $D_s\{F_\mu\}$ and defined by

$$D_s\{F_\mu\} = \frac{F_\mu'''}{F_\mu'} - \frac{3}{2}\left(\frac{F_\mu''}{F_\mu'}\right)^2.$$

It follows from the expansion of $F_\mu(x) = ax + bx^2 + cx^3 + \cdots$ and $a(\mu_c) = -1$ that

$$D_s\{F_{\mu_c}\}(0) = -6\left[c(\mu_c) + (b(\mu_c))^2\right] = -6\beta(\mu_c).$$

Hence, the flip bifurcation is:

- supercritical if $D_s\{F_{\mu_c}\}(0) < 0$
- subcritical if $D_s\{F_{\mu_c}\}(0) > 0$

It is often simpler to determine the nature of a flip bifurcation by calculating the Schwarzian derivative of $F_\mu(x)$ directly rather than by expanding $F_\mu(x)$ in power series. The Schwarzian derivative may also be used in the detailed study of the logistic map (see below) and other similar maps.

4.3. Logistic map: period-doubling cascade and chaos

So far we have studied systems whose dynamics is relatively simple. Here we show how complicated dynamics can emerge after a succession of bifurcations. We consider the logistic map

$$x_{n+1} = F_\mu(x_n) = \mu x_n(1 - x_n)$$

for $\mu \in (0, 4]$ (which implies $F_\mu : [0, 1] \to [0, 1]$) and study the changes in the dynamics as μ increases from 0; see figure 4.3.5. Note that the map is non-invertible.

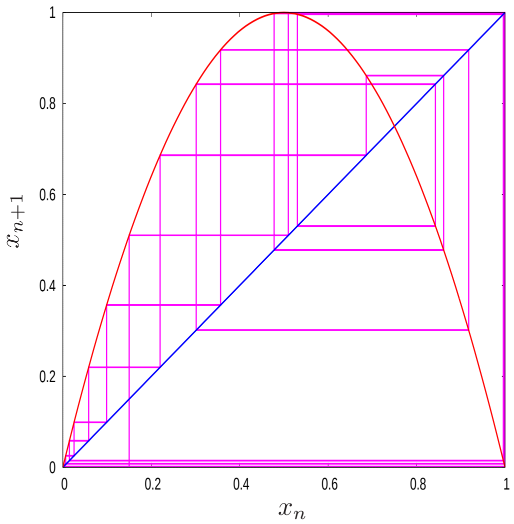


FIGURE 4.3.5. Cobweb diagram for the logistic map with $\mu = 4$.

The map has two fixed points: $x^*_{(i)} = 0$ and $x^*_{(ii)} = (\mu - 1)/\mu$, the latter being in [0,1] for $\mu \geq 1$ only. The stability of the fixed points depends on $F'_\mu(x) = \mu(1 - 2x)$ evaluated at the fixed point. It is found that $x^*_{(i)}$ is stable for $0 < \mu < 1$, and $x^*_{(ii)}$ is stable for $1 < \mu < 3$. In these ranges of μ, the stable fixed points are attractors; in fact they are attracting the whole interval [0,1] minus the unstable fixed point.

For $\mu = \mu_1 = 3$, we have $F'_{\mu_1}(x^*_{(ii)}) = -1$, thus there is a flip bifurcation. To test whether it is super- or subcritical using the theory of the previous section, we calculate the Schwarzian derivative of the map:

$$D_s\{F_\mu\}(x) = -\frac{6}{(1 - 2x)^2} < 0.$$

Thus the flip bifurcation is supercritical and when $x^*_{(ii)}$ becomes unstable for $\mu > \mu_1$ a stable period-2 orbit appears in the vicinity of $x^*_{(ii)}$. In fact, this orbit can be found explicitly. It is a fixed point of the iterated map

$$F^2_\mu(x) = \mu^2 x(1 - x)(1 - \mu x(1 - x)).$$

Solving for fixed points we find the four solutions

$$\left.\begin{array}{l} x = x^*_{(i)} \\ x = x^*_{(ii)} \\ x = x_{(iii)} = \frac{\mu+1+[(\mu+1)(\mu-3)]^{1/2}}{2\mu} \\ x = x_{(iv)} = \frac{\mu+1-[(\mu+1)(\mu-3)]^{1/2}}{2\mu} \end{array}\right\}.$$

The first two, $x^*_{(i)}$ and $x^*_{(ii)}$, are fixed points of F_μ; the other two, $x_{(iii)}$ and $x_{(iv)}$, are such that $F(x_{(iii)}) = x_{(iv)}$ and thus form the period-2 orbit. As expected, this orbit exists for $\mu > 3$ only.

The stability of this orbit can be studied by computing

$$\frac{d}{dx} F_\mu(F_\mu(x))\bigg|_{x=x_{(iii)}} = F'_\mu(x_{(iv)})F'_\mu(x_{(iii)}) = 4 + 2\mu - \mu^2.$$

This is equal to -1 for $\mu = 1 + \sqrt{6}$. Thus $x_{(iii)}$, regarded as a fixed point of F_μ^2, becomes unstable through a flip bifurcation for $\mu = \mu_2 = 1 + \sqrt{6} = 3.44949\cdots$. It can be verified that it is again a supercritical bifurcation, so that for μ greater than μ_2 a stable period-2 orbits of F_μ^2 appears. This orbit is a period-4 orbit of F_μ (i.e., of the original map), and it is attracting. Thus at $\mu = \mu_2$, the period-2 orbit becomes unstable and it is replaced as an attractor by a period-4 orbit; this is the *period-doubling* phenomenon.

A detailed study of the sequence of bifurcations of the logistic map can be performed using numerical simulations. Typical results of simulations are shown in figure 4.3.6; this figure displays as a function of μ a number of iterates $(x_N, x_{N+1}, \cdots)$ obtained from an initial value $x_0 = 1/2$. (The first iterates $(x_0, x_1, \cdots, x_{N-1})$ for $N = 300$, say, are not plotted, so that the plotted iterates can be assumed to have converged to the set attracting most orbits.) `Worksheet 5`

Figure 4.3.7 shows a zoom on a smaller range of values of μ revealing that successive period-doubling bifurcations take place as μ is increased. Further zooms allow one to identify a sequence μ_n, $n = 1, 2, \cdots$ such that a stable period-2^n orbit exists for $\mu_n < \mu < \mu_{n+1}$, becomes unstable at μ_{n+1} when a stable period-2^{n+1} orbit appears. This is the *period-doubling cascade*, which exists for a wide class of mappings including the logistic map. (Note that when a stable period-2^n orbit exists, all the orbits with period 2^i, $i = 1, \cdots, n-1$ also exist; however, they are repelling and have thus a secondary importance for the dynamics.)

Importantly, the sequence μ_n of period doubling parameters converges to a finite $\mu_\infty = 3.5700\cdots$. Thus the period-doubling cascade occupies only a finite range of μ. The convergence is of the form $\mu_n = \mu_\infty - A\delta^{-n}$, for some constants A, δ, so that

$$\lim_{n\to\infty} \frac{\mu_{n-1} - \mu_n}{\mu_n - \mu_{n+1}} = \delta.$$

A remarkable observation due to Feigenbaum and Coullet & Tresser is that the constant $\delta = 4.6692$ for a wide range of different mappings F_μ. Thus, the period-doubling cascade and the scaling of the bifurcation parameters are generic to a wide class of DS and in that sense are universal.

For $\mu > \mu_\infty$, the dynamics become very complicated for most μ. However, for certain ranges of μ, periodic orbits (e.g., with period 3 or 5) are again attracting; see figure 4.3.8. There is an infinite number of repelling periodic orbits, and for most initial conditions the orbits behave in an irregular way, with no discernible long-time patterns. The logistic map for $\mu > \mu_\infty$ is thus an example of a *chaotic system.*

Several qualitative features of the transition to chaos[3] of the logistic map can be understood theoretically. These features are in fact common to a range of maps (typically continuous, one-dimensional maps with a unique maximum or minimum); you can carry out numerical experiments with `Worksheet 5` for a variety of maps to confirm this. Here we present three theoretical results which help explain the behaviour of these maps. Later we will explore further these matters using *kneading theory.*

Result 1. Firstly, we have:

Theorem 4.3.1. *For each μ, there exists at most one attracting periodic orbit for the logistic mapping.*

Thus, we find in numerical experiments that the attractor for the logistic map is essentially independent of the initial value chosen for the iterations. In fact, the above theorem can be proved for a wide class of mappings F_μ which have the property that the Schwarzian derivative of F_μ is negative for all x.

Result 2. Secondly, all period-doubling bifurcations (which we regard as flip bifurcations of the iterated maps F^{2^n}) are supercritical, and therefore lead to the appearance of stable orbits of periods 2^{n+1}. This follows from the fact that $D_s\{F_\mu\} < 0$ and from the following result:

---

[3] No definition of the term *chaos* is universally accepted as yet. However, a useful working definition is: **Chaos** is *aperiodic long-term behaviour* in a *deterministic system* which exhibits *sensitive dependence on initial conditions.*

(i) *Aperiodic long-term behaviour* means that there are orbits (trajectories) which do not eventually settle down to periodic orbits (trajectories) or fixed points.

(ii) *Deterministic* means that the system's irregular behaviour arises from its intrinsic nonlinearity rather than from noisy 'input' data. All systems we consider in this course are determinisitic.

(iii) *Sensitive dependence on initial conditions* means that orbits (trajectories) which are initially close together will ultimately diverge. We discuss this matter in Section 4.4, in terms of Liapunov exponents.

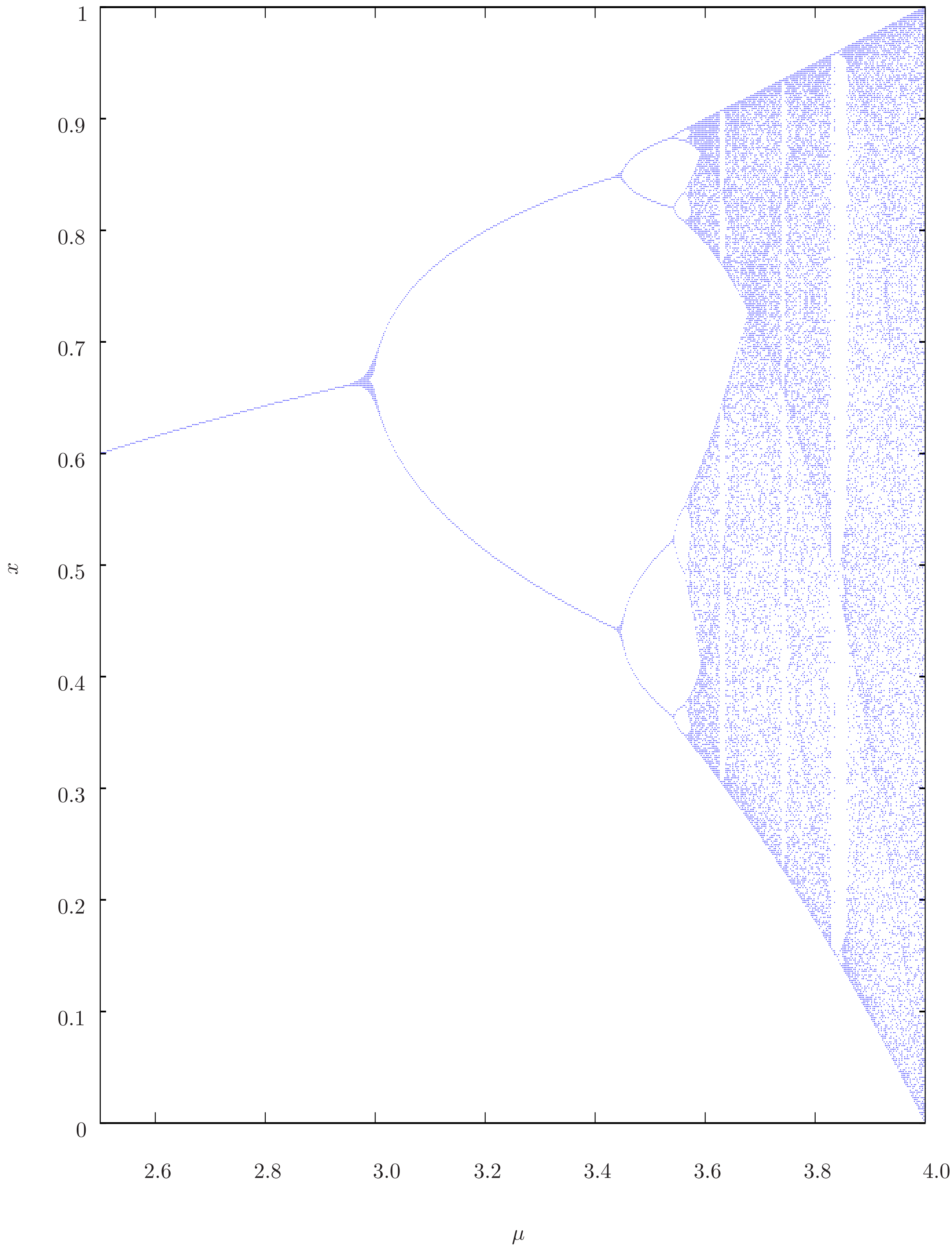


FIGURE 4.3.6. Period-doubling bifurcations for the logistic map: x (vertical axis) as a function of μ (horizontal axis). The stable period 2^n periodic orbits are found by following the orbit x_n with $x_0 = 0.1$ for a large number of iterations.

Theorem 4.3.2. *If a mapping F satisfies $D_s\{F\} < 0$, then the iterated mappings F^k for $k = 2, 3, 4, \ldots$ satisfy $D_s\{F^k\} < 0$.*

This theorem is a consequence of the property satisfied by the Schwarzian derivative of the composition of two functions, $P(x)$ and $Q(x)$ say, namely

$$D_s\{P \circ Q\}(x) = [Q'(x)]^2 \, D_s\{P\}\,(Q(x)) + D_s\{Q\}(x)$$

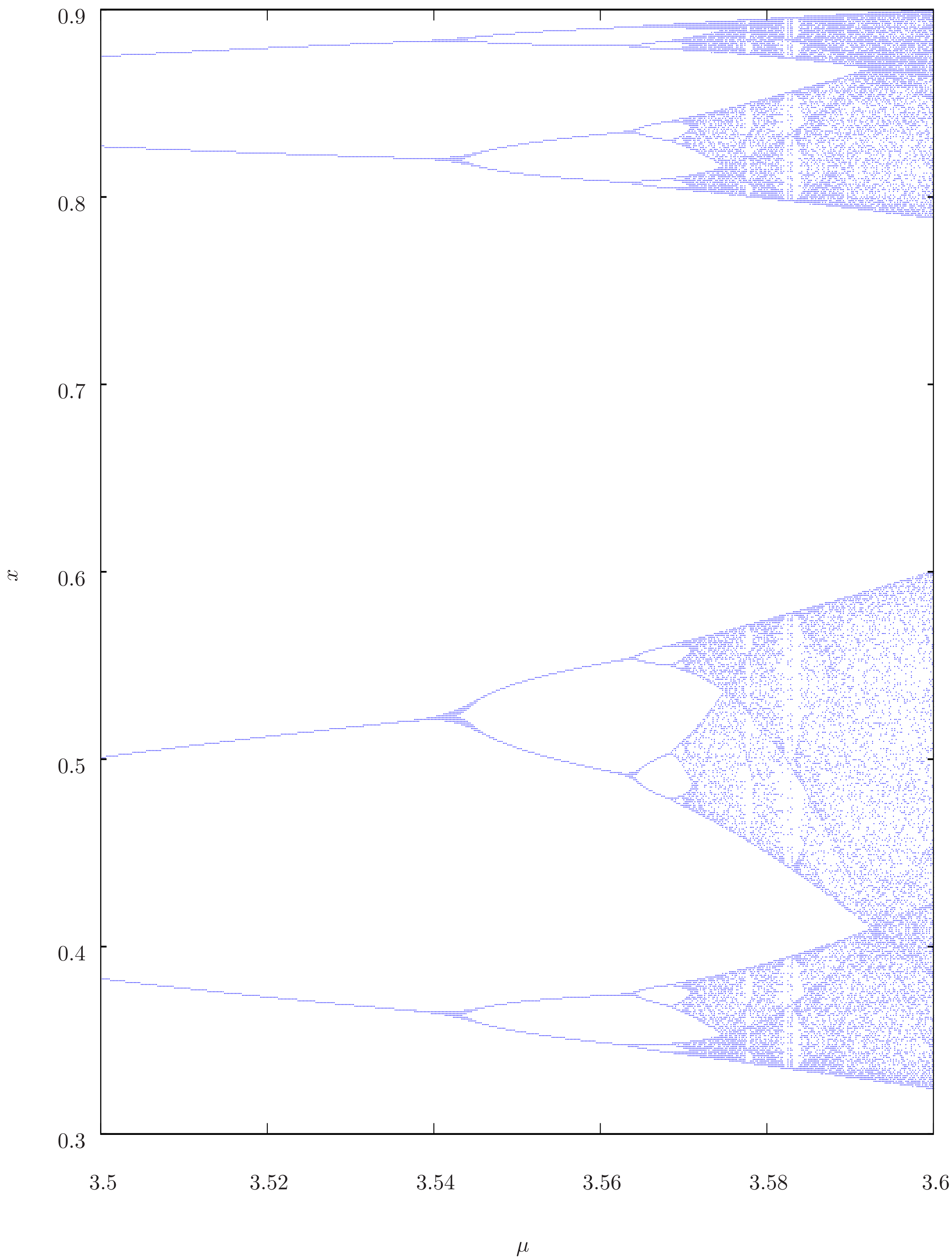


FIGURE 4.3.7. Same as figure 4.3.6 but for a smaller range of values of μ.

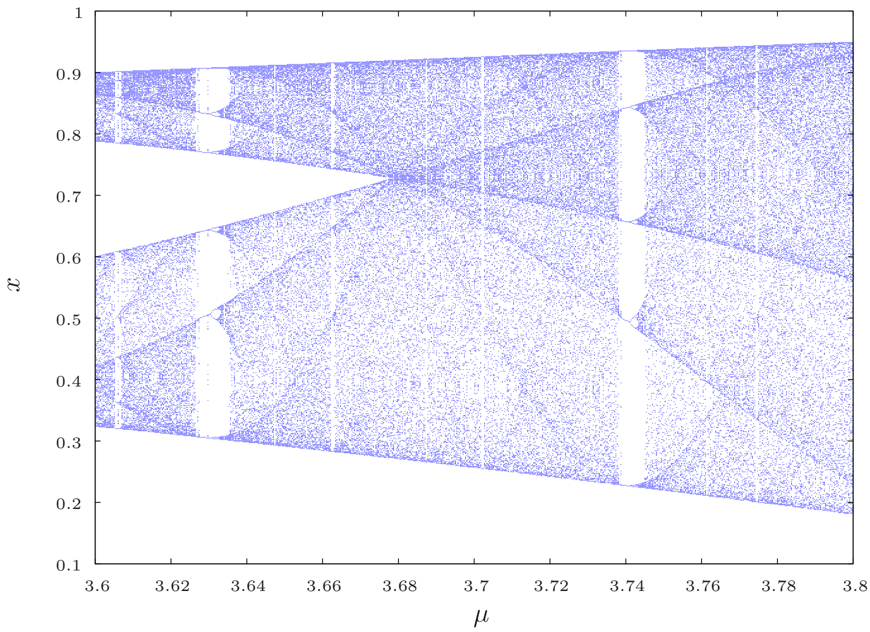


FIGURE 4.3.8. 'Chaotic region' for the logistic map with $x_0 = 0.1$; x (vertical axis) as a function of μ (horizontal axis). Note the windows of attracting periodic orbits.

(check this as an exercise). It follows that if $D_s\{P\} < 0$ and $D_s\{Q\} < 0$, then $D_s\{P \circ Q\} < 0$; in particular $D_s\{F_\mu^k\} < 0$ if $D_s\{F_\mu\} < 0$ It is then clear that all the flip bifurcations of $F_\mu^{2^n}$ are supercritical since $D_s\{F_\mu^{2^n}\} < 0$.

Result 3. Thirdly, we examine the order of the appearance of periodic orbits for the logistic map. In fact, there is a natural ordering of the possible periodic orbits which is described by the following important result:

Theorem 4.3.3 (Sarkovskii's theorem.). *Consider the following ordering of the natural numbers:*

$$3 \rhd 5 \rhd 7 \rhd \cdots \rhd 2 \cdot 3 \rhd 2 \cdot 5 \rhd 2 \cdot 7 \rhd \cdots \rhd 2^2 \cdot 3 \rhd 2^2 \cdot 5 \rhd 2^2 \cdot 7 \rhd \cdots$$
$$\cdots \rhd 2^3 \cdot 3 \rhd 2^3 \cdot 5 \rhd 2^3 \cdot 7 \rhd \cdots \rhd 2^n \rhd 2^{n-1} \rhd \cdots \rhd 2^3 \rhd 2^2 \rhd 2 \rhd 1.$$

If a continuous map F has a prime period-k orbit, then it has prime period-n orbits for all $n \lhd k$.

We remark that there is a converse to Sarkovskii's theorem: there are maps with period-k orbits and no period-l orbits for all $l \rhd k$. We prove the following corollary to Sarkovskii's theorem:

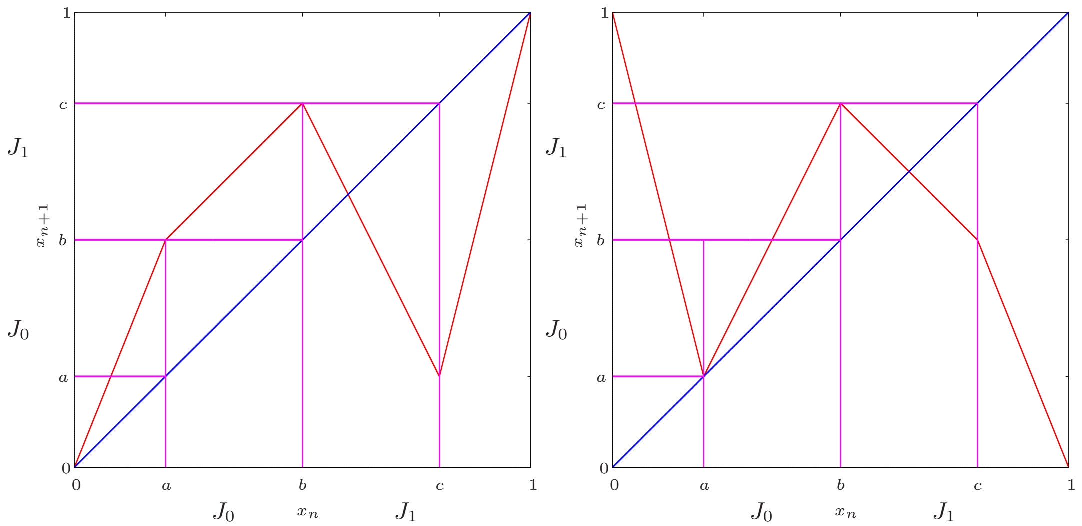
Corollary 4.3.1. *If a continous map F has a prime period-3 orbit, then it has prime period-n orbits for all $n \in \mathbb{N}$.*

PROOF. To prove this, we consider the mapping F with period-3 orbit (a, b, c); i.e., we have $F(a) = b, F(b) = c, F(c) = a$. We shall assume that $a < b < c$ (the case $a < c < b$ is treated similarly, see figure 4.3.9).

Let us define $I_0 = [a, b]$ and $I_1 = [b, c]$. Four observations are used in the proof:

(i) $F(I_0) \supseteq I_1$.
(ii) $F(I_1) \supseteq I_0 \cup I_1$.
(iii) If I is a closed interval and $F(I) \supseteq I$, then F has a fixed point in I.
(iv) Suppose I, J are closed intervals. If $F(I) \supseteq J$, then there exists a closed interval $K \subseteq I$ such that $F(K) = J$.

The last two observations can be established using the intermediate value theorem. We start the proof by noting that (ii) and (iii) imply that F has a fixed point in I_1. Also, (i), (ii) and (iii) imply that F^2 has a fixed point in I_0, so that F has a period-2 orbit. Thus, the $n = 1$ and $n = 2$ cases are proven and



(a) The first case, with $J_0 \to J_1$ and $J_1 \to J_0 \cup J_1$. (b) The second case, with $J_1 \to J_0$ and $J_0 \to J_0 \cup J_1$.

(c) The two cases: left is (a) above; right is (b) above.

FIGURE 4.3.9. The two cases of the "prime period-3" implies all prime periods theorem.

henceforth we assume $n > 3$. Now we construct a nested sequence of closed intervals A_n: let $A_0 = I_1$, (ii) and (iv) imply that there is a $A_1 \subseteq A_0$ with $F(A_1) = A_0 = I_1$. Similarly, there is a $A_2 \subseteq A_1$ with $F(A_2) = A_1$ and so $F^2(A_2) = A_0$. Proceeding similarly, the sequence

$$A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots \supseteq A_{n-2}, \quad \text{with} \quad F^k(A_k) = A_0, \quad k = 1, 2, \dots, n-2,$$

can be constructed. The next interval in the sequence, A_{n-1} is constructed by noting that $F^{n-1}(A_{n-2}) = F(A_0) \supseteq I_0$ (using (ii)). Then, (iv) implies that there is a $A_{n-1} \subseteq A_{n-2}$ with $F^{n-1}(A_{n-1}) = I_0$. Finally since $F^n(A_{n-1}) = F(I_0) \supseteq I_1$ (using (i)), there exists a $A_n \subseteq A_{n-1}$ with $F^n(A_n) = A_0 = I_1$. Now, by construction $A_n \subseteq A_0$, so that $F^n(A_n) \supseteq A_n$. So (iii) then implies that there exists a fixed point $x^\star \in A_n$ with $F^n(x^\star) = x^\star$. This is a prime period-n point unless it is also fixed point of F^k for $k < n$. But this is impossible since $x^\star \in A_k$, $k = 0, 1, \cdots, n$ gives that $F^k(x^\star) \in I_1$ for $k = 1, 2, \dots, n-2$ and we also have $F^{n-1}(x^\star) \in I_0$. (The case $F^{n-1}(x^\star) \in I_0 \cap I_1 = \{b\}$ can be excluded since it would imply $n = 3$.) This completes the proof. □

A more detailed understanding of the period-doubling cascade and transition to chaos of the logistic map and analogous maps requires the introduction of the *symbolic dynamics* and *kneading theory*. We start by introducing symbolic dynamics in simpler maps.

4.4. Chaotic systems and symbolic dynamics

Complicated, chaotic behaviour can be found in very simple (nonlinear) systems. In this section, this is illustrated by three systems. We start by the simplest, the times 2 mod 1 map.

4.4.1. Times 2 mod 1 map. The times 2 mod 1 is the map[4]

$$x_{n+1} = 2x_n \bmod 1;$$

[4]'(mod 1)' means we only consider the noninteger part; e.g., $5.43 = 0.43 \bmod 1$.

thus,

$$\left.\begin{array}{lll} x_{n+1} = 2x_n & \text{if} & 0 \leq x_n < 1/2 \\ x_{n+1} = 2x_n - 1 & \text{if} & 1/2 \leq x_n < 1 \end{array}\right\},$$

etc (see figure 4.4.10). Note that for any x_0, we have $x_1 \in [0,1)$ (i.e., $[0,1)$ is an attractor), so we can take $x_0 \in [0,1)$. The effect of the map is seen to be a stretching of $[0,1)$ onto $[0,2)$ followed by a cut. One can also think of the map as acting on the circle S^1, with $2\pi x_n$ as an angle.

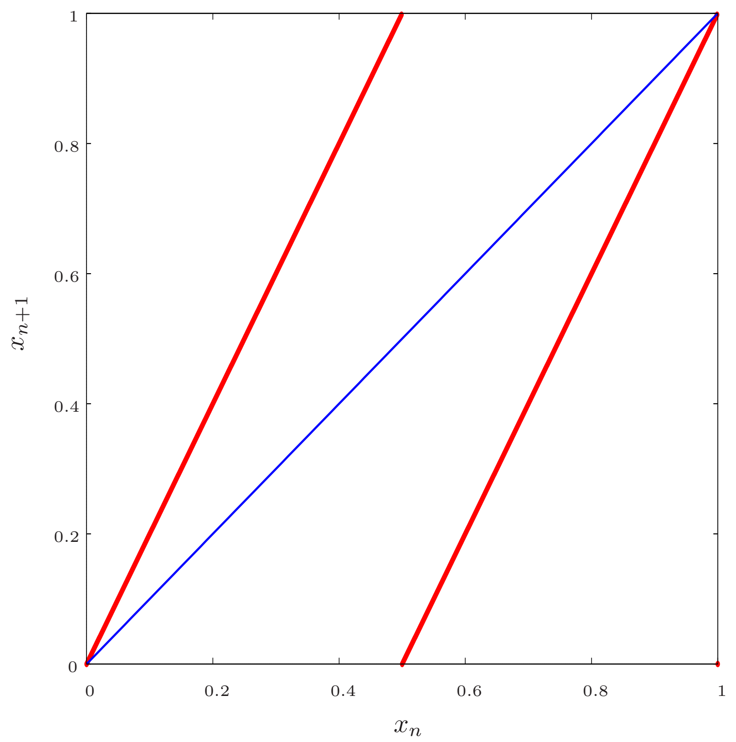


FIGURE 4.4.10. The times 2 mod 1 map.

The map has a single fixed point at $x^* = 0$, which is unstable. If you perform a numerical integration of the map, you will realise that the orbits are complicated, at least for initial conditions with a large number of decimal digits. What is remarkable about the times 2 mod 1 is that its behaviour can be completely understood analytically. Consider the binary expansion of the initial condition

$$x_0 = d_1/2 + d_2/2^2 + d_3/2^3 + \cdots \tag{4.4.1}$$

where $d_j \in \{0,1\}$ for $j = 1,2,3,4,\cdots$. The expansion is unique if we exclude sequences with $d_n = 1$ for all large n. Applying the times 2 mod 1 gives

$$\begin{aligned} x_1 &= 2x_0 \bmod 1 \\ &= d_2/2 + d_3/2^2 + d_4/2^3 + \cdots; \end{aligned}$$

i.e., the binary expansion of x_1 is obtained from that of x_0 by removing the first digit d_1 and shifting all the remaining digits to the left. This procedure allows us to construct all the x_n once x_0 has been written as a binary expansion. It is easy then to predict the nature of the orbit associated with a given x_0. If x_0 has an expansion (4.4.1) of finite length N, then $x_n = 0$, $\forall n > N$; this is the case when $x_0 = p/2^N$ for some odd $p \in \mathbb{N}$ (x_0 then belongs to the stable manifold of the origin). If x_0 is rational but has an infinite binary expansion, then the expansion is eventually periodic, so the orbit is also eventually periodic. If x_0 is irrational, then the orbit is aperiodic. Because there are irrational numbers arbitrarily close the rational numbers, the periodic orbits are unstable.

4.4.2. Sensitive dependence on initial conditions. An important property of chaotic systems appears clearly in the times 2 mod 1: namely, the *sensitive dependence on initial conditions*. This means that orbits with starting points which are close together will ultimately diverge after a sufficient number of iterations of the map; thus the result after a large number of iterations depends sensitively on the exact initial condition. This is seen clearly by considering two initial conditions x_0 and x_0'. If they are close, their binary expansions are identical up to the N^{th} term, for a large N, but different thereafter. After N iterations, the binary expansions of x_N and x_N' will differ from the first term, i.e., will be completely different. In practice, sensitive dependence on initial conditions means that a system, although deterministic, is unpredictable, since the initial conditions are never known perfectly. This has important consequences in many areas of sciences, e.g., in meteorology.

More formally, we may define an invariant set S of a map $F : \mathbb{R} \to \mathbb{R}$ to have *sensitive dependence on initial conditions* if $\exists r > 0$ such that for each $x \in S$ and for each $\epsilon > 0$, there is a $y \in S$ with $|x - y| < \epsilon$ and a $k > 0$ with $|F^k(x) - F^k(y)| \geq r$. So, neighbouring orbits, which are arbitrarily close initially, eventually separate from one another, although each keeps close to the invariant set.

4.4.3. Liapunov exponent. The sensitive dependence on initial conditions is often quantified using the Liapunov exponent. The Liapunov exponent λ measures the (exponential) rate of divergence of orbits.

Here is the intuition: Given some initial condition x_0, consider a nearby point $x_0 + \delta_0$, where the initial separation δ_0 is extremely small. Let δ_n be the separation after n iterates. If $|\delta_n| \approx |\delta_0| \exp(n\lambda)$, then λ is called the Liapunov exponent. A positive Liapunov exponent indicates sensitive dependence on initial conditions and is a signature of chaos.

A more precise and useful formula for λ is derived as follows: By taking logarithms and noting that $\delta_n = F^n(x_0 + \delta_0) - F^n(x_0)$, we obtain

$$\begin{aligned} \lambda &\approx \frac{1}{n} \ln \left| \frac{\delta_n}{\delta_0} \right| \\ &= \frac{1}{n} \ln \left| \frac{F^n(x_0 + \delta_0) - F^n(x_0)}{\delta_0} \right| \\ &= \frac{1}{n} \ln \left| (F^n)'(x_0) \right| \end{aligned}$$

where the limit $\delta_0 \to 0$ has been taken in the last step. Now, the chain rule gives us

$$(F^n)'(x_0) = \prod_{i=0}^{n-1} F'(x_i),$$

thus we may write

$$\lambda \approx \frac{1}{n} \ln \left| \prod_{i=0}^{n-1} F'(x_i) \right| = \frac{1}{n} \sum_{i=0}^{n-1} \ln |F'(x_i)| .$$

If this expression has a limit as $n \to \infty$, then we define that limit to be the Liapunov exponent for the orbit starting at x_0 :

$$\lambda = \lim_{n \to \infty} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} \ln |F'(x_i)| \right\} .$$

Observe that λ depends on x_0. However, it is the same for all x_0 in the basin of attraction for a given attractor. For the times 2 mod 1 it can be checked that $\lambda = \ln 2$. In general, numerical calculation is required to compute λ.

4.4.4. Transitivity and chaos. A map F acting on set S is said to be *transitive* if there is an orbit $(x_0, x_1, x_2, \cdots)$ which is *dense* in S ; i.e., if the iterates $F^k(x_0)$ come arbirarily close to any point in S. Transitivity implies a certain complexity of the dynamics. It is sometimes used to define chaos (see earlier footnote on definition of chaos): A map may be said to be chaotic if it has sensitive dependence on initial conditions and it is transitive. An alternative definition is that a map is said to be chaotic if it has a dense set of periodic points and is transitive. (This can be shown to imply that the map has sensitive dependence on initial conditions, so that this second definition is stronger than the first.)

4.4.5. Symbolic dynamics. To capture the essence of the dynamics of a chaotic system, it is sometimes useful to consider a very simple partition of the phase space and to study how the orbits wander from one part to the other. More precisely, it is an invariant set S of the state space which is partioned. To keep track of the location of the orbit in the partition, each component part of the partition is denoted by a symbol, usually an integer. Thus, one can attribute a symbol to any given point of S depending on which part of S it belongs to. For example, a useful partition of the invariant set $S = [0, 1)$ of the times 2 mod 1 is $S_1 \cup S_2$, where $S_1 = [0, 1/2)$, $S_2 = [1/2, 1)$, and the symbol $j = 1, 2$ can be attached to points inside S_j.

To a sequence of points constituting an orbit $(x_0, x_1, x_2, \cdots)$, we can associate a sequence of symbols $(j_0, j_1, j_2, \cdots)$ such that $x_n \in S_{j_n}$, $n = 0, 1, 2, \cdots$. More formally, we can define a map $h : S \to \Sigma$, where Σ is the space of sequences of symbols (e.g., $\Sigma = \{1, 2\}^{\mathbb{N}}$ for the Bernoulli shift map; $\Sigma = \{1, 2, \cdots, p\}^{\mathbb{N}}$ when there are p different symbols), as follows:

$$h(x_0) = (j_0, j_1, j_2, \cdots), \quad \text{with} \quad x_0 \in S_{j_0} \quad \text{and} \quad F^n(x_0) \in S_{j_n} \quad (n = 1, 2, 3, \cdots).$$

The map h is called the *itinerary map* since it gives, for specified initial conditions, the successive locations of x_n in the partition. By definition, we see

$$h(x_0) = (j_0, j_1, j_2, \cdots) \Rightarrow h(F(x_0)) = (j_1, j_2, j_3, \cdots);$$

that is, the action of F on a sequence of symbols is simply the backward shift of the sequence. Formally, one writes

$$h \circ F = \sigma \circ h,$$

where σ represents the backward shift.

Now, one can get an idea of how complicated the dynamics generated by F are by analysing which sequences of symbols (or 'words') can be realised; i.e., by asking the following question: given a sequence $(j_0, j_1, j_2, \cdots) \in \Sigma$, does there exist an initial condition x_0 such that $h(x_0) = (j_0, j_1, j_2, \cdots)$? If this is true for all sequences of Σ (i.e., if h is surjective), then the complexity of the dynamics (as reflected by the symbolic dynamics) is maximum. Often, not all sequences of Σ can be realised, but only the ones which obey certain rules (i.e., some 'words' are forbidden). For example, one can imagine a partition of S into 3 parts ($\Sigma = \{1, 2, 3\}^{\mathbb{N}}$) and a dynamics that does not allow two consecutive symbols to be the same. If the rules restricting the sequences are known, one can then consider a restricted set $\Sigma' \subset \Sigma$ that contains only the allowed sequences, so that $h : S \to \Sigma'$ is surjective. The complexity of the dynamics then depends on whether Σ' contains sequences that are non-trivial (e.g., periodic repetition of symbols).

The symbolic dynamics can be simpler than the original dynamics if different initial conditions lead to the same sequences of symbol. A challenge is thus to find the partition S_j such that h is one-to-one (and thus a bijection from S to Σ'). When this is possible, one has a very good description of the dynamics generated by F.

For the times 2 mod 1, the symbolic dynamics are easily constructed. Recall the binary expansion of initial condition: $x_0 = d_1/2 + d_2/2^2 + d_3/2^3 + \cdots$. The first digit in the binary expansion of x_n gives its position in either S_1 and S_2. Since the first digit in the binary expansion of x_n is the nth digit in the binary expansion of x_0, we have

$$h(x_0) = (d_1 + 1, d_2 + 1, d_3 + 1, \cdots).$$

If we exclude from Σ sequences that end with an infinite number of 2's, h is bijective: any sequence of 1's and 2's corresponds to a unique initial condition x_0.

For the times 2 mod 1, the invariant set S is simply the whole interval $[0, 1)$. In the next section we consider a system which has an invariant set of a different type.

4.4.6. Logistic map with $\mu > 4$. We consider the logistic map

$$x_{n+1} = F(x_n) = \mu x_n(1 - x_n)$$

for $\mu > 4$ and restrict our attention to $x_n \in [0, 1]$. Since $\mu x(1 - x) > 1$ for $x \in [a, b]$ for some $0 < a < b < 1$, points in $[a, b]$ leave the interval $[0, 1]$. On the other hand, points of $I_1 = [0, a]$ and $I_2 = [b, 1]$ are mapped into the interval $[0, 1]$ (see figure 4.4.11).

Before introducing the appropriate symbolic dynamics, we need to study the invariant set S for the map. Clearly, this set consists of the points x for which all the successive iterates $F^n(x)$ are either in

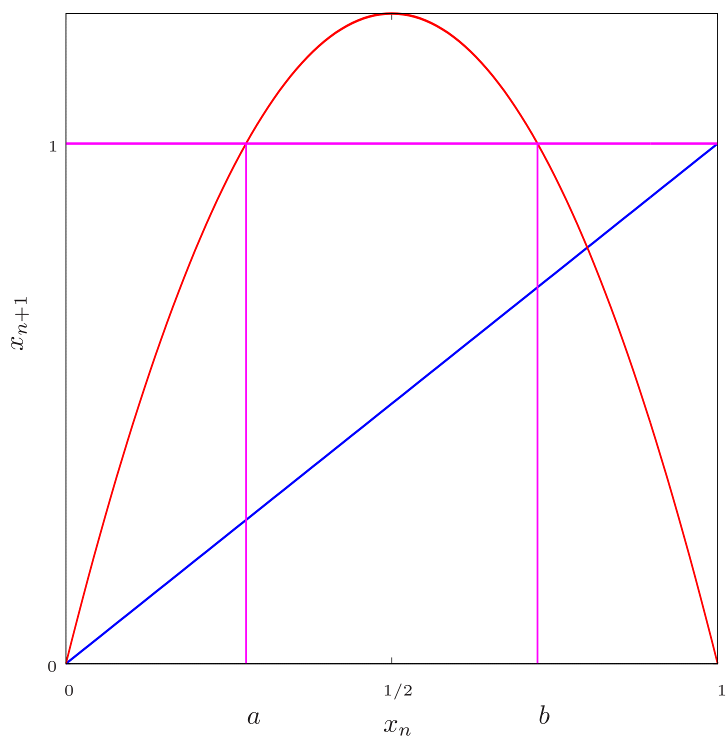


FIGURE 4.4.11. Logistic map for $\mu > 4$. The intervals $[0, a]$ and $[b, 1]$ are mapped onto $[0, 1]$.

I_1 or in I_2. Let

$$I_{j_0 j_1 j_2 \cdots j_n} = \{x : x \in I_{j_0} \text{ and } F^k(x) \in I_{j_k} \text{ for } 1 \leq k \leq n\},$$

where $j_k \in \{1, 2\}$. The first step leads to the construction of:

$$\begin{aligned} I_{11} &= \{x : x \in I_1 \text{ and } F(x) \in I_1\} \\ I_{12} &= \{x : x \in I_1 \text{ and } F(x) \in I_2\} \\ I_{21} &= \{x : x \in I_2 \text{ and } F(x) \in I_1\} \\ I_{22} &= \{x : x \in I_2 \text{ and } F(x) \in I_2\} \end{aligned}$$

with $I_{11}, I_{12} \subset I_1$ and $I_{21}, I_{22} \subset I_2$. Each successive stage multiplies the number of intervals by 2. We can define

$$S_n = \bigcup_{j_0, j_1, \cdots, j_n = 1,2} I_{j_0 j_1 j_2 \cdots j_n},$$

which is the set of points remaining in $[0, 1]$ after n iterations. The size of S_n decreases at each step; i.e, we have $S_n \subset S_{n-1}$. Then the invariant set is $S = \cap_{n=0}^{\infty} S_n$. It turns out that this set is a *Cantor set* (equivalent to the famous *Cantor middle-third set*)[5]. It is on this set that the symbolic dynamics can be defined.

The symbolic dynamics are defined by associating to any $x_0 \in S$ the sequence of symbols $(j_0, j_1, j_2, \cdots)$ where $j_k \in \{1, 2\}$ and $x_k = F^k(x_0) \in I_{j_k}$. Thus, the sequence of symbols records the passage of the successive iterates of x_0 in I_1 or I_2. Formally, this means that the function $h : S \to \Sigma = \{1, 2\}^{\mathbb{N}}$ is defined by

$$h(x_0) = (j_0, j_1, j_2, \cdots),$$

with

$$\left. \begin{array}{l} x_0 \in I_{j_0} \\ x_k = F^k(x_0) \in I_{j_k} \\ j_k \in \{1, 2\} \end{array} \right\}, \qquad (k \in \mathbb{N}).$$

[5]i.e., a *fractal set* — to be discussed later in Section 6

Equivalently, $h(x_0) = (j_0, j_1, j_2, \cdots)$ may be defined with

$$\left.\begin{array}{l} x_0 \in I_{j_0 j_1 j_2 \cdots} \\ j_k \in \{1,2\} \end{array}\right\}, \qquad (k \in \mathbb{N}).$$

Clearly, $h \circ F = \sigma \circ h$. We now show that h is:

(i) surjective

Consider the sequence of symbols $(j_0, j_1, j_2, \cdots) \in \Sigma$. We form the sequence of sets $J_n = I_{j_0 j_1 j_2 \cdots j_n}$. Observe that $J_0 \supset J_1 \supset J_2 \supset \cdots \supset J_n$. Since the J_n form a nested family of closed intervals, by the Nested Intervals Theorem[6], $\exists x_0 \in \cap_{n=0}^{\infty} J_n$. Thus, $h(x_0) = (j_0, j_1, j_2, \cdots)$. Hence, h is surjective.

(ii) bijective, under the additional hypothesis $|F'(x)| > 1$ in $I_1 \cup I_2$ (i.e., $\mu > 2 + \sqrt{5}$)

We prove that h is injective (one-to-one). Firstly, observe that $|\left(F^k\right)'(x)| > 1$ in S, since $(F^n)'(x_0) = \prod_{i=0}^{n-1} F'(x_i)$ and we have assumed $|F'(x)| > 1$ in $I_1 \cup I_2$. Secondly, suppose there exists $x, y \in S$ and $h(x) = h(y)$. This implies that $F^k(x)$ and $F^k(y)$ are in the same subinterval of the pair I_1, I_2. But, from $|F'| \geq q > 1$, we find that

$$\left| \frac{F^{k+1}(y) - F^{k+1}(x)}{F^k(y) - F^k(x)} \right| \geq q$$

i.e.,

$$\begin{aligned} \left|F^{k+1}(y) - F^{k+1}(x)\right| &\geq q\left|F^k(y) - F^k(x)\right| \\ &\geq q^2\left|F^{k-1}(y) - F^{k-1}(x)\right| \\ &\geq q^3\left|F^{k-2}(y) - F^{k-2}(x)\right| \\ &\vdots \\ &\geq q^{k+1}|y - x|. \end{aligned}$$

But $|F^{k+1}(y) - F^{k+1}(x)| \leq 1$ (since $x, y \in S$). So

$$|y - x| \leq \frac{1}{q^{k+1}}$$

which implies that $x = y$. Hence, h is injective and therefore, by (i), h is bijective.

Since there is a bijection between F and σ, the qualitative properties of the logistic map (with $\mu > 4$) can be derived from those of the shift. In particular, the number of periodic orbits of a given period can be studied by considering recurrent successions of the symbols 1, 2. For instance, consider orbits of period 5; they are represented by sequences of symbols which repeat when shifted by 5. There are $2^5 = 32$ ways to choose such sequences, which are thus fixed points of F^5; two of those correspond to fixed points ($11111\cdots$ and $22222\cdots$), so that there are 30 points of prime period 5. Each periodic orbit contains 5 points, hence there are 6 different orbits of period 5.

The fact that the map F has a dense set of periodic points and is transitive (and so is chaotic) can also be deduced from the fact that these properties hold for the shift σ acting on Σ. To prove these properties for σ we need to define an appropriate 'distance' between elements of Σ. Consider two sequences of symbols $\mathbf{s} = (s_0, s_1, s_2, \cdots) \in \Sigma$ and $\mathbf{t} = (t_0, t_1, t_2, \cdots) \in \Sigma$; the 'distance' between them, $d(\mathbf{s}, \mathbf{t})$, is defined by

$$d(\mathbf{s}, \mathbf{t}) = \sum_{i=0}^{\infty} \frac{|s_i - t_i|}{3^i}.$$

We show that periodic points are dense by constructing a periodic point $\mathbf{t}$ arbitrarily close to any $\mathbf{s} = (s_0, s_1, s_2, \cdots)$. Let $\mathbf{t} = (s_0, s_1, \cdots, s_n, s_0, s_1, \cdots, s_n, \cdots)$. Clearly, $\mathbf{t}$ is periodic and $d(\mathbf{s}, \mathbf{t}) < \epsilon$ for any $\epsilon > 0$ provided that n is large enough.

Transitivity is established by constructing a point $\mathbf{t}$ whose iterates $\sigma^k(\mathbf{t})$ come arbitrarily close to any $\mathbf{s}$. Such a $\mathbf{t}$ is given by

$$\mathbf{t} = (1, 2; 1, 1, 2, 1, 2, 2; 1, 1, 1, 2, 1, 1, 2, 2, 1, 2, 2, 2; \cdots),$$

which lists all possible blocks of one symbol, then (after the ';' used for clarity) all possible blocks of two symbols. (Verify: for any $\epsilon > 0$ there is a $k \geq 0$ such that $d\left(\sigma^k(\mathbf{t}), \mathbf{s}\right) < \epsilon$).

[6] The Nested Intervals Theorem asserts: "Suppose B_i form a nested sequence of closed nonempty intervals. Then $\cap_{i=0}^{\infty} B_i$ is nonempty."

CHAPTER 5

Two-dimensional dynamics

5.1. Introduction

This Chapter considers the global, nonlinear dynamics of two-dimensional systems

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathbb{R}^2, \qquad \text{i.e.,} \quad \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} f(x_n, y_n) \\ g(x_n, y_n) \end{pmatrix}.$$

We assume that the map $\mathbf{F}$ is invertible. The phenomena that will be illustrated also occur in higher-dimensional maps.

From §2 recall that the stability of a fixed point $\mathbf{x}_* = \mathbf{F}(\mathbf{x}_*)$ depends on the eigenvalues of the Jacobian matrix $D\mathbf{F}|_{\mathbf{x}_*}$. The Jacobian matrix is also useful for points which are not fixed: its determinant at some $\mathbf{x}_n$ gives the change in area of a surface element centred on $\mathbf{x}_n$ when it is mapped by $\mathbf{F}$:

$$\mathrm{d}x_{n+1}\mathrm{d}y_{n+1} = \det(D\mathbf{F}|_{\mathbf{x}_n})\mathrm{d}x_n\mathrm{d}y_n.$$

It can also be shown that an infinitesimal circle centred on $\mathbf{x}_n$ is mapped onto an ellipse centred on $\mathbf{x}_{n+1}$ with semi-axes given by the eigenvalues of $(D\mathbf{F}|_{\mathbf{x}_n})^{\mathrm{T}}\, D\mathbf{F}|_{\mathbf{x}_n}$ (times the radius of the circle) and in the direction of the corresponding eigenvectors.

For an *area–preserving* map, we have $|\det(D\mathbf{F})| = 1$ everywhere. Such maps are associated with conservative systems, particularly with the Hamiltonian systems of classical mechanics. Note that area–preservation is incompatible with the existence of an attractor.

5.2. Stable and unstable manifolds

Saddle points and their stable and unstable manifolds are important for sensitive dependence on initial conditions and chaos. In two-dimensional systems, in particular, the behaviour of orbits for $n \to \infty$ can depend crucially on which side of the stable manifold they originate from. Also, the unstable manifold is generally attracting so that orbits diverge to infinity along the unstable manifold for $n \to \infty$.

In section §3 we discussed the local nature of the stable and unstable manifolds in a small neighbourhood of a saddle point $\mathbf{x}_*$. These are global objects, however: let $W^-_{loc}(\mathbf{x}_*)$ be a local unstable manifold in some neighbourhood of the saddle $\mathbf{x}_*$, where it is defined by an expansion $u^+ = a_2(u^-)^2 + a_3(u^-)^3 + \cdots$. Then the unstable manifold can be defined globally as the union of all forward images of $W^-_{loc}(\mathbf{x}_*)$:

$$W^-(\mathbf{x}_*) = \cup_{n=0}^{\infty}\mathbf{F}^n(W^-_{loc}(\mathbf{x}_*)).$$

Similarly, the global stable manifold $W^+(\mathbf{x}_*)$ can be defined as the union of the backwards images of a small (local) segment $W^+_{loc}(\mathbf{x}_*)$ near the saddle point; i.e.,

$$W^+(\mathbf{x}_*) = \cup_{n=0}^{\infty}\mathbf{F}^{-n}(W^+_{loc}(\mathbf{x}_*)).$$

The attracting nature of the unstable manifold for $n \to \infty$, and similarly of the stable manifold for $n \to -\infty$, provides a technique to find numerical approximations to the manifolds. The forward orbits of a set of points along the unstable subspace very near the saddle approximate the unstable manifold; and the backward orbits of a set of points along the stable subspace near the saddle approximate the stable manifold.

Figure 5.2.1 shows approximations to the global stable and unstable manifolds obtained using this technique for the Hénon map

$$x_{n+1} = a - x_n^2 + by_n, \quad y_{n+1} = x_n,$$

with $a = 2.12$ and $b = -0.3$ `Worksheet 6`. The complicated shape of the manifolds, which are convoluted curves, appears in spite of the low resolution.

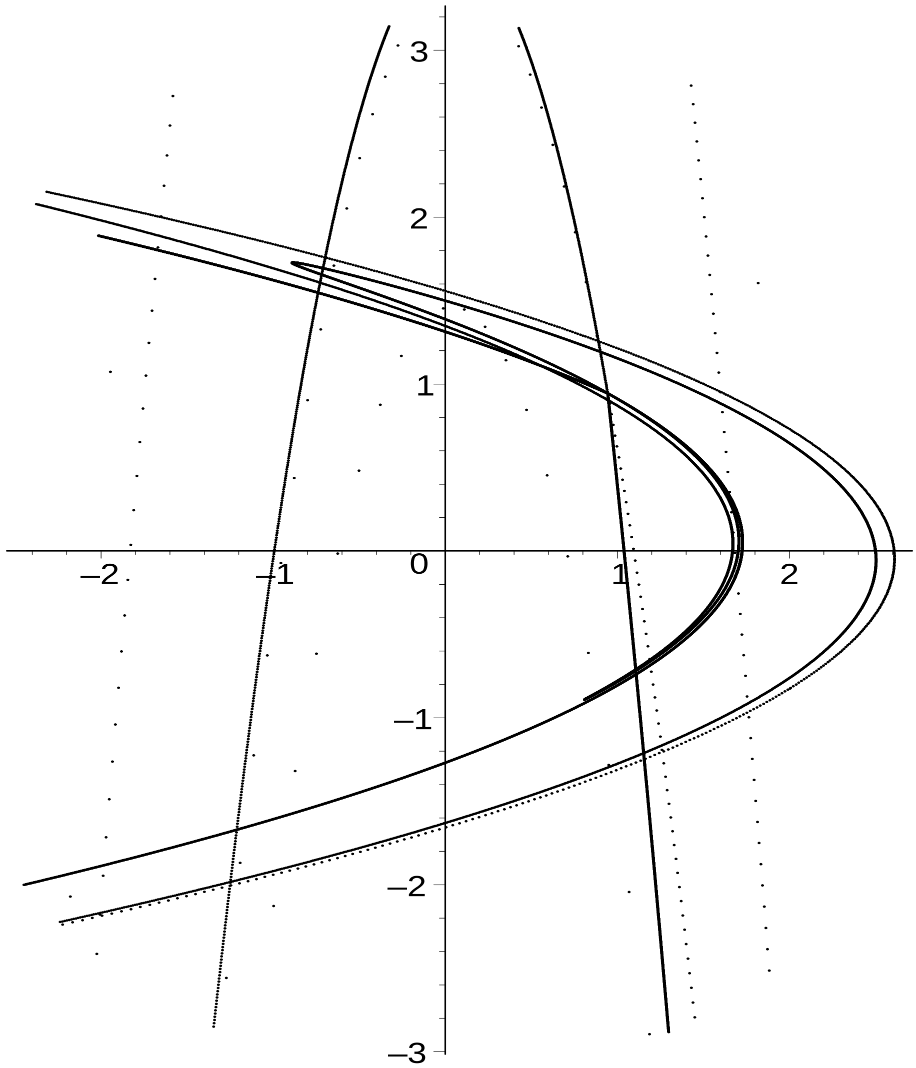


FIGURE 5.2.1. Global stable and unstable manifolds for the Hénon map for $a = 2.12$ and $b = -0.3$: the saddle point has coordinates $\mathbf{x}_* = (0.94, 0.94)$, $W^-(\mathbf{x}_*)$ is the mostly horizontal curve and $W^+(\mathbf{x}_*)$ is the mostly vertical (and poorly resolved) curve.

The following should be noted and can be proved: the manifold $W^-(\mathbf{x}_*)$ does not intersect itself and the manifold $W^+(\mathbf{x}_*)$ does not intersect itself, but $W^-(\mathbf{x}_*)$ and $W^+(\mathbf{x}_*)$ do intersect one another. When stable and unstable manifolds intersect, they do so at a *homoclinic point* $\mathbf{x}_\mathrm{h}$ which satisfies

$$\lim_{n\to\infty} \mathbf{F}^n(\mathbf{x}_\mathrm{h}) = \lim_{n\to\infty} \mathbf{F}^{-n}(\mathbf{x}_\mathrm{h}) = \mathbf{x}_*.$$

If the stable and unstable manifolds intersect at a homoclinic point, they must intersect an infinite number of times since

$$\mathbf{x}_\mathrm{h} \in W^+(\mathbf{x}_*) \cap W^-(\mathbf{x}_*) \quad \Rightarrow \quad \mathbf{F}^k(\mathbf{x}_\mathrm{h}) \in W^+(\mathbf{x}_*) \cap W^-(\mathbf{x}_*), \qquad \text{for } k = \pm 1, \pm 2, \pm 3, \ldots.$$

Because $\mathbf{F}^n(\mathbf{x}_h) \to \mathbf{x}_*$ for $n \to \pm\infty$, the stable and unstable manifold have a complicated form, as illustrated in figure 5.2.2. This is the *homoclinic tangle* discovered by Poincaré.

Stretching and folding. We remarked in §4 that sensitive dependence on initial conditions is a central characteristic of chaotic systems. In particular, we observed that neighbouring orbits diverge exponentially fast from one another (at least initially). How can this be if the orbits remain confined to a bounded region of phase space? The answer to this question lies in the basic mechanism of repeated

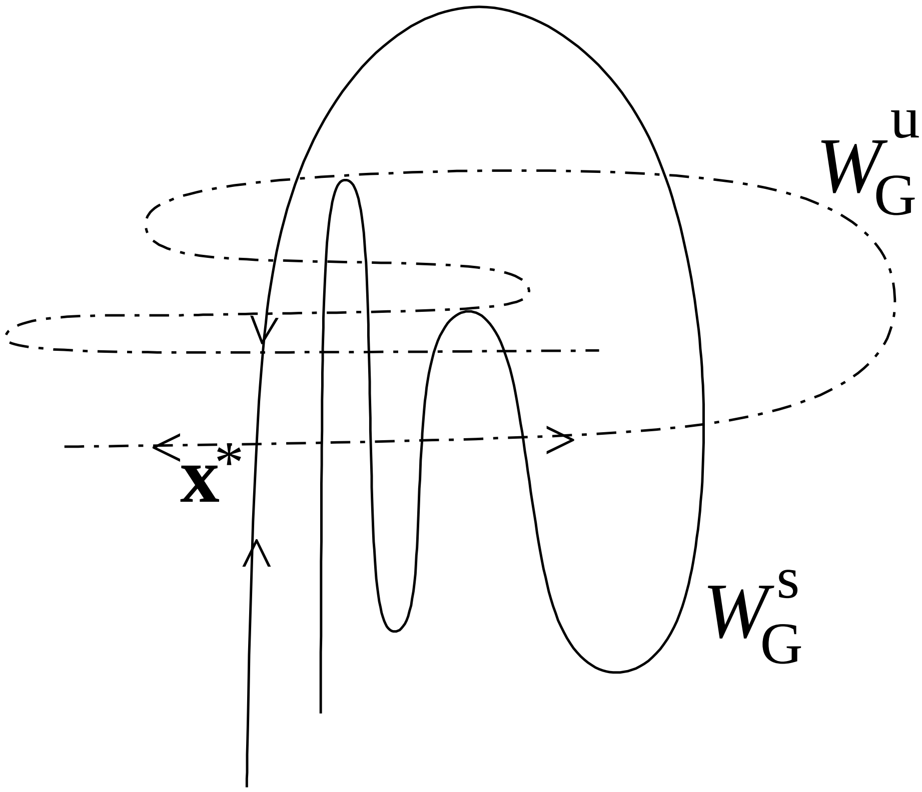


FIGURE 5.2.2. Homoclinic tangle of stable and unstable manifolds.

stretching and folding of the phase space. We illustrate this process by an important example: *Smale's horseshoe map*.

In the 1960's, Smale studied the dynamics of maps with saddles and homoclinic points and showed that these are (locally) equivalent to the dynamics of the horseshoe map whose construction is given schematically in figure 5.2.3. The horseshoe map can be described as a two-dimensional version of the logistic map with $\mu > 4$: like the logistic map, it has a fractal invariant set on which the dynamics is equivalent to the shift and it is chaotic.

5.3. Bifurcation theory

The theory of bifurcations for one-parameter families of two-dimensional maps $\mathbf{F}_\mu$ is similar to that of one-dimensional maps discussed in §4.2. If the map has a fixed point $\mathbf{x}_*$, then the fixed point persists provided that the eigenvalues λ_1 and λ_2 of $D\mathbf{F}|_{\mathbf{x}_*}$ are both different from 1.

Suppose, as we continuously vary μ, we find that an eigenvalue becomes equal to 1 at $\mu = \mu_c$. Then a *saddle–node* bifurcation may occur, with two fixed points on one side of the bifurcation (e.g., for $\mu < \mu_c$) which collide for $\mu = \mu_c$ and no fixed points on the other side of the bifurcation (for $\mu > \mu_c$).

The stability of the fixed point can change (as we vary μ) if one eigenvalue, λ_1 say, becomes such that $|\lambda_1| = 1$ for $\mu = \mu_c$. One possibility is that $\lambda_1 = -1$, in which case the system may undergo a *flip* bifurcation similar to that studied for one-dimensional systems, with the destabilisation of the fixed point associated with the birth of a period-2 orbit.

But $|\lambda_1| = 1$ can also occur for a complex value of λ_1 (with nonzero imaginary part). In this case, $\lambda_1 = \overline{\lambda}_2 = \lambda = \mathrm{e}^{\mathrm{i}\alpha}$ for some $\alpha \neq 0, \pi$, and the bifurcation is the:

Hopf bifurcation (or Andronov–Hopf bifurcation)

To study this bifurcation we use the normal form derived in §3.3. Under the assumption that $\lambda^k \neq 1$ for $k = 1, 2, 3, 4$ at $\mu = \mu_c$, the successive variable transformations used in §3.3 can be carried out to write the two-dimensional system in the form

$$\zeta_{n+1} = \lambda \zeta_n + q\zeta_n^2 \overline{\zeta}_n + \cdots .$$

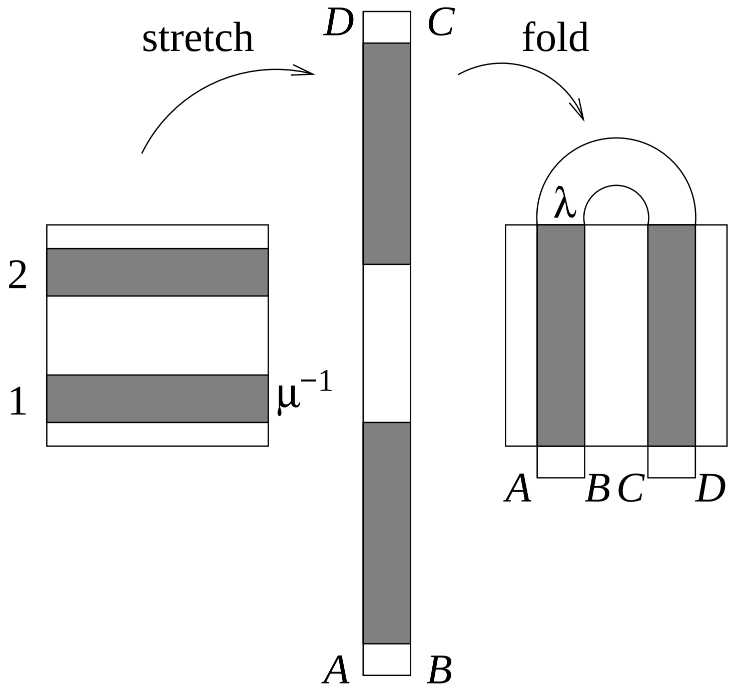


FIGURE 5.2.3. The two transformations defining the Smale horseshoe: the unit square (i.e., the domain of the transformation) is first stretched by a factor μ in the vertical direction and contracted by a factor λ in the horizontal direction, then folded to form a horseshoe. The dynamics of points mapped on the unit square are considered. The two horizontal bands are mapped onto the two vertical bands.

(Note:

(i) the variable transformations depend on μ, so that $q = q(\mu)$;
(ii) in deriving the normal form, we first translated the coordinate system so that the fixed point $\mathbf{x}_* = \mathbf{0}$).

Introducing polar coordinates $\zeta_n = r_n e^{i\theta_n}$ brings the system into the form[1]

$$r_{n+1} = |\lambda|(r_n + h r_n^3) + \cdots, \quad \text{where} \quad h = \text{Re}\left(\frac{q}{\lambda}\right).$$

For $\mu \approx \mu_c$ (i.e., $|\lambda| \approx 1$), terms of order higher than r_n^3 are neglected. The dynamics of r_n are thus seen to be those of a one-dimensional map with the fixed points

$$r^*_{(i)} = 0 \qquad \text{and} \qquad r^*_{(ii)} = \sqrt{\frac{1-|\lambda|}{h}}.$$

[1]Substituting $\zeta_n = r_n e^{i\theta_n}$ into the normal form gives

$$r_{n+1} e^{i\theta_{n+1}} = e^{i\theta_n}(\lambda r_n + q r_n^3 + \cdots),$$

and taking the modulus of both sides gives

$$\begin{aligned}
r_{n+1} &= \sqrt{(\lambda r_n + q r_n^3 + \cdots)(\overline{\lambda} r_n + \overline{q} r_n^3 + \cdots)} \\
&= \sqrt{r_n^2|\lambda|^2 + r_n^4(\lambda\overline{q} + \overline{\lambda} q) + r_n^6|q|^2 + \cdots} \\
&= |\lambda|\, r_n \sqrt{1 + r_n^2\left(\frac{q}{\lambda} + \frac{\overline{q}}{\overline{\lambda}}\right) + r_n^4\frac{|q|^2}{|\lambda|^2} + \cdots} \\
&\approx |\lambda|\, r_n \left[1 + \frac{1}{2}\left(r_n^2\left(\frac{q}{\lambda} + \frac{\overline{q}}{\overline{\lambda}}\right) + r_n^4\frac{|q|^2}{|\lambda|^2}\right)\right] \\
&\qquad \left[\text{Recall Taylor expansion: } \sqrt{1+x} \approx 1 + \frac{1}{2}x \quad \text{for } |x| \ll 1\right] \\
&\approx |\lambda|\, r_n \left[1 + \frac{1}{2}\left(r_n^2\left(\frac{q}{\lambda} + \frac{\overline{q}}{\overline{\lambda}}\right)\right)\right] \quad \text{for small } r_n, |q|.
\end{aligned}$$

The fixed point $r^*_{(ii)}$ only exists for either (a) $|\lambda| \leq 1$ if $h > 0$; or (b) $|\lambda| \geq 1$ if $h < 0$; i.e., it exists only on one side of the bifurcation. In the two-dimensional system, this fixed point $r^*_{(ii)}$ corresponds to orbits that lie on an *invariant curve* (almost a circle in the complex plane ζ). This heuristic argument can be made rigorous, leading to the classification of the Hopf bifurcation in two cases:

- *Supercritical case:* $h(\mu_c) < 0$
  there is an attracting fixed point and no invariant curve for $|\lambda| < 1$; and a repelling fixed point and an (attracting) invariant curve for $|\lambda| > 1$.
- *Subcritical case:* $h(\mu_c) > 0$
  there is an attracting fixed point and a (repelling) invariant curve for $|\lambda| < 1$; and a repelling fixed point and no invariant curve for $|\lambda| > 1$.

5.4. Attractors and basins of attraction: the Hénon map

The large n behaviour of orbits can be characterised by their *ω-limit set*, defined for a point $\mathbf{x}_0$ by

$$\omega(\mathbf{x}_0) = \{\mathbf{x} : \forall N > 0 \text{ and } \forall \epsilon > 0, \quad \exists n > N \text{ such that } |\mathbf{F}^n(\mathbf{x}_0) - \mathbf{x}| < \epsilon\}.$$

Thus, the iterates $\mathbf{F}^n(\mathbf{x}_0)$ of $\mathbf{x}_0$ come arbitrarily close to points of $\omega(\mathbf{x}_0)$ for sufficiently large n.

For fixed points $\mathbf{x}_*$, it is clear that $\omega(\mathbf{x}_*) = \{\mathbf{x}_*\}$. Furthermore, the points of a periodic orbit make up their own ω-limit set. However, in general, points do not belong to their ω-limit set: an attracting fixed point, for example, constitutes the ω-limit set of all points which are attracted by it. ω-limit sets are important when they attract a large number of orbits. This motivates the definition of *attractors* as ω-limit sets attracting a set of initial conditions with non-zero *measure* (i.e., for the two-dimensional case, a set of initial conditions with non-zero *area*). Attracting fixed points and periodic orbits are attractors already encountered in one-dimensional systems. There are also *chaotic attractors*, defined as attractors containing a dense chaotic orbit; for example, the logistic map has a chaotic attractor for most μ with $\mu_\infty < \mu < 4$. Chaotic attractors are also known as *strange* attractors. (NB: The invariant set that exists for the logistic map with $\mu > 4$ is not an attractor. This is because the invariant set has *zero measure* in the sense that it can be covered by intervals whose total length is arbitrarily small).

For two-dimensional systems, the structure of attractors can be richer. For example, the Hénon map

$$x_{n+1} = a - x_n^2 + by_n, \quad y_{n+1} = x_n,$$

with $b = -0.3$ undergoes a period-doubling cascade and transition to chaos as a increases (see `Worksheet 6`). Thereby, it possesses fixed and periodic points as attractors as well as chaotic attractors made up of curves or *fractal* sets. Fractal sets — which will be discussed in §6 — are typically characterised as possessing a degree of *self-similarity*; i.e., if we magnify a tiny part of a fractal, we observe features reminiscent of the whole. Sometimes the similarity is exact (e.g., Cantor sets), but more often it is only approximate or statistical. The fractal attractor (which is a chaotic attractor or strange atractor) of the Hénon map with $a = 1.4$ and $b = 0.3$ is illustrated in figure 5.4.4. The right panel shows a zoom on a part of the attractor and hints at the self-similar structure; further zooms would demonstrate this more convincingly.

The set of initial conditions which is attracted to a given attractor is called its *basin of attraction*. Basins of attractions can have complicated structures even when the attractors themselves are simple. As an example, figure 5.4.5 shows (in black) an approximation of the basin of attraction of infinity (i.e., $\{\mathbf{x} : \mathbf{F}^n(\mathbf{x}) \to \infty \text{ for } n \to \infty\}$) for the Hénon map with $a = 1.39$ and $b = 0.3$. In this case, orbits which do not tend to infinity are attracted to a period-2 orbits. The boundary between the two basins of attractions is, in fact, a fractal. Basins of attraction with extremely complicated structures can be found.

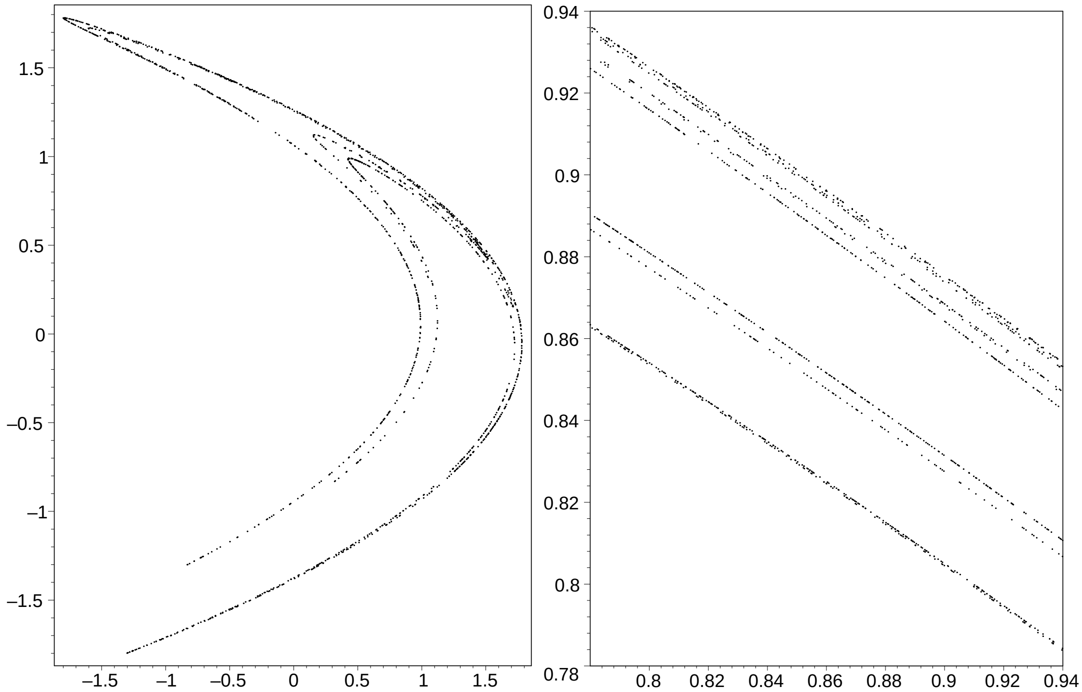


FIGURE 5.4.4. Attractor of the Hénon map with $a = 1.4$ and $b = 0.3$. The right panel shows a zoom which suggests the fractal structure of the attractor.

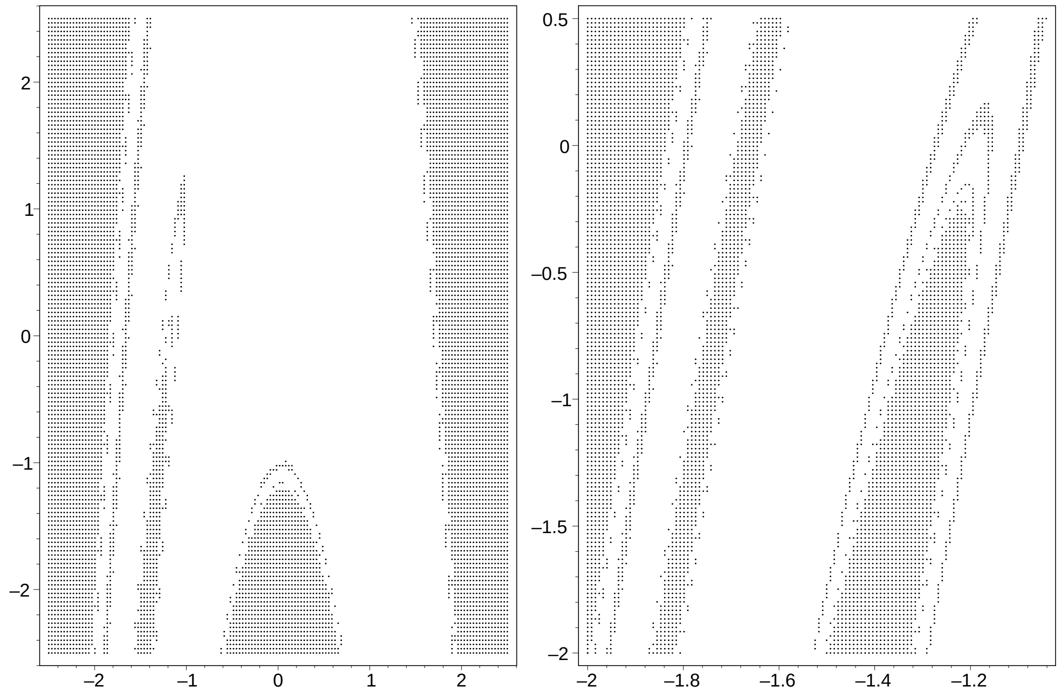


FIGURE 5.4.5. Basin of attraction of infinity for the Hénon map with $a = 1.39$ and $b = 0.3$ (in black).

CHAPTER 6

Fractals and their dimensions

Recall the invariant set for the logistic map with $\mu > 4$ which we encountered in §4.4.3. This Cantor set is an example of a *fractal set*.[1]

Cantor's middle-third set

The middle-third set S is defined recursively as follows (see figure 6.0.1). Starting with the unit interval $S_0 = [0,1]$, we construct a sequence of sets S_n by removing from each segment of S_{n-1} its middle third. Thus,

$$S_0 = [0,1],$$
$$S_1 = [0,1/3] \cup [2/3,1],$$
$$S_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1],$$

etc, and

$$S = \bigcap_{n=0}^{\infty} S_n.$$

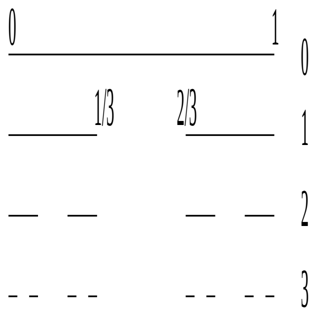


FIGURE 6.0.1. Construction of Cantor's middle-third set: S_0, S_1 S_2 and S_3 are shown.

Note the following:

(i) S_n contains 2^n segments, each of of length 3^{-n}.
(ii) The set S contains an uncountable number of points.
(iii) The set S has zero 'length' in the sense that it can be covered by intervals whose total length is arbitrarily small. (This follows from (i)).
(iv) The set S is nowhere dense.
(v) S can also be defined using the ternary representation:

$$S = \{x \in [0,1] : x = d_1/3 + d_2/3^2 + d_3/3^3 \cdots \quad \text{with} \quad d_i = 0 \text{ or } 2 \ \ \forall i\}.$$

Fractal sets appear often in dynamical systems, particularly as attractors for chaotic systems. For example, the attractor for some Hénon maps and the butterfly–shaped attractor of the Lorenz equations are both fractals. Basins of attractions can also be fractals. In this section we introduce the notion of fractals and their dimensions, informally.

We are familiar with the notion that geometrical objects are attributed a *topological dimension* d, which is an integer for standard objects. For example, for points $d = 0$, for curves $d = 1$, for surfaces

[1]The invariant set for the Smale horsehoe map is also a fractal set.

$d = 2$, for volumes $d = 3$, etc. The topological dimension d specifies the minimum number of coordinates necessary to locate a point on each of these objects. E.g., a smooth curve is one–dimensional because every point on it is determined by one number, namely the arc length from some fixed reference point on the curve.

Fractals are objects for which this correspondence does not hold; a dimension which (roughly) indicates the number of coordinates required can be defined, but it is not an integer. For example, consider *von Koch's snowflake*. This fractal is constructed recursively: starting with an equilateral triangle T_0, the middle-third of each segment is replaced by two segments of equal length forming an equilateral triangle with the replaced middle-third segment. The fractal T is obtained by repeating this operation an infinite number of times; i.e.,

$$T = \lim_{i \to \infty} T_i \,,$$

as indicated in figure 6.0.2.

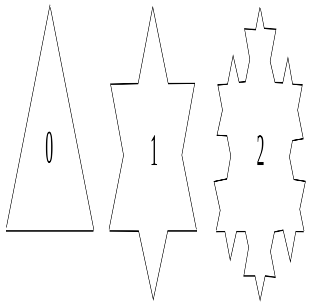


FIGURE 6.0.2. von Koch's snowflake: T_0, T_1 and T_2 .

Since the von Koch snowflake is a (closed) curve, one might be tempted to think that it is one–dimensional. However, the von Koch snowflake has an infinite arc length. Furthermore, the arc length between any two points on the von Koch snowflake is infinite. (Exercise: convince yourself that this is true!). Hence, points on the snowflake cannot be determined by their arc length from a particular reference point, because every point is infinitely far from every other. Intuitively, we might guess that the von Koch snowflake is more than one–dimensional but less than two–dimensional.

Box dimension

The topological dimension d of Cantor's middle-third set is 0, since it is a collection of points. However, more useful concepts of dimensions can be introduced for fractals. Here we consider the *box dimension*.

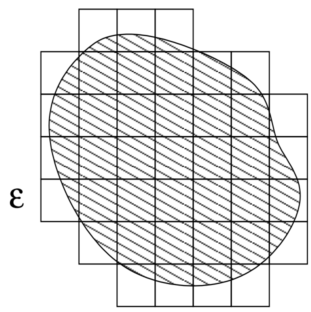


FIGURE 6.0.3. Box dimension: $N(\epsilon)$ is the number of boxes of size ϵ necessary to cover an object.

Consider a curve of finite length l and cover it by boxes (or hypercubes) of size ϵ. For ϵ small, the minimum number of boxes required for such a covering is $N(\epsilon) \sim l\epsilon^{-1}$. Simarly, the minimum number of boxes required to cover a surface of area A is $N(\epsilon) \sim A\epsilon^{-2}$. In general, for simple (i.e., non-fractal) objects of topological dimension d, we have $N(\epsilon) \sim \alpha\, \epsilon^{-d}$, where α is a constant. See figure 6.0.3. For fractal objects, the same form of relation holds, but with an exponent different from d; i.e.,

$$N(\epsilon) \sim \alpha\, \epsilon^{-D}$$

for some D. This leads to the definition of the box dimension:

$$D = -\lim_{\epsilon \to 0^+} \frac{\ln N(\epsilon)}{\ln \epsilon}.$$

Examples

Cantor's middle-third set. S is covered entirely by the set S_n obtained after n iterations. This set consists of 2^n segments, each of length 3^{-n}. Thus, taking $\epsilon = 3^{-n}$ and $N(\epsilon) = 2^n$, we find that $D = \ln 2/\ln 3 = 0.6309$.

von Koch's snowflake. The box dimension is derived by noting that at stage n, there are 3×4^n segments, each of length 3^{-n}. Taking $\epsilon = 3^{-n}$ and $N(\epsilon) = 3 \times 4^n$ gives a box dimension $D = \ln 4/\ln 3 = 1.2619$ (which is consistent with our earlier intuitive guess).

Hénon map. The attractor of the Hénon map for certain a and b is a fractal whose box dimension can be computed numerically. A dimension $1 < D < 2$ is found, with the precise value depending on the parameters.

Remarks:

- The same box dimension D is obtained if the boxes of size ϵ are taken on a fixed grid, as in figure 6.0.3, or if they are placed freely to minimize their number. The same dimension is also obtained if (hyper)spheres of diameter ϵ are employed.
- An alternative definition of D considers the limit

$$\lim_{\epsilon \to 0^+} N(\epsilon)\epsilon^p$$

for different values of p. D is then characterized by the fact that this limit is ∞ for $0 < p < D$ and 0 for $p > D$.

- A similar characterisation is used to define the Hausdorff dimension D_H. For this dimension, one covers the object with sets which are not necessarily cubes but can take any shape. Starting with

$$H(\epsilon, p) = \inf \sum_i |U_i|^p,$$

where the sets U_i, with diameters $|U_i| < \epsilon$, cover the object and the infimum is taken over all possible covers U_i, the Hausdorff dimension is D_H if

$$\lim_{\epsilon \to 0^+} H(\epsilon, p) = \begin{cases} \infty & \text{for } 0 \le p < D_H \\ 0 & \text{for } p > D_H \end{cases}.$$

In general $D_H \le D$, but frequently $D_H = D$.