$f(x) := \sqrt{x}$ near $x = 16$ by $L(x) = f(16) + f'(16)(x - 16) = 4 + \frac{1}{8}(x - 16)$, in particular, $\sqrt{17} \approx 4\frac{1}{8}$. This corresponds to the first step in the Newton method for approximating $\sqrt{17}$, as explained by Exercise 1.3.6 (p. 22).

Such linear approximation can sometimes be useful for dynamics when the orbits of a nonlinear map stay near enough to the reference point for the linear approximation to be relevant. There are examples of this throughout the book. For now we give a special case of Proposition 2.2.17:

PROPOSITION 2.1.1. *Suppose F is a differentiable map of the line and $F(b) = b$. If all orbits of the linearization of F at b are asymptotic to b then all orbits of F that start near enough to b are asymptotic to b as well.*

The quintessential linearization result in analysis is the Mean Value Theorem A.2.2 (see also Lemma 2.2.12), and it is used numerous times already in the next section (for example, Proposition 2.2.3 and Section 2.2d4). Linearization also plays an important role in highly complicated dynamical systems (see for example, Chapter 7 and Chapter 10).

Exercises

EXERCISE 2.1.1. Show that the change of variable $y = x - \dfrac{b}{1-k}$ transforms the recursion $x_{i+1} = f(x_i) = kx_i + b$ to the recursion $y_{i+1} = ky_i$.

EXERCISE 2.1.2. Describe what asymptotic behaviors appear for the maps $f(x) = kx$ when $k \pm 1$.

EXERCISE 2.1.3. Describe what asymptotic behaviors appear for the maps $f(x) = kx + b$ when $k = \pm 1$ and $b \neq 0$.

2. Contractions in euclidean space

Traditionally, scientists and engineers have had a preference for dynamical systems that have stable asymptotic behavior, ideally settling into a steady state, maybe after a short period of "transient behavior". Simple real-life examples abound. A desk lamp, when turned on, settles into the "on" state of constant light intensity instantly (after a very short heating period for the filament). Unless it is broken it does not blink or flicker erratically. Likewise escalators are preferred in a steady state of constant speed. Radios, when first turned on, have complicated transient behavior for a remote fraction of a second, but then settle into a steady state of reception. Our tour of dynamical systems begins with those that display such simple behavior.

Corresponding to the above continuous time real-world examples the simplest imaginable kind of asymptotic behavior of a discrete time dynamical system is represented by the convergence of iterates of any given point to a particular point, a steady state. There is an important general class of situations where this kind of behavior can be established, namely for contracting maps. These are presented here not only because their simple dynamics provides an ideal starting point, but also because we will use contractions as a tool in numerous problems in analysis, differential equations, as well as for study of dynamical systems with more complicated behavior. These applications appear throughout this book, and Chapter 9 concentrates on such applications.

We now define contractions and clarify the usage of the words "map" and "function".

a. Definitions. When we use the word "map" we usually mean that the domain and range lie in the same space, and even more often that the range is in the domain—we iterate maps and in this way they generate a dynamical system. Fibonacci's rabbits, Maine lobsters, phyllotaxis, butterflies, and methods for finding roots all provided examples of such dynamical systems. Time is discrete, and the laws of nature (or of an algorithm) have been distilled into a rule that determines present data from prior data, and the next "state" of the system from the present state. All this is achieved by applying one map that encodes these laws. So, discrete-time dynamical systems are maps of a space to itself. Maps are almost always continuous.

"Functions", on the other hand, have numerical values even if they are defined on a rather different space, and they are not iterated. still we will sometimes use the conventional word "function" to denote a map of the ral line or its part into itself. There is a third possibility of transformations used for a change of variables. These are called *coordinate changes* or *conjugacies* (and maybe sometimes also maps). One map that we always have at our disposal is the identity, which we denote by Id. It is defined by $\mathrm{Id}(x) = x$.

Now we define contracting maps:

DEFINITION 2.2.1. A map f of a subset X of euclidean space is said to be *Lipschitz continuous* with Lipschitz constant λ, or λ-Lipschitz if

$$d(f(x), f(y)) \leq \lambda d(x, y) \tag{2.2.1}$$

for any $x, y \in X$. f is said to be a *contraction* or a *λ-contraction* if $\lambda < 1$. If a map f is Lipschitz-continuous then we define $\mathrm{Lip}(f) := \sup_{x \neq y} d(f(x), f(y))/d(x, y)$.

EXAMPLE 2.2.2. The function $f(x) = \sqrt{x}$ defines a contraction on $[1, \infty)$. To prove this, we show that for $x \geq 1$ and $t \geq 0$ we have $\sqrt{x+t} \leq \sqrt{x} + (1/2)t$ (why is this enough?). This is most easily seen by squaring:

$$(\sqrt{x} + \frac{t}{2})^2 = x + xt + \frac{t^2}{4} \geq x + xt \geq x + t.$$

b. The case of one variable. We now give an easy way of checking the contraction condition that uses the derivative.

PROPOSITION 2.2.3. *Let I be an interval and $f\colon I \to \mathbb{R}$ a differentiable function with $|f'(x)| \leq \lambda$ for all $x \in I$. Then f is λ-Lipschitz.*

PROOF. By the Mean Value Theorem A.2.2, for any two points $x, y \in I$ there exists a point c between x and y such that

$$d(f(x), f(y)) = |f(x) - f(y)| = |f'(c)(x - y)| = |f'(c)|d(x, y) \leq \lambda d(x, y).$$

□

Note that we need no information about f' at the endpoints of I.

EXAMPLE 2.2.4. This criterion makes it easier to check that $f(x) = \sqrt{x}$ defines a contraction on $I = [1, \infty)$ because $f'(x) = 1/2\sqrt{x} \leq 1/2$ for $x \geq 1$.

Let us point out that by Exercise 2.2.14 the weaker condition $|f'(x)| < 1$ does not suffice to obtain (2.2.1). However, sometimes it does:

PROPOSITION 2.2.5. *Let I be a closed bounded interval and $f\colon I \to I$ a continuously differentiable function with $|f'(x)| < 1$ for all $x \in I$. Then f is a contraction.*

PROOF. The maximum λ of $|f'(x)|$ is attained at some point x_0 because f' is continuous. It is less than 1 because $|f'(x_0)| < 1$. □

The difference is that the real line is not closed and bounded (see also Exercise 2.2.13 for a related fact).

In calculus a favorite example of a recursively defined sequence is of the form $a_{n+1} = f(a_n)$, with a_0 given and f a function with $|f'| \leq \lambda < 1$. This is a simple dynamical system given by the map f. For each initial value a_0 a sequence is uniquely defined by $a_{n+1} = f(a_n)$. If f is invertible then this sequence is defined for all $n \in \mathbb{Z}$.

DEFINITION 2.2.6. For a map f and a point x the sequence $x, f(x), f(f(x)), \dots, f^n(x), \dots$ (if f is not invertible) or the sequence $\dots f^{-1}(x), x, f(x), \dots$ is called the *orbit* of x under f. A *fixed point* is a point such that $f(x) = x$. The set of fixed points is denoted by $\mathrm{Fix}(f)$. A *periodic point* is a point x such that $f^n(x) = x$ for some $n \in \mathbb{N}$, that is, a point in $\mathrm{Fix}(f^n)$. Such an n is said to be *a* period of x. The smallest such n is called the *prime period* of x.

EXAMPLE 2.2.7. If $f(x) = -x^3$ on $\mathbb{R}$ then 0 is the only fixed point and ± 1 is a periodic orbit, that is, 1 and -1 are periodic points with prime period 2.

The reason the calculus examples of such sequences always converge is the following important fact:

PROPOSITION 2.2.8 (Contraction Principle). *Let $I \subset \mathbb{R}$ be a closed interval, possibly infinite on one or both sides and $f\colon I \to I$ a λ-contraction. Then f has a unique fixed point x_0 and $|f^n(x) - x_0| \leq \lambda^n |x - x_0|$ for every $x \in \mathbb{R}$, that is, every orbit of f converges to x_0 exponentially.*

PROOF. By iterating $|f(x) - f(y)| \leq \lambda |x - y|$, one sees that

$$|f^n(x) - f^n(y)| \leq \lambda^n |x - y| \tag{2.2.2}$$

for $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$, so for $x \in I$ and $m \geq n$ we can use the triangle inequality to show

$$\begin{aligned} |f^m(x) - f^n(x)| &\leq \sum_{k=0}^{m-n-1} |f^{n+k+1}(x) - f^{n+k}(x)| \\ &\leq \sum_{k=0}^{m-n-1} \lambda^{n+k} |f(x) - x| \leq \frac{\lambda^n}{1-\lambda} |f(x) - x|. \end{aligned} \tag{2.2.3}$$

Here we used the familiar fact

$$(1-\lambda) \sum_{k=l}^{n-1} \lambda^k = \lambda^l + \lambda^{l+1} + \dots + \lambda^{n-1} - \lambda^{l+1} + \lambda^{l+2} + \dots + \lambda^n = (\lambda^l - \lambda^n)$$

about partial sums of geometric series. Since the right hand side of (2.2.3) becomes arbitrarily small as n gets large, (2.2.3) shows that $(f^n(x))_{n \in \mathbb{N}}$ is a Cauchy sequence. Thus for any $x \in I$ the limit of $f^n(x)$ as $n \to \infty$ exists because Cauchy sequences converge. The limit is in I because I is closed. By (2.2.2) this limit is the same for all x. We denote this limit by x_0 and show that x_0 is a fixed point for f. If $x \in I$ and $n \in \mathbb{N}$ then

$$\begin{aligned} |x_0 - f(x_0)| &\leq |x_0 - f^n(x)| + |f^n(x) - f^{n+1}(x)| + |f^{n+1}(x) - f(x_0)| \\ &\leq (1+\lambda)|x_0 - f^n(x)| + \lambda^n |x - f(x)|. \end{aligned} \tag{2.2.4}$$

Since $|x_0 - f^n(x)| \to 0$ and $\lambda^n \to 0$ as $n \to \infty$, we have $f(x_0) = x_0$.

That $|f^n(x) - x_0| \le \lambda^n |x - x_0|$ for every $x \in \mathbb{R}$ follows from (2.2.2) with $y = x_0$. □

EXAMPLE 2.2.9. In contemplating his rabbits, Leonardo of Pisa, also known as Fibonacci, came up with a model according to which the number of rabbit pairs in the nth month is given by the number b_n defined by the recursive relation $b_0 = 1$, $b_1 = 2$, $b_n = b_{n-1} + b_{n-2}$ for $n \ge 2$ (Section 1.2b). Expecting that the growth of these numbers should be exponential we would like to see how fast these numbers grow by finding the limit of $a_n := b_{n+1}/b_n$ as $n \to \infty$. To that end we use the Contraction Principle. Since

$$a_{n+1} = \frac{b_{n+2}}{b_{n+1}} = \frac{b_{n+1} + b_n}{b_{n+1}} = \frac{1}{b_{n+1}/b_n} + 1 = \frac{1}{a_n} + 1,$$

$(a_n)_{n=1}^{\infty}$ is the orbit of 1 under iteration of the map $g(x) := (1/x) + 1$. Since $g(1) = 2$ we are in fact considering the orbit of 2 under g. Now $g'(x) = -x^{-2}$. This tells us that g is not a contraction on $(0, \infty)$. Therefore we need to find a suitable (closed) interval where this is the case and that is mapped inside itself.

Since $g' < 0$, g is decreasing on $(0, \infty)$. This implies that $g([3/2, 2]) \subset [3/2, 2]$ because $3/2 < g(3/2) = 5/3 < 2$ and $g(2) = 3/2$. Furthermore $|g'(x)| = 1/x^2 \le 4/9 < 1$ on $[3/2, 2]$, so g is a contraction on $[3/2, 2]$. By the Contraction Principle the orbit of 2 and hence that of 1 is asymptotic to the unique fixed point x of g in $[3/2, 2]$. Thus $\lim_{n\to\infty} b_{n+1}/b_n = \lim_{n\to\infty} a_n$ exists. To find the limit we solve the equation $x = g(x) = 1 + 1/x = (x+1)/x$, which is equivalent to $x^2 - x - 1 = 0$. There is only one positive solution: $x = (1 + \sqrt{5})/2$. (This solves Exercise 1.2.3.) Another way of obtaining this ratio and an explicit formula for the Fibonacci numbers is given in Section 3.1i.

c. The case of several variables. We now show that the Contraction Principle holds in higher dimension as well, and we use the same proof, replacing absolute values by the Euclidean distance $d(x, y) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.

PROPOSITION 2.2.10 (Contraction Principle). *Let $X \subset \mathbb{R}^n$ be closed, $f\colon X \to X$ a λ-contraction. Then f has a unique fixed point x_0 and $d(f^n(x), x_0) = \lambda^n d(x, x_0)$ for every $x \in X$.*

PROOF. Iterating $d(f(x), f(y)) \le \lambda d(x, y)$ shows

$$d(f^n(x), f^n(y)) \le \lambda^n d(x, y) \tag{2.2.5}$$

for $x, y \in X$ and $n \in \mathbb{N}$. Thus $(f^n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence because

$$\begin{aligned} d(f^m(x), f^n(x)) &\le \sum_{k=0}^{m-n-1} d(f^{n+k+1}(x), f^{n+k}(x)) \\ &\le \sum_{k=0}^{m-n-1} \lambda^{n+k} d(f(x), x) \le \frac{\lambda^n}{1-\lambda} d(f(x), x) \end{aligned} \tag{2.2.6}$$

for $m \ge n$, and $\lambda^n \to 0$ as $n \to \infty$. Thus $\lim_{n\to\infty} f^n(x)$ exists (because Cauchy sequences in $\mathbb{R}^n$ converge) and is in X because X is closed. By (2.2.5) it is the same for all x. Denote this limit by x_0. Then

$$d(x_0, f(x_0)) \le d(x_0, f^n(x)) + d(f^n(x), f^{n+1}(x)) + d(f^{n+1}(x), f(x_0)) \\ \le (1+\lambda)d(x_0, f^n(x)) + \lambda^n d(x, f(x)) \tag{2.2.7}$$

for $x \in X$ and $n \in \mathbb{N}$. Now $f(x_0) = x_0$ because $d(x_0, f^n(x)) \xrightarrow[n\to\infty]{} 0$. □

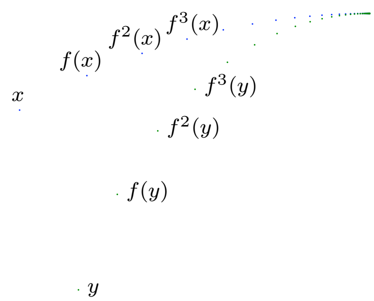

FIGURE 2.2.1. Convergence of iterates

Taking the limit in (2.2.6) as $m \to \infty$ we obtain $d(f^n(x), x_0) \le \dfrac{\lambda^n}{1-\lambda} d(f(x), x)$. This means that after n iterations we can say with certainty that the fixed point is in the $\dfrac{\lambda^n}{1-\lambda} d(f(x), x)$-ball around $f^n(x)$. In other words, if we make numerical computations then we can make a rigorous conclusion about where the fixed point must be (after accounting for roundoff errors).

DEFINITION 2.2.11. We say that two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ of points in $\mathbb{R}^n$ *converge exponentially* (or *with exponential speed*) *to each other* if $d(x_n, y_n) < cd^n$ for some $c > 0$, $d < 1$. In particular, if one of the sequences is constant, that is, $y_n = y$, we say that x_n *converges exponentially* to y.

d. The derivative test. We now show, similarly to the case of one variable, that the contraction property can be verified using the derivative.

To that end we recall some pertinent tools from the calculus of several variables, namely the differential and the Mean Value Theorem.

1. *The differential*. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a map with continuous partial derivatives. Then at each point one can define the derivative or differential of $f = (f_1, \dots, f_m)$ as the linear map defined by the matrix of partial derivatives

$$Df := \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}$$

We say that the map is regular at x_0 if this map is invertible. We define the norm (see Definition A.1.26) of the differential by the norm of the matrix Df. In linear algebra the norm of a matrix A is defined by looking at its action as a linear map:

$$\|A\| := \max_{v \neq 0} \frac{\|A(v)\|}{\|v\|} = \max_{\|v\|=1} \|A(v)\|. \tag{2.2.8}$$

Geometrically this is easy to visualize by considering the second of these expressions: Consider the unit sphere $\{v \in \mathbb{R}^n \mid \|v\| = 1\}$ and notice that the second maximum is just the size of the largest vectors in the image of this unit sphere. The image of the unit sphere under a linear map is an ellipsoidal figure and in a picture the largest vector is easy to find. Calculating this norm in particular cases may not always be easy, but there are easy ways of finding upper bounds (see Exercise 2.2.9 and Lemma 3.3.2).

2. *The Mean Value Theorem.*

LEMMA 2.2.12. *If $g\colon [a,b] \to \mathbb{R}^m$ is continuous and differentiable on (a,b) then there exists $t \in [a,b]$ such that $\|g(b) - g(a)\| \leq \|\frac{d}{dt}g(t)\|(b-a)$.*

PROOF. Let $v = g(b) - g(a)$, $\varphi(t) = \langle v, g(t) \rangle$. By the Mean Value Theorem A.2.2 for one variable there exists a $t \in (a,b)$ such that $\varphi(b) - \varphi(a) = \varphi'(t)(b-a)$ and so

$$\begin{aligned}(b-a)\|v\|\|\frac{d}{dt}g(t)\| \geq (b-a)\langle v, \frac{d}{dt}g(t)\rangle = \frac{d}{dt}\varphi(t)(b-a) = \varphi(b) - \varphi(a)\\ = \langle v, g(b)\rangle - \langle v, g(a)\rangle = \langle v, v\rangle = \|v\|^2.\end{aligned}$$

Divide by $\|v\|$ to finish the proof. □

3. *Convexity*. A further notion we need is that of a convex set.

DEFINITION 2.2.13. A *convex set* in $\mathbb{R}^n$ is set C such that for all $a, b \in C$ the line segment with endpoints a, b is entirely contained in C. It is said to be *strictly convex* if for any points a, b in the closure of C the segment from a to b is contained in C, except possibly for one or both endpoints.

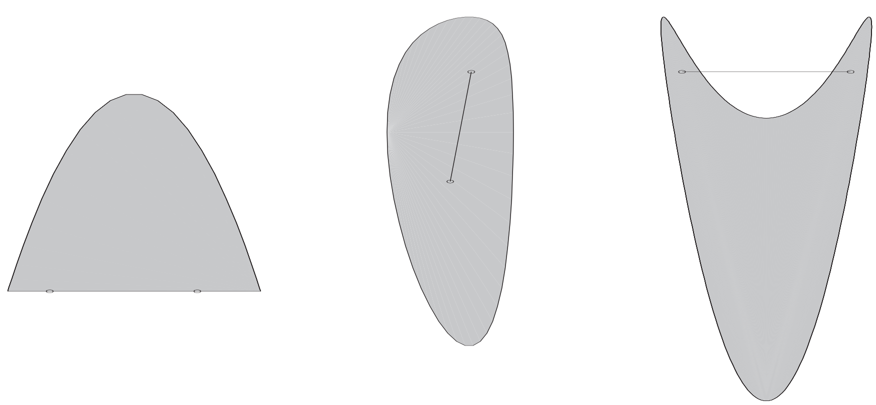

FIGURE 2.2.2. A convex, strictly convex, and nonconvex set

For example, the disk $\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2<1\}$ is strictly convex. The open upper half plane $\{(x,y) \in \mathbb{R}^2 \mid y>0\}$ is convex but not strictly convex. A kidney shape $\{(r,\theta) \mid 0 \le r \le 1+(1/2)\sin\theta\}$ (in polar coordinates) is not convex. Neither is the *annulus* $\{(x,y) \in \mathbb{R}^2 \mid 1<x^2+y^2<2\}$.

4. *The derivative test*. We can now give two versions of a derivative test for contractions in several variables.

THEOREM 2.2.14. *If $C \subset \mathbb{R}^n$ is convex and open and $f\colon C \to \mathbb{R}^m$ is differentiable with $\|Df(x)\| \le M$ for all $x \in C$ then $\|f(x)-f(y)\| \le M\|x-y\|$ for $x,y \in C$.*

PROOF. The line segment connecting x and y is given by $c(t)=x+t(y-x)$ for $t \in [0,1]$ and it is contained in C by convexity. Let $g(t) := f(c(t))$. Then by the chain rule

$$\|\frac{d}{dt}g(t)\| = \|Df(c(t))\frac{d}{dt}c(t)\| = \|Df(c(t))(y-x)\| \le M\|y-x\|.$$

By Lemma 2.2.12 this implies $\|f(y)-f(x)\| = \|g(1)-g(0)\| \le M\|y-x\|$. □

COROLLARY 2.2.15. *If $C \subset \mathbb{R}^n$ is a convex open set, $f\colon C \to C$ a map with continuous partial derivatives and $\|Df\| \le \lambda < 1$ at every point $x \in \mathbb{R}^n$ then f is a λ-contraction.*

The role of convexity in Theorem 2.2.14, and hence in Corollary 2.2.15, is elucidated in Exercise 2.2.12. In particular, it is not sufficient to assume that any two points of C can be connected by a curve, however nice. It is really necessary to use a single line segment.

The preceding corollary does not quite seem to be geared towards applying the Contraction Principle because an open set may not contain the limits of Cauchy sequences in it. Therefore we give a result that holds for the closure of such a set. It is proved exactly like Theorem 2.2.14.

THEOREM 2.2.16. *If $C \subset \mathbb{R}^n$ is an open strictly convex set, $\bar{C}$ its closure, $f\colon \bar{C} \to \mathbb{R}^n$ differentiable on C and continuous on $\bar{C}$ with $\|Df\| \le \lambda < 1$ on C then f has a unique fixed point $x_0 \in \bar{C}$ and $d(f^n(x),x_0) \le \lambda^n d(x,x_0)$ for every $x \in \bar{C}$.*

PROOF. For $x,y \in \bar{C}$ we parameterize the line segment connecting x and y by $c(t) = x+t(y-x)$ for $t \in [0,1]$ and let $g(t) := f(c(t))$. Then $c((0,1))$ is contained in C by strict convexity and

$$\|\frac{d}{dt}g(t)\| = \|Df(c(t))\frac{d}{dt}c(t)\| = \|Df(c(t))(y-x)\| \le \lambda\|y-x\|.$$

By Lemma 2.2.12 this implies $\|f(y)-f(x)\| \le \lambda\|y-x\|$. Thus f is a λ-contraction and has a unique fixed point x_0. Furthermore, $d(f^n(x),x_0) = \lambda^n d(x,x_0)$ for every $x \in \bar{C}$. □

e. Local contractions. Now we discuss maps that are not contracting on their entire domain but on a part of it. A prime example of a map that contracts only locally is given by the following:

PROPOSITION 2.2.17. *Let f be a continuously differentiable map with a fixed point x_0 where $\|Df_{x_0}\| < 1$. Then there is a closed neighborhood U of x_0 such that $f(U) \subset U$ and f is a contraction on U.*

DEFINITION 2.2.18. By a closed neighborhood of x we mean the closure of an open set containing x.

PROOF. Since Df is continuous there is a small closed ball $U = \overline{B(x_0, \eta)}$ around x_0 on which $\|Df_x\| \le \lambda < 1$ (Exercise 2.2.11). If $x, y \in U$ then $d(f(x), f(y)) \le \lambda d(x, y)$ by Corollary 2.2.15, so f is contraction on U. Furthermore, taking $y = x_0$ shows that if $x \in U$ then $d(f(x), x_0) = d(f(x), f(x_0)) \le \lambda d(x, x_0) \le \lambda\eta < \eta$ and hence $f(x) \in U$. □

Unfortunately the definition of $\|Df\|$ is inconvenient for calculations. However, it is easy to avoid having to use it:

PROPOSITION 2.2.19. *Let f be a continuously differentiable map with a fixed point x_0 such that all eigenvalues of Df_{x_0} have absolute value less than 1. Then there is a closed neighborhood U of x_0 such that $f(U) \subset U$ and f is a contraction on U.*

PROOF. In the next chapter (Proposition 3.3.3) we will show that the assumption on the eigenvalues implies that one can choose a norm which we denote by $\|\cdot\|'$ for which $\|Df\|' < 1$. Now Proposition 2.2.17 applies. In other words a sufficiently small closed "ball" around x_0 with respect to the norm $\|\cdot\|'$ can be chosen as the set U. This ball is in fact an ellipsoid in $\mathbb{R}^n$. □

This particular situation is interesting because of some robustness under perturbation.

f. Perturbations. We now study what happens to the fixed point when one perturbs a contraction.

PROPOSITION 2.2.20. *Let f be a continuously differentiable map with a fixed point x_0 where $\|Df_{x_0}\| < 1$ and let U be a closed neighborhood of x_0 such that $f(U) \subset U$. Then any map g sufficiently close to f is a contraction on U.*

Specifically, if $\epsilon > 0$ then there is a $\delta > 0$ such that any map g with $\|g(x)-f(x)\| \le \delta$ and $\|Dg(x) - Df(x)\| \le \delta$ on U maps U into U and is a contraction on U with its unique fixed point y_0 in $B(x_0, \epsilon)$.

PROOF. Since the linear map Df_x depends continuously on the point x there is a small closed ball $U = \overline{B(x_0, \eta)}$ around x_0 on which $\|Df_x\| \le \lambda < 1$ (Exercise 2.2.11). Assume $\epsilon < 1$ and take $\delta = \epsilon\eta(1-\lambda)/2$. Then

$$\|Dg\| \le \|Dg - Df\| + \|Df\| \le \delta + \lambda \le \lambda + (1-\lambda)/2 = (1+\lambda)/2 =: \mu < 1$$

on U, so g is a contraction on U by Corollary 2.2.15. If $x \in U$ then $d(x, x_0) \le \eta$ and

$$\begin{aligned} d(g(x), x_0) &\le d(g(x), g(x_0)) + d(g(x_0), f(x_0)) + d(f(x_0), x_0) \\ &\le \mu d(x, x_0) + \delta + 0 \le \mu\eta + \delta \le \eta(1+\lambda)/2 + \eta(1-\lambda)/2 = \eta, \end{aligned} \tag{2.2.9}$$

so $g(x) \in U$ also, that is, $g(U) \subset U$. Finally, since $g^n(x_0) \to y_0$ we have

$$d(x_0, y_0) \le \sum_{n=0}^{\infty} d(g^n(x_0), g^{n+1}(x_0)) \le d(g(x_0), x_0) \sum_{n=0}^{\infty} \mu^n \le \frac{\delta}{1-\mu} = \frac{\epsilon\eta(1-\lambda)}{1-\lambda},$$

which is less than ϵ. □

The previous result in particular tells us that the fixed point of a contraction depends continuously on the contraction. This part can be proved without differentiability:

PROPOSITION 2.2.21. *If $f\colon \mathbb{R}\times(a,b)\to\mathbb{R}$ is continuous and $f_y := f(\cdot,y)$ satisfies $|f_y(x_1)-f_y(x_2)|\le\lambda|x_1-x_2|$ for all $x_1,x_2\in\mathbb{R}$ and all $y\in(a,b)$, then the fixed point $g(y)$ of f_y depends continuously on y.*

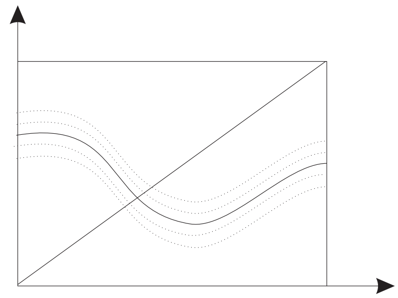

FIGURE 2.2.3. Continuous dependence of the fixed point

PROOF. Since

$$|x-g(y)|\le\sum_{i=0}^{\infty}|f_y^i(x),f_y^{i+1}(x)|\le\frac{1}{1-\lambda}|x-f_y(x)|$$

we take $x=g(y')=f_{y'}(g(y'))$ to get

$$|g(y')-g(y)|\le\frac{1}{1-\lambda}|f_{y'}(g(y'))-f_y(g(y'))|.$$

□

This also works in greater generality (Proposition 2.6.14), and an even stronger result in this direction is given by Theorem 9.2.4.

g. Attracting fixed points. At this point we have encountered two kinds of *stability*: Given a contraction, each individual orbit exhibits stable behavior in that every nearby orbit (actually, every orbit) has precisely the same asymptotics. Put differently, a little perturbation of the initial point has no effect on the asymptotic behavior. This constitutes stability of orbits. On the other hand, Proposition 2.2.20 and Proposition 2.2.21 tell us that contractions are stable as a system, that is, when we perturb the contracting map itself, then the qualitative behavior of all orbits remains the same, and the fixed point changes only slightly.

This is a good time to make precise what we mean by a stable fixed point. As we said, we want every nearby orbit to be asymptotic to it. However, this is not all we want, as Figure 2.2.4 shows, where we have a semistable fixed point. Such a map can be given for example as $f(x)=x+(1/4)\sin^2 x$ if the circle is represented as $\mathbb{R}/\mathbb{Z}$ (see Section 2.6b). We need to make sure that no nearby points ever stray far. But, as the example

$$f(x)=\begin{cases}-2x & x\le 0\\ -x/4 & x>0\end{cases}$$

(or Figure 3.1.3) shows, we must allow points to go a little further for a while.

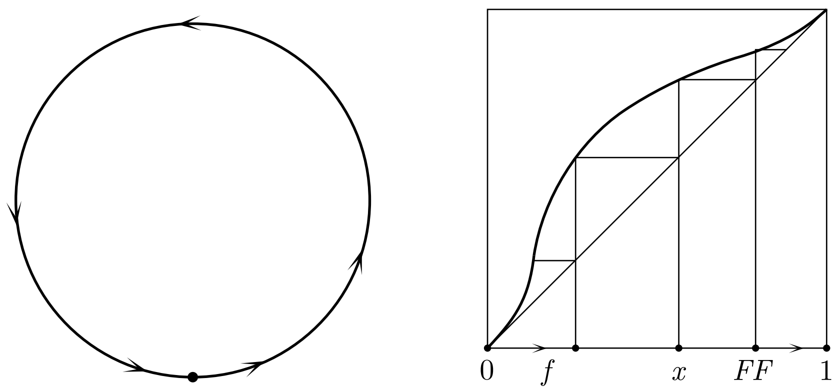

FIGURE 2.2.4. Not an attracting fixed point

DEFINITION 2.2.22. A fixed point p is said to be *Poisson stable* if for every $\epsilon > 0$ there is a $\delta > 0$ such that if a point is within δ of p then its positive semiorbit is within ϵ of p. The point p is said to be *asymptotically stable* or an *attracting fixed point* if it is Poisson stable and there is an $a > 0$ such that every point within a of p is asymptotic to p.

h. The Newton method. A refined application of linear approximation to an otherwise difficult problem is the Newton method for finding roots of equations, which we saw in Section 1.3b2. Doing this exactly is often difficult or impossible and the roots are rarely expressible in closed form. The Newton method can work well to find a root in little computational time given a reasonable initial guess. To see how, consider a function f on the real line and suppose we have a reasonable guess x_0 for a root. Unless the graph intersects the x-axis at x_0, that is, $f(x_0) = 0$, we need to improve our guess. To that end we take the tangent line and see at which point x_1 it intersects the x-axis by setting $f(x_0) + f'(x_0)(x_1 - x_0) = 0$. Thus the improved guess is

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

EXAMPLE 2.2.23. When we start from $x_0 = 4$ for the function $x^2 - 17$ this improved guess is

$$x_1 = x_0 - \frac{x_0^2 - 17}{2x_0} = \frac{x_0}{2} + \frac{17}{2x_0} = \frac{33}{8}.$$

One further step gives

$$x_2 = \frac{33}{16} + \frac{17 \cdot 8}{2 \cdot 33} = \frac{33^2 + 17 \cdot 64}{16 \cdot 33}.$$

Iteratively one can improve the guess to x_3,... using the same formula. With a good initial guess few steps usually give a rather accurate solution. (Indeed, x_2 is already off by less than 10^{-6}.) It is easy to see why: We are applying the map $F(x) := x - \dfrac{f(x)}{f'(x)}$ repeatedly, and the desired point has the following property:

DEFINITION 2.2.24. A fixed point x of a differentiable map F is said to be *superattracting* if $F'(x) = 0$.

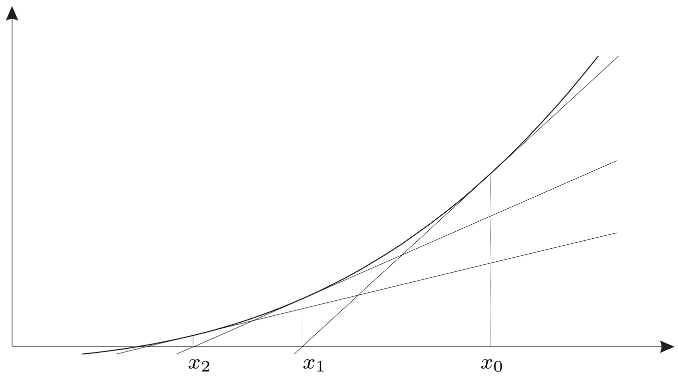

FIGURE 2.2.5

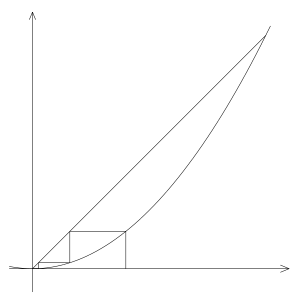

FIGURE 2.2.6. A superattracting fixed point

PROPOSITION 2.2.25. *If* $|f'(x)| > \delta$ *and* $|f''(x)| < M$ *on a neighborhood of the root* r *then* r *is a superattracting fixed point of* $F(x) := x - \dfrac{f(x)}{f'(x)}$.

PROOF. $F(r) = r$ and $F'(x) = f(x)f''(x)/(f'(x))^2$. □

REMARK 2.2.26. A small first derivative might cause the intersection of the tangent line with the x-axis to go quite far from x_0. The hypothesis $|f''(x)| < M$ holds whenever f'' is continuous.

At a superattracting fixed point we have quadratically (that is, superexponentially) converging iterates, as in the case of the fixed point of the quadratic map f_2 in Section 2.5. In other words, the error is approximately squared in every iteration.

This argument only works if the initial guess is fairly good. With an unfortunate initial choice the iterates under F can behave rather erratically. In other words, F has an attracting fixed point, but may otherwise have quite complicated dynamics.

The special case of extracting roots by the Newton method had an ancient precursor.

PROPOSITION 2.2.27. *Approximating* $\sqrt{z}$ *by the Newton method with initial guess 1 is the same as using the first components of the Greek root extraction method in* (1.3.1).

PROOF. With initial guess 1 the Newton method gives the recursion

$$x_0 = 1, \quad x_{n+1} = x_n - \frac{x_n^2 - z}{2x_n} = \frac{1}{2}(x_n + \frac{z}{x_n}).$$

The Greek method starts with $(x_0, y_0) = (1, z)$, and the recursion $(x_{n+1}, y_{n+1}) = f(x_n, y_n)$ defined by (1.3.1) has the property that $y_n = z/x_n$. Therefore we have

$$x_{n+1} = \frac{x_n + y_n}{2} = \frac{1}{2}(x_n + \frac{z}{x_n}).$$

□

i. Applications of the Contraction Principle. The foremost tool we introduced in this chapter is the Contraction Principle. It is one of the most important individual facts in analysis and dynamical systems. Its applications are both diverse and fundamental. We do not only have numerous applications of it in the course of our development of dynamics, but several basic facts that underly the theory are consequences of the Contraction Principle. Chapter 9 is devoted to such applications. That chapter provides the Inverse and Implicit Function Theorems, which are fundamental to analysis (Theorem 9.2.2 and Theorem 9.2.3). As we mentioned, the fixed point of a contraction persists under perturbations, and Chapter 9 gives the most general condition on a fixed point for such persistence (Proposition 9.3.1). Also, the theory concerning existence and uniqueness of differential equations, on which in a manner of speaking half of dynamics is based, is derived there from the Contraction Principle (Theorem 9.4.3). A result central to dynamical systems of the type discussed in Chapter 7 is the stable manifold theorem (Theorem 9.5.2, see the comments at the end of Section 10.1). It also depends crucially on the Contraction Principle.

Exercises

EXERCISE 2.2.1. Show that entering any number on a calculator and repeatedly pressing the sin button gives a sequence that goes to zero. Prove that convergence is not exponential if we use the radian setting and exponential if we use the degree setting. In the latter case find out, how many iterates are needed to obtain a number less than 10^{-10} times the initial input.

EXERCISE 2.2.2. If one enters a number greater than 1 on a calculator and repeatedly hits the square root key, the resulting numbers settle down eventually. Prove that this always happens and determine the limit. If the calculator keeps k binary digits internally, roughly how long does it take for the sequence to settle down to this limit of accuracy?

EXERCISE 2.2.3. Do the previous exercise for initial values in $(0, 1]$.

EXERCISE 2.2.4. Show that x^2 defines a λ-contraction on $[-\lambda/2, \lambda/2]$.

EXERCISE 2.2.5. This is a variation on Fibonacci's problem of rabbit populations, taking mortality into account.

A population of polar lemmings evolves according to the following rules. There are equal numbers of males and females. Each lemming lives for two years and dies in the third winter of its life. Each summer each female lemming produces an offspring of four. In the first summer there is one pair of one-year old lemmings. Let x_n be the total number of lemmings during the nth year. Use the Contraction Principle to show that x_{n+1}/x_n converges to a limit $\omega > 1$. Calculate ω.

EXERCISE 2.2.6. Let x be a fixed of point of a map f on the real line such that $|f'(x)| = 1$ and $f''(x) \neq 0$. Show that arbitrarily close to x there is a point y such that the iterates of y do not converge to x.

EXERCISE 2.2.7. Which of these are convex: $\{(x,y) \in \mathbb{R}^2 \mid xy > 1\}$, $\{(x,y) \in \mathbb{R}^2 \mid xy < 1\}$, $\{(x,y) \in \mathbb{R}^2 \mid x + y > 1\}$, $\{(x,y) \in \mathbb{R}^2 \mid x > y^2\}$.

EXERCISE 2.2.8. Prove that the norm of a matrix defined in (2.2.8) is a norm in the sense of Definition A.1.26.

EXERCISE 2.2.9. Show that $\|A\| \leq \sqrt{\sum_{i,j} a_{ij}^2}$ for any $n \times n$ matrix $A = (a_{ij})_{1 \leq i,j \leq n}$.

EXERCISE 2.2.10. Show that $\|A\| \geq |\det A|^{1/n}$ for any $n \times n$ matrix $A = (a_{ij})_{i,j=1,\dots,n}$.

EXERCISE 2.2.11. Prove that the norm of a matrix is a continuous function of its coefficients.

Problems for further study

EXERCISE 2.2.12. Construct an example of an open connected subset U of the plane $\mathbb{R}^2$ and a continuously differentiable map $f\colon U \to U$ such that $\|Df_x\| < \lambda < 1$ for all $x \in U$ but f is not a contraction. (Such a set cannot be convex.)

EXERCISE 2.2.13. Suppose that I is a closed bounded interval and $f\colon I \to I$ is such that $d(f(x), f(y)) < d(x,y)$ for any $x \neq y$ (this is weaker than the assumption of the Contraction Principle). Prove that f has a unique fixed point $x_0 \in I$ and $\lim_{n\to\infty} f^n(x) = x_0$ for any $x \in I$.

EXERCISE 2.2.14. Show that the assertion of the previous exercise is not valid for $I = \mathbb{R}$ by constructing a map $f\colon \mathbb{R} \to \mathbb{R}$ such that $d(f(x), f(y)) < d(x,y)$ for $x \neq y$, f has no fixed point, and $d(f^n(x), f^n(y))$ does not converge to zero for some x, y.

3. Nondecreasing maps of an interval and bifurcations

As a next step we look at maps that may have several fixed points, but otherwise hardly more complicated behavior than contractions. Here we will see examples that, unlike contractions, can change radically under perturbations.

a. Nondecreasing interval maps. We now study the situation where the dynamics is similar to that of a contraction, but there is no guarantee of exponentially fast convergence to a fixed point. This situation is instructive because it demonstrates an important method in low-dimensional dynamics, the systematic use of the Intermediate Value Theorem.

DEFINITION 2.3.1. If $I \subset \mathbb{R}$ is an interval then $f\colon I \to \mathbb{R}$ is said to be *increasing* if $x > y \implies f(x) > f(y)$ and *decreasing* if $x > y \implies f(x) < f(y)$. We say that f is *nondecreasing* if $x \geq y \implies f(x) \geq f(y)$ and non*increasing* if $x \geq y \implies f(x) \leq f(y)$.

The simple example situation, and a useful building block in the theory of nondecreasing maps is the following observation.