Ordinary Differential Equations, an intuitive introduction

Mark Levi
Department of Mathematics
Penn State University
levi@math.psu.edu.

November 28, 2017

Contents

Chapter 1

Background

The subject of ordinary differential equations (ODEs) combines calculus, geometry, linear algebra, and more. This chapter emphasizes ideas from these subjects that are sometimes lost in standard courses (on calculus and linear algebra mostly). ODEs are easy if you are comfortable with this background; they are not if you are not.

1.1 Functions as deformations.

Definition. The *function* is a pair of sets X, Y and a *rule* which assigns to every element of X one element (no more, no less) of Y. Figure 1.1 illustrates this concept on three examples.[†] *The function can therefore be thought of as a deformation* of one set X into another set Y, where each point x gets moved to the new position $f(x)$. This is a way to think of function avoiding graphs. For example, the function $y = 2x$ is the stretching of the x–axis by the factor of 2; $y = -x$ is the reflection with respect to 0; and $y = x^2$ folds the line and maps it to the positive half–line.

Question: how to visualize $f(x) = \sin x$ as a deformation of a line?

Graphs are often useful, but not always. One problem is the dimension of the space in which the graph lies. For example, for $f : \mathbb{R}^2 \to \mathbb{R}^2$, the graph lives in $\mathbb{R}^4$, and is not easy to visualize. Here is another example, quite apart from the dimensional aspect: the familiar chain rule $\frac{d}{dx} f(g(x)) = f'(g(x))\, g'(x)$ is easier to explain without using graphs than with.

[†]It is common practice not to mention X and Y and just give the rule (for example, $y = \frac{1}{x}$), with the understanding that X includes all x–values for which the rule makes sense. In this example, $X = \mathbb{R} \setminus \{0\}$, $Y = \mathbb{R} \setminus \{0\}$.

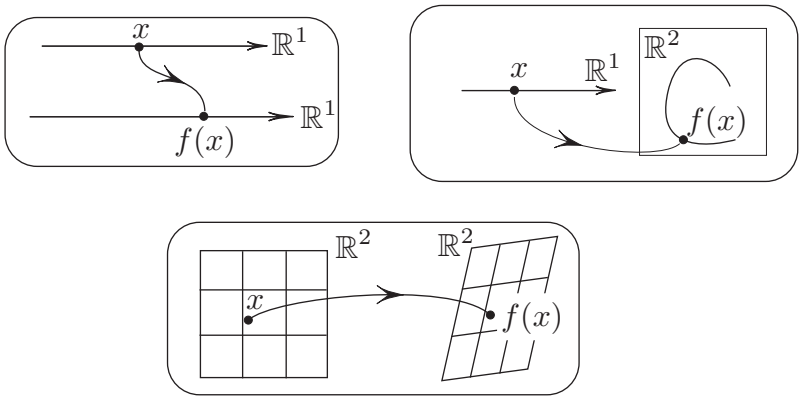

Figure 1.1: Illustrating functions for the cases when X and Y are sets in $\mathbb{R}$ or $\mathbb{R}^2$, in different combinations.

Some intuition Here is another useful way to think of a function: think of the t–axis as a (one dimensional) trackpad, and think of the x–axis as a screen (also one–dimensional). Then you can get a feel of a function by imagining moving the "finger" t and watching the "cursor" $x(t)$ respond. For example, if $x = t^2$, as the finger t moves from $-\infty$ to $+\infty$, the cursor t^2 moves from $+\infty$, touches 0 and goes back to $+\infty$.

A question. What is a common sense interpretation of the derivative $x'(t)$ in the context of the preceding paragraph?

Exercise. Build in your mind a dynamical visualization of the following familiar functions: $1/t$, $\sin t$, $\cos t$, $\tan t$, e^t.

1.2 Derivative as a stretching factor.

There is another meaning of the derivative, besides the slope or the velocity. The derivative

$$f'(t) = \lim_{h \to 0} \frac{f(t+h) - f(t)}{h} \tag{1.1}$$

is the limit of the ratio of the lengths of two segments, the image and the premage (Figure 1.2) – in other words, $f'(t)$ is simply *the stretching coefficient* at t. To test your intuition, see if it is obvious that $(2t)' = 2$, or that $(t^2)'_{t=0} = 0$. In the "trackpad–cursor" interpretation, $f'(t)$ is the sensitivity coefficient. When you set the trackpad sensitivity in the computer preferences, you are prescribing the value of the derivative.

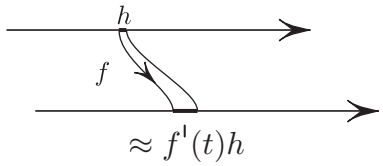

Figure 1.2: $f'(t)$ as the stretching factor (of an infinitesimal neighborhood).

Here is an English-to-math translation problem, with an interesting view of the derivative.

Problem 1.1. Find the integer value of 1.0001^{10000} and prove your claim.

Problem 1.2. A pen is moving along the x–axis; the pen's position at time t is a given function of t, i.e. $x = x(t)$. Assume that x is a monotone function. The pen leaks ink into the paper at the constant rate r grams per second. Let us define the *thickness* of the trace left by the pen at point x as the *amount of ink per unit length* at x (more precisely, as the appropriate limit over a shrinking segment containing x.) Denote this thickness by $w(x)$ (w for *width*).

1. Express the definition of $w(x)$ by a formula (involving the limit).
2. Express $w(x)$ in terms of the derivative x'.
3. What happens to $w(x)$ at the point x_m where the pen reverses the direction of its motion? Can you explain the answer by a common sense argument, not using the formula?

Remark 1.1. *The above problem is related to the brightness of the rainbow, and to the caustics – the brightly lit curves seen on the bottom of an empty cup in the sun.*

1.3 The chain rule by stretching

If I stretch a piece of rubber band by the factor of (say) 2, and then follow this by another stretching by the factor of 3, then the net result is the $2 \cdot 3 = 6$ – fold stretching. The chain rule

$$\frac{d}{dt} f(g(t)) = f'(g(t))g'(t). \tag{1.2}$$

boils down to this fact, as explained by Figure 1.3: Think of the segment $[t, t+h]$ as the rubber band being stretched* by applying function g to

*The word "stretch" is meant in the generalized sense: stretching by a factor < 1 is really a contraction. And if the factor is negative, it involves also a reversal of direction.

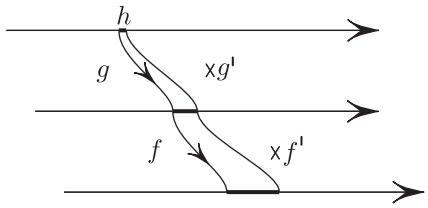

Figure 1.3: The chain rule: consecutive stretchings by factors g' (evaluated at t) and f' (evaluated at $g(t)$) amount to one stretching by factor $f'g'$.

it, with a subsequent stretch by applying the function f. The stretching coefficients are first $g'(t)$ and then $f'(g(t))$, so that the net stretching is the product in (1.2).

1.4 The definition of $\exp(t)$ and of e

My college textbook's definition of

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n, \tag{1.3}$$

while simple, seemed to me somewhat unmotivated and arbitrary. We then spent a few pages learning what exactly does the power e^t means for t being a fraction or even an irrational number. Following this, we proved the immunity to differentiation property:

$$\frac{d}{dt}e^t = e^t. \tag{1.4}$$

In plain language, the exponential grows at the rate equal to the amount present. For example, five pounds grow at the rate of five pounds per second (if t is in seconds)* The presentation just outlined is logically optimal, but it puts the cart (1.3) before the horse (1.4). It is (1.4) that is the workhorse in science (physics, biology, engineering). This is because many things in life change at the rate that is proportional to the amount present. The only rule of change that is simpler is the one with the *constant* rate of change.

The definition of $\exp(t)$

Since the exponential is omnipresent, it is worth rethinking afresh. So let us pretend we never saw e, i.e. (1.3), and instead define the exponential

*only at the instant where $e^t = 5$.

function $\exp(t)$ not as a power of e, but as the function that satisfies the growth property (1.4), as well as the condition suggested by $e^0 = 1$. In short, we make the following definition.

Definition 1.1. $\exp(t)$ *is the function* f *which satisfies*

$$\begin{aligned} f'(t) &= f(t) \\ f(0) &= 1 \end{aligned} \tag{1.5}$$

There is one and only one function (the proof of this will be given the chapter on existence and uniqueness), and this function is denoted by $\exp(t)$. In summary, exp is defined uniquely by the properties $\exp'(t) = \exp(t)$ and $\exp(0) = 0$.

Problem 1.3. Let $A(t)$ denote the amount of money in a bank account at time t (measured in years). The starting balance is \$1 and the account is compounded continuously at the 100% annual rate. Translate the preceding sentence into a mathematical notation.

Problem 1.4. Given that $f'(t) = f(t)$ (without $f(0) = 1$) how is $f(t)$ related to exp? Same question for f satisfying $f'(t) = 5f(t)$. (Since the goal of this short section is to rethink the exponentials from scratch, all you are allowed to use is the definition (1.5)), and not your prior knowledge of e^t).

Defining e and proving $\exp(t) = e^t$

Eq. (1.5) must be our starting point, since this is the only information on the exponential at this point. The plan is to replace the derivative by it difference, and then take an appropriate limit. Let us treat t as fixed, and divide $[0, t]$ into n short subintervals of equal lengths, which must then be $h = t/n$. Replacing the derivative in (1.5) by the ratio we get an approximation to (1.5):

$$\frac{f_{k+1} - f_k}{h} = f_k, \quad f_0 = 1, \tag{1.6}$$

for $k = 0, \ldots, n-1$. Here $f_k \approx \exp(k\frac{t}{n})$, and in particular $f_n \approx f(n\frac{t}{n}) = f(t)$. More precisely, $\lim_{n\to\infty} f_n = f(t)$; we omit the rigorous proof.* From (1.6) we get the recursion relation

$$f_{k+1} = (1 + h)f_k, \quad f_0 = 1,$$

*The last statement claims convergence of Euler's method of numerical solution of ODEs. In fact, our approximation of the solution of (1.5) follows Euler's method, as will be discussed in the section on numerical solutions of ODEs.

implying that $f_n = (1+h)^n$. Since $n = t/h$ by the definition, we obtain

$$f_n = (1+h)^{t/h} = \big((1+h)^{1/h}\big)^t. \tag{1.7}$$

Finally,

$$f(t) = \lim_{h\to 0} f_n = \lim_{h\to 0}\big((1+h)^{1/h}\big)^t \overset{(1.3)}{=} e^t.$$

Now that we proved that $\exp(t) = e^t$, we know that $\exp(t+s) = \exp(t)\exp(s)$, using the property of the powers. It would be interesting to see how to prove this property *directly* from the defining property (1.5). This is done next.

The following two problems outline an alternative approach to showing that $\exp(t)$ defined by the initial value problem (1.5) is an exponential function e^t.

Problem 1.5. Show that if $f(t)$ satisfies

$$f'(t) = f(t) \quad \text{for all} \quad t \in \mathbb{R}, \tag{1.8}$$

and

$$f(0) = 1, \tag{1.9}$$

then

$$f(t+s) = f(t)f(s) \quad \text{for all} \quad t,\ s \in \mathbb{R} \tag{1.10}$$

without using the fact that $f(t) = ce^t$. You can use the fact (proven later on) two if two functions $x(t)$ and $y(t)$ satisfy the same ODE $\dot z = F(z)$, where F is a differentiable function, with the same initial condition then they are identically equal.* are equal (this is a very special consequence of the uniqueness theorem proven later on).

Hint. treat s as an arbitrary constant, and t as a variable. Show that each side of (1.10) satisfies the same ODE. Then the uniqueness theorem.

Solution. Pick an arbitrary s and fix it. Denote the two sides of (1.10) by $x = x(t)$ and $y = y(t)$ respectively (they depend also on s, but we treat it as fixed and so do not include in the notation). We will show that both x and y satisfy the same initial value problem

$$\begin{cases} \dot z = z \\ z(0) = f(s); \end{cases} \tag{1.11}$$

by the uniqueness theorem, two solutions of the same IVP coincide, which would imply that $x = y$ for all t (and any choice of s), completing the proof.

The initial conditions coincide: indeed,

$$x(0) = f(0+s) = f(s), \quad \text{and} \quad y(0) = f(0)f(s) \overset{(1.9)}{=} f(s).$$

*To repeat, the theorem you can use states: if $\dot x(t) = F(x(t))$, $\dot y(t) = F(y(t))$ (where F is a given function) for all t and if $x(0) = y(0)$, then $x(t) = y(t)$ for all t.

Verifying that x and y satisfy the same ODE in (1.11):

$$\dot{x} = \frac{d}{dt} f(t+s) \overset{\text{chain rule}}{=} f'(t+s)\frac{d}{dt}(t+s) \overset{(1.8)}{=} f(t+s) = x;$$

$$\dot{y} = \frac{d}{dt} f(t)f(s) = f'(t)f(s) \overset{(1.8)}{=} f(t)f(s) = y.$$

This completes the proof.

Problem 1.6. Show that if a contiuous function $f : \mathbb{R} \to \mathbb{R}$ satisfies (1.10) then

$$f(t) = a^t, \quad \text{where} \quad a = f(1). \tag{1.12}$$

Proof. From (1.10) it follows by induction that

$$f(t_1 + \ldots + t_n) = f(t_1) \ldots f(t_n) \tag{1.13}$$

for any collection of $n > 2$ of real numbers t_k. Choosing n equal values $t_k = 1/n$, where n is any positive integer, we obtain:

$$f(1) = f\Big(\underbrace{\frac{1}{n} + \ldots + \frac{1}{n}}_{n \text{ terms}}\Big) \overset{(1.13)}{=} f\left(\frac{1}{n}\right)^n,$$

and extracting nth root we get

$$f\left(\frac{1}{n}\right) = f(1)^{\frac{1}{n}}, \tag{1.14}$$

proving (1.12) for special $t = 1/n$, where $n > 0$ is an integer. Taking (1.14) to an integer power $m > 0$, we get

$$f\left(\frac{1}{n}\right)^m = f(1)^{\frac{m}{n}};$$

the left–hand side

$$f\left(\frac{1}{n}\right)^m \overset{(1.13)}{=} f\left(\frac{m}{n}\right),$$

and thus (1.12) holds for any positive rational t. Setting $s = t = 0$ in (1.10) shows that $f(0) = 1$ or 0. In the latter case $f(t) = f(t+0) = f(t)f(0) = 0$ for all t and (1.12) holds trivially, so we focus on $f(0) = 1$. How to deal with negative t? We have

$$1 = f(0) = f(t + (-t)) \overset{(1.10)}{=} f(t)f(-t),$$

concluding that the sign change of t inverts $f(t)$. Let us now change t from positive rational to negative rational in both sides of (1.12); since both sides undergo inversion, the equality remains true. It remains to prove (1.12) for irrational t. For an irrational t, take a sequence of rationals: $t_n \to t$; we proved that

$$f(t_n) = f(1)^{t_n}.$$

Taking the limit of both sides

$$\lim_{n\to\infty} f(t_n) = \lim_{n\to\infty} f(1)^{t_n};$$

by continuity of f the left–hand side is $f(t)$; by continuity of a^t ($a = f(1)$), the right–hand side is $f(1)^t$. This completes the proof.

Problem 1.7. Let $A(t)$ be the number of dollars in a bank account, compounded continuously at the annual rate of 10%.

1. Write the ODE satisfied by A.
2. How does the form of the ODE change if instead of dollars we use yen?
3. How does the form of the ODE change if instead of one year as the unit of time we use one day?

Answer. (1) A satisfies the ODE $\dot{x} = 0.1x$. (2) $B = rA$, where r is the exchange rate, satisfies the same ODE. (3) Let t be measured in days, so $t = 365t_y$ where t_y is the number (possibly even irrational) of years. The money at time t days is $D(t) = A(t_y) = A(t/365)$, and so $\dot{D}(t) = \frac{1}{365}\dot{A}(t/365) = \frac{1}{365}\dot{A}(t/365)$, or $\dot{D} = \frac{1}{365}$.

Problem 1.8. Consider the direction field in the (t, x)–plane associated with the ODE $\dot{x} = x$; in other words, the slope of the field at the point (t, x) is x.

1. Consider the stretching $(t, x) \mapsto (t, ax)$ in the x–direction, deforming the curves of the vector field. What ODE do the deformed curves satisfy?
2. The same question: the curves are deformed by stretching in the t–direction by factor b: every point (t, x) on the curve is moved to (bt, x). What is the ODE for the curves thus stretched?

The preceding two problems give a geometrical view of the fact that (i) any solution of $\dot{x} = x$ multiplied by a constant a is still a solution, and (ii) if the time t is measured in new units $\tau = bt$, then in the new units x evolves according to $\frac{d}{d\tau}x = b^{-1}x$ (think of t measured in days and τ measured in years, i.e. $b = 365$, and it stands to reason that the per annum rate is 365 times greater than the daily rate.)

Problem 1.9. Write a difference equation which is the analog of the ODE $\dot{x} = ax$ and solve it.

Divide the time interval $[0, t]$ into n equal pieces of length $h = t/n$, separated by equally spaced points $x_k = k\frac{t}{n}$. Let x_k denote the approximate values of x at t_k, where $k = 0, \ldots, n$. Replacing the derivative by the difference in our ODE, we get

$$\frac{x_{k+1} - x_k}{h} = ax_k,$$

or

$$x_{k+1} = (1 + ah)x_k.$$

Thus

$$x_n = (1+ah)x_{n-1} = (1+ah)(1+ah)x_{n-2} = \ldots = (1+ah)\ldots(1+ah)x_0,$$

or

$$x_n = (1+ah)^n x_0. \tag{1.15}$$

To find the limit as $n \to \infty$, i.e. as the partition of the interval gets to be finer and finer, we call $ah = \varepsilon$, and note that $n = at/\varepsilon$, and $\varepsilon \to 0$ as $n \to \infty$, so that

$$\lim_{n\to\infty} x_n = \lim_{\varepsilon\to 0}\left((1+\varepsilon)^{\frac{1}{\varepsilon}}\right)^{at} = e^{at}.$$

1.5 The Fundamental Theorem of Calculus

The fundamental theorem of calculus is a generalization of the fact that the addition and subtraction are inverse operations of each other. For the same deep–down reason, the integration and the differentiation are inverses of each other, as we shall see. We first state the theorem in its two versions, give a word–free explanation of the first version, then establish their equivalence, and finally prove the theorem.

Two versions and a word–free proof

Theorem 1.1 (The fundamental theorem of calculus–version 1)**.** *If f is a continuously differentiable function defined on $[a,b] \subset \mathbb{R}$, where $a < b$, then*

$$\int_a^b f'(s)\,ds = f(b) - f(a), \tag{1.16}$$

or equivalently, denoting $f'(x) = g(x)$ and $f(x) = G(x)$ (the antiderivative of g):

$$\int_a^b g(s)\,ds = G(b) - G(a). \tag{1.17}$$

Figure 1.4 explains (1.16) without words. Here is an equialent version of FTC:

Theorem 1.2 (The fundamental theorem of calculus–version 2)**.** *Let g be a continuous function an an interval $[a,b] \subset \mathbb{R}$. Then*

$$\frac{d}{dt}\int_a^t g(s)\,ds = g(t), \tag{1.18}$$

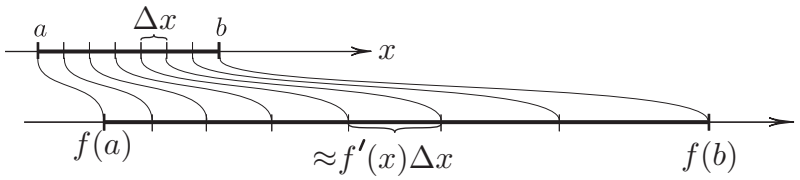

Figure 1.4: An explanation of the fundamental theorem of calculus (1.16). This can be turned into a rigorous proof (see Problem 1.10 on page 18.

for all $t \in (a, b)$*, see Figure 1.6.*

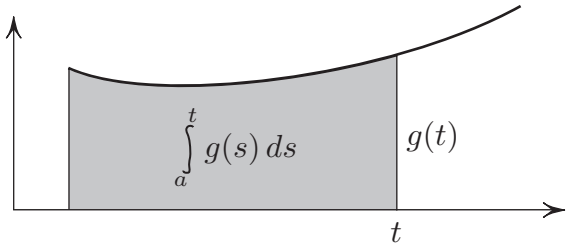

Figure 1.5: The fundamental theorem (1.18): derivative of the area equals the height of the moving boundary.

Some intuition The relation (1.18) states that *as the right wall of the shaded area in Figure* **??** *advances to the right with speed* 1*, the area grows at the rate equals to the length of the wall.* The same idea is behind the fact that the area $A(t) = \pi t^2$ and the length $L(t) = 2\pi t$ of the circle of radius t are related: $\frac{d}{dt}A(r) = L(t)$, an exact analog of (1.18). For the circle, just as for Figure 1.6, the area grows at the rate equal to the length of the advancing boundary times the speed of the boundary, which is 1.

A historical perspective The theorem is fundamental because it links two fundamental concepts of calculus: integration (defined as the limit of a sum, and NOT as an antiderivative) and differentiation, showing that the two operations are inverses of each other in the sense that differentiation undoes integration, see (1.18). Historically, integration, although a harder task than differentiation, is about two thousand years older: Archimedes computed integrals 2,400 years ago, but it required his genius to do so.

Nowadays millions of people can now do more than Archimedes could, all thanks to (1.17).

The equivalence of the two versions

Before proving either of the two versions, let us show that they are equivalent. To prove (1.17)⇒(1.18), we start with (1.17), and think of b as a variable, which we emphasize by denoting $b = t$. Then we differentiate both sides and use $G' = g$, obtaining (1.18) as desired. To show the reverse implication (1.18)⇒(1.17), let G be an antiderivative of g, i.e. let $G' = g$, so that (1.18) amounts to

$$\frac{d}{dt}\int_a^t g(s)\,ds = \frac{d}{dt}G(t),$$

which shows that $\int_a^t g(s)\,ds$ and G differ by a constant:

$$\int_a^t g(s)\,ds = G(t) + c;$$

we find that $c = -G(a)$ by setting $t = a$, and thus

$$\int_a^t g(s)\,ds = G(t) - G(a), \tag{1.19}$$

which gives (1.17) after setting $t = b$. ♢

In the next subsection we give a more formal proof of the fundamental theorem. An alternative proof can be obtained by formalizing the idea of Figure 1.4.

Proof of the fundamental theorem

Since the two versions are equivalent, it suffices to prove, say, (1.18), which we do here.* We now begin the formal proof of (1.18), i.e. of the claim that

$$\lim_{h\to 0}\frac{1}{h}\int_t^{t+h} g(s)\,ds = g(t). \tag{1.20}$$

Wishing to estimate the integral, we set $\underline{g} = \inf_{s\in[t,t+h]} g(s)$ and $\overline{g} = \sup_{s\in[t,t+h]} g(s)$ (see Figure 1.6), so that

$$\underline{g} \le g(s) \le \overline{g} \quad \text{for all} \quad s \in [t, t+h];$$

*The version (1.16) can be proved independently by formalizing the idea of Figure 1.4.

integration gives $\underline{g}\,h \le \int_t^{t+h} g(s)\,ds \le \overline{g}\,h$ (see Figure 1.6), and thus

$$\underline{f} \le \frac{1}{h}\int_t^{t+h} g(s)\,ds \le \overline{f}. \tag{1.21}$$

Intuitively, (1.21) says that the average height of the graph of g is bounded by the height of the highest and the lowest points. By continuity of g, $\underline{f} \to g(t)$ and $\underline{f} \to g(t)$ as $h \to 0$, and by the sandwich theorem (1.20) holds. $\diamondsuit$

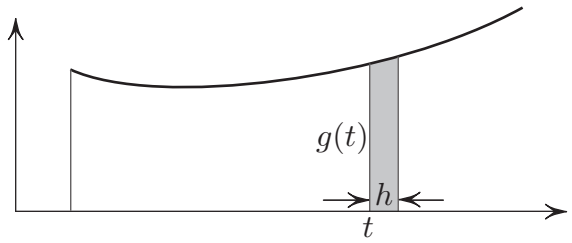

Figure 1.6: Proof of the FTC.

Problem 1.10. Prove the FTC using Figure 1.4 as a guide, by writing the length of the image interval as the sum of subintervals and then expressing the latter using the definition of the derivative $f(x+\Delta x) - f(x) = f'(x)h + o(\Delta x)$:

$$f(b) - f(a) = \sum\big(f(t_{i+1} - f(t_i)\big) = \sum\big(f'(t_i)\Delta t + r_i\big).$$

Show that if f' is continuous, then r_i are uniformly small for small Δx (that is, for any prescribed ε there exists $\delta > 0$ such that $|r_i| < \varepsilon$ for all $0 < \Delta x < \delta$. Take the limit of the above to get (1.16).

Problem 1.11. Find a geometrical explanation/interpretation of the familiar facts

1. $\frac{d}{dx}x^2 = 2x$
2. $\frac{d}{dx}x^3 = 3x^2$
3. $\frac{d}{dt}uv = u'v + uv'$
4. $\frac{d}{dt}uvw = u'vw + u'w + uvw'$

Solution of #1: When increasing x by h we increase the area of the square by *two* rectangles each of area xh plus a square of area h^2, so that the change in area is $2xh + h^2$. The derivative is, by the definition, the limit of the ratio of the increment of area to the increment of x:

$$\frac{d}{dx}x^2 = \lim_{h\to 0}\frac{2xh + h^2}{h} = \lim_{h\to 0}(2x + h) = 2x.$$

Problem 1.12. State and prove the discrete equivalent of the fundamental theorem of calculus.

The following problem involves the discrete analog of antiderivative: given the first difference of a sequence, we are asked to find the sequence itself.

Problem 1.13. Find A_n in finite form, i.e. in the form such that the number of terms does not grow with n in each of the following cases:

1. $A_{n+1} - A_n = \frac{1}{n(n+1)}$
2. $A_{n+1} - A_n = \frac{2n+1}{n^2(n+1)^2}$
3. $A_{n+1} - A_n = n$
4. $A_{n+1} - A_n = n^2$

Problem 1.14. Use the discrete version of the FTC to find the sums

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}, \quad \sum_{k=1}^{\infty} \frac{2k+1}{k^2(k+1)^2}, \quad \sum_{k=1}^{n} k^2, \quad \sum_{k=1}^{n} k^3.$$

Solution for $\sum_{k=1}^{n} k^2$. We will find a sequence A_k whose differences give the terms of our sum:

$$A_{k+1} - A_k = k^2, \quad k - 1, 2, \ldots, \tag{1.22}$$

so that the sum telescopes. Seeking A_k in the form

$$A_k = ak^3 + bk^2 + ck,$$

we substitute it in (1.22), and group by powers of k:

$$3ak^2 + (2a + 2b)k + (a + b + c) = k^2;$$

this identity, and thus the desired (1.22), will hold if we choose

$$3a = 1, \ 2a + 3b = 0, \ a + b + c = 0.$$

This dictates the choice of $a = 1/3, \ b = -1/2, \ c = 1/6$, and thus the desired

$$A_k = \frac{1}{3}k^3 - \frac{1}{2}k^2 + \frac{1}{6}k.$$

Finally,

$$\sum k^2 = (\cancel{A_2} - A_1) + \ldots + (A_{n+1} - \cancel{A_n}) = A_{n+1} - A_1,$$

and the answer is

$$\frac{1}{3}(n+1)^3 - \frac{1}{2}(n+1)^2 + \frac{1}{6}(n+1) - \left(\frac{1}{3} - \frac{1}{2} + \frac{1}{6}\right)$$

Simplification is left out.

1.6 The discrete–continuous analogy

This short section gives the discrete analogs of some continuous concepts. In particular, the analogy brings out the fact that the FTC amounts to the collapse of telescoping sums. The same collapse explains Green's Theorem, the divergence theorem and Stokes's theorem, as explained in sections on these theorems.

Continuous	**Discrete**
$t \in \mathbb{R}$ independent variable	$n \in \mathbb{Z}$ indpendent variable
$f(t)$ value of f	a_n element of a sequence
$f'(t) = \lim_{h\to 0} \frac{f(t+h)-f(t)}{h}$	$a_{n+1} - a_n$ first difference
$f''(t)$ second derivative	$(a_{n+1} - a_n) - (a_n - a_{n-1}) = a_{n-1} - 2a_n + a_{n+1}$, second difference
$\int_0^t f(s)\, ds$	$\sum_{k=0}^n a_k$
$\int_0^t f'(s)\, ds = f(t) - f(0)$, the FTC	$\sum_{k=0}^n (a_{k+1} - a_k) = a_{n+1} - a_0$, a telescoping sum
$\frac{d}{dt}\int_0^t f(s)\, ds = f(t)$, the FTC	$\sum_{k=0}^n a_k - \sum_{k=0}^{n-1} a_k = a_n$
$F'(t) = f(t)$: F is an antiderivative of f	$A_{n+1} - A_n = a_n$; A_n is an "antidifference" of a_n

1.7 Parametric equations of basic curves: ellipses, spirals, hyperbolas.

Problem 1.15. Identify the curve $ax^2 + 2bxy + cy^2 = 1$ (e.g. if it's an ellipse, find its major and minor axes and their orientation).

Problem 1.16. Consider the parametric curve

$$\begin{aligned} x &= a\cos t + b\sin t \\ y &= c\cos t + d\sin t. \end{aligned} \tag{1.23}$$

Show that:

1. the curve is an ellipse iff the coefficient matrix has a nonzero determinant.
2. Show that the directions of the ellipse's axes are the eigendirections of the matrix AA^T, where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.
3. Find the length and the direction of the semiaxes of the ellipse.

To answer questions 2 and 3 at once, let $\mathbf{r}$ be an arbitrary point on the alleged ellipse, i.e. let $\mathbf{r} = A\mathbf{u}$, where $\mathbf{u}$ is a unit vector. Assuming A to be invertible, we have $\mathbf{u} = A^{-1}\mathbf{r} = B\mathbf{r}$, where $B = A^{-1}$. The equation of the alleged ellipse is

$$\langle B\mathbf{r}, B\mathbf{r}\rangle = 1, \quad \text{i.e.} \quad \langle B^T B\mathbf{r}, \mathbf{r}\rangle = 1. \tag{1.24}$$

Since $B^T B$ is symmetric matrix, and a positive one, it has an orthonormal basis $\mathbf{e}_1,\ \mathbf{e}_2$, and positive eigenvalues λ_1, λ_2. Decomposing $\mathbf{r} = X\mathbf{e}_1 + Y\mathbf{e}_2$, where X and Y are the Cartesian coordinates in the eigenbasis, we rewrite (1.24) as

$$\lambda_1 X^2 + \lambda_2 Y^2 = 1,$$

which is an equation of an ellipse with semiaxes $1/\sqrt{\lambda_k}$, $k = 1,\ 2$. Finally, $B^T B = (AA^T)^{-1}$, so that $1/\lambda_k$ are the eigenvalues of AA^T. In short, *the semiaxes of the ellipse are square roots of the eigenvalues of AA^T. And the directions of the two semiaxes are given by the eigenvectors of AA^T.*

1.8 The linearization of a function $f : \mathbb{R} \to \mathbb{R}^2$

In Calculus III the linearization $L(x, y)$ of a function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ at the point (x_0, y_0) was defined by the property that the graph of L is the tangent plane to the graph of f at $(x_0, y_0, f(x_0, y_0)) \in \mathbb{R}^3$, or equivalently, by

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0). \tag{1.25}$$

Equivalently, L can be defined by the property that $L = a + b(x - x_0) + c(y - y_0)$ is the linear function that approximates f best, in the sense that

$$|f(x, y) - L(x, y)| = o(|x - x_0| + |y - y_0|). \tag{1.26}$$

This a less secretive definition than (1.25), since it gives the whole reason for introducing L; the explicit formula (1.25) is an immediate consequence of (1.26).

To re–emphasize this approximation property of L, one often rewrites (1.26) as

$$f(x,y) = f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)+o(|x-x_0|+|y-y_0|). \quad (1.27)$$

Recall that the differential of f is simply the "increment part" of L, i.e. the linear approximation to the change of f; in the commonly used notations $dx = x - x_0$, $dy = y - y_0$ for the deviations from x_0, y_0,

$$df = f_x(x_0,y_0)dx + f_y(x_0,y_0)dy.$$

Prove that (1.26) implies (1.25). Hint. (i) Show that $a = f(x_0,y_0)$ by setting $x = x_0,\ y = y_0$ in (1.26); (ii) show that $b = f_x(x_0,y_0)$ by dividing (1.26), setting $y = y_0$ and taking the limit as $x \to x_0$.

Problem 1.17. The dimensions of a rectangular solid were with dimensions a, b, c were changed by amounts da, db, dc. Use the differential to find an approximate resulting change of the volume, both absolute and relative.

The approximation idea (1.27) can be used to prove the chain rule.

Problem 1.18. Prove the chain rule $\frac{d}{dt}f(x(t),y(t)) = f_x x' + f_y y'$.

1.9 The directional derivative and the gradient

Definition 1.2. *The directional derivative*

The following problem captures two concepts in one formula: the directional derivative and the gradient.

Problem 1.19. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function with continuous partial derivatives at $\mathbf{x} \in \mathbb{R}^2$, and let $\mathbf{u} \in \mathbb{R}^2$ be a *unit* vector. Show that there exists a vector $\mathbf{V} \in \mathbb{R}^2$ such that

$$\frac{d}{ds}f(\mathbf{x}+s\mathbf{u})_{|\,s=0} = \mathbf{V}\cdot\mathbf{u}, \quad (1.28)$$

where $\cdot$ denotes the dot product, and find the expression for $\mathbf{V}$. Does $\mathbf{V}$ depend on $\mathbf{u}$?

Solution. Denoting $\mathbf{x} = \langle x_1, x_2\rangle$ and $\mathbf{u} = \langle u_1, u_2\rangle$, we have

$$\frac{d}{ds}f(\mathbf{x}+s\mathbf{u}) = \frac{d}{ds}f(x_1+su_1, x_2+su_2) = f_{x_1}\frac{d}{ds}(x_1+u_1 s) + f_{x_2}\frac{d}{ds}(x_2+su_2),$$

by the chain rule; setting $s = 0$ gives

$$\frac{d}{ds} f(\mathbf{x} + s\mathbf{u})_{|\, s=0} = f_{x_1} u_1 + f_{x_2} u_2 = \langle f_x, f_y \rangle \cdot \mathbf{u},$$

proving the claim (1.28) with $\mathbf{V} = \langle f_x, f_y \rangle$. ♢

The left–hand side of (1.28) is the directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{u}$, denoted by $D_{\mathbf{u}} f(\mathbf{x})$. It is the rate of change of f *per unit length* traveled along $\mathbf{u}$, when passing (x, y). The vector $\langle f_x, f_y \rangle$ in (1.28) is called the gradient of f at $\mathbf{x}$, denoted by $\nabla f(\mathbf{x})$, or simply ∇f if $\mathbf{x}$ is clear from context.

Here is a self–contained definition of ∇f, motivated by (1.28).

Definition. Given a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, $\nabla f(\mathbf{x})$ is that vector $\mathbf{V} \in \mathbb{R}^n$ which satisfies

$$D_{\mathbf{u}} f(\mathbf{x}) = \mathbf{V} \cdot \mathbf{u} \quad \text{for all} \quad \mathbf{u} \in \mathbb{R}^n,\ |\mathbf{u}| = 1. \tag{1.29}$$

This definition is sometimes better than the formula $\nabla f = \langle f_x, f_y \rangle$, particularly if we want to keep things in vector form and do not want to stoop to discussing coordinates.

Example: find the gradient of $f(\mathbf{x}) = A\mathbf{x} \cdot \mathbf{x}$, where A is a $n \times n$ matrix (with any integer $n > 0$) with $\mathbf{x} \in \mathbb{R}^n$ and where $\cdot$ denotes the dot product.* **Solution:** we have to identify $\mathbf{V}$ in (1.29); to that end, we substitute the given f, obtaining

$$\frac{d}{ds} A(\mathbf{x} + s\mathbf{u}) \cdot (\mathbf{x} + s\mathbf{u}) \Big|_{s=0} = A\mathbf{x} \cdot \mathbf{u} + A\mathbf{u} \cdot \mathbf{x} = (A + A^T)\mathbf{x} \cdot \mathbf{u}.$$

According to the definition, whatever multiplies $\mathbf{u}$ is ∇f, so that

$$\nabla (A\mathbf{x} \cdot \mathbf{x}) = (A + A^T)\mathbf{x}. \tag{1.30}$$

We managed to avoid mentioning coordinates, and the proof works for any n, not just $n = 2$.

Problem 1.20. Prove the characteristic properties of ∇f:

1. $|\nabla f(x, y)|$ gives the maximal rate of change of f per unit length at (x, y).
2. ∇f points in the direction of the maximal rate of increase of f per unit length at (x, y).

*matrix multiplication is taken before the dot product, i.e. $A\mathbf{x} \cdot \mathbf{x} \equiv (A\mathbf{x}) \cdot \mathbf{x}$.

3. $\nabla f(x, y)$ is perpendicular to the level curve of f passing through (x, y).

Problem 1.21. A given function $f : \mathbb{R}^2 \to \mathbb{R}$ governs the motion of a bug as follows. (i) The direction of the bug's speed aligns with ∇f, and (ii) the magnitude of his speed is the reciprocal of the magnitude $|\nabla f|$. Find $\frac{d}{dt} f(x(t))$, where $x(t)$ is the bug's position at time t.

1.10 Derivative of the map $f : \mathbb{R}^2 \to \mathbb{R}^2$

This short section contains (1) the definition, (2) an explicit expression for the derivative matrix, and (3) a motivation of the definition.

Definition. Given a function $f : \mathbb{R}^2 \to \mathbb{R}^2$, its derivative at $\mathbf{x}_0 \in \mathbb{R}^2$ is the linear map, denoted by $f'(\mathbf{x}_0)$, from $\mathbb{R}^2$ to $\mathbb{R}^2$ which approximates the increment of f in the sense that

$$f(\mathbf{x}) - f(\mathbf{x}_0) = f'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(|\mathbf{x} - \mathbf{x}_0|), \tag{1.31}$$

where $|\mathbf{x}|$ denotes the norm (i.e. the length) of the vector $\mathbf{x}$.

A motivation of this definition is given at the end of the section.

Theorem. Let $f_1,\ f_2$ denote the coordinate functions of $f : \mathbb{R}^2 \to \mathbb{R}^2$:

$$f(\mathbf{x}) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} \quad \text{where} \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.32}$$

If $f_1,\ f_2$ are continuously differentiable at $\mathbf{x} = \mathbf{x}_0$, then $f'(\mathbf{x}_0)$ exists and is given by the matrix

$$f'(\mathbf{x}_0) = \begin{pmatrix} f_{1x} & f_{1y} \\ f_{2x} & f_{2y} \end{pmatrix}_{\mathbf{x}=\mathbf{x}_0} \tag{1.33}$$

whose rows, we note, are the gradients of $f_1,\ f_2$.

Proof. Abbreviating $f_k(\mathbf{x}) - f_k(\mathbf{x}_0) = \Delta f_k$, $k = 1,\ 2$ we have

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \end{pmatrix} \overset{(1.27)}{=} \begin{pmatrix} f_{1x}\Delta x + f_{1y}\Delta y \\ f_{2x}\Delta x + f_{2y}\Delta y \end{pmatrix} + o(|\mathbf{x} - \mathbf{x}_0|),$$

or

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \begin{pmatrix} f_{1x} & f_{1y} \\ f_{2x} & f_{2y} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} + o(|\mathbf{x} - \mathbf{x}_0|). \tag{1.34}$$

Matching this term–by–term with the definition (1.31) proves (1.33). ♢

Motivating the definition. The standard definition of the derivative of $f : \mathbb{R} \to \mathbb{R}$:

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{1.35}$$

is not generalizable for $\mathbf{x} \in \mathbb{R}^n$ with $n > 1$ as it stands, since we cannot divide by a vector. But we avoid the division by rewriting (1.35) in an equivalent form:

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + o(x - x_0), \tag{1.36}$$

where $o(x - x_0)$ is small compared to $x - x_0$ in the sense that

$$\frac{o(x - x_0)}{x - x_0} \to 0 \quad \text{as} \quad x \to x_0.$$

To summarize, we could define $f'(x_0)$ as the number which makes (1.36) true. To see that the new form is indeed equivalent to (1.35), note that the latter amounts to

$$f'(x_0) = \frac{f(x) - f(x_0)}{x - x_0} + O(x - x_0), \tag{1.37}$$

where the quantity $O(x - x_0) \to 0$ as $x \to x_0$, and that in turn is the same as (1.36), as claimed.

A general remark on derivatives and gradient

The gradient ∇f of $f : \mathbb{R}^2 \to \mathbb{R}$ was defined by the same formula as the derivative of a function from $\mathbb{R}^2$ to $\mathbb{R}^2$, namely

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + o(\mathbf{x} - \mathbf{x}_0),$$

and so the gradient is also the derivative, if we extend the concept of derivative to functions between two spaces of different dimensions, as one in fact does; the above definition applies to all dimensions of the domain and the range; it is just the multiplication in $\nabla f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ must be understood in matrix sense. The matrix f' is rectangular, with as many rows as the dimension of the range and with the number of columns given by the dimension of the domain.

Problem 1.22. Show that (1.36) defines $f'(x_0)$ uniquely.

Solution. (1.36) $\Rightarrow$ (1.37) $\Rightarrow$ (1.35).

Problem 1.23. Find the derivative matrix of the following maps, and use this matrix to sketch the image of a small square centered the point at which the derivative is taken and explain what happens to the square (Rotated? Sheared? Dilated? By how much? What is the new area?). Where is the smallness of square is used in getting the answer?

1. $f = (x, y) \mapsto (x^2 - y^2, 2xy)$ at the point $(1, 1)$.
2. $f = (x, y) \mapsto (x - y^2, y)$ at the point $(0, 1)$.
3. $f = (x, y) \mapsto (ax + by, cx + dy)$ at any point.

1.11 Simplest examples.

The following vector fields are building blocks to which any linear vector field reduces, in a certain sense. This reduction is done later.

1. dilation field: $\mathbf{F}_d(\mathbf{x}) = \langle x, y \rangle$.
2. rotation field: $\mathbf{F}_r(\mathbf{x}) = \langle -y, x \rangle$.
3. Combination of the above two fields: $\lambda \mathbf{F}_d(\mathbf{x}) + \omega \mathbf{F}_r(\mathbf{x})$.
4. hyperbolic rotation field: $\mathbf{F}_h = \langle x, -y \rangle$.
5. shear field: $\mathbf{F}_s = \langle y, 0 \rangle$.

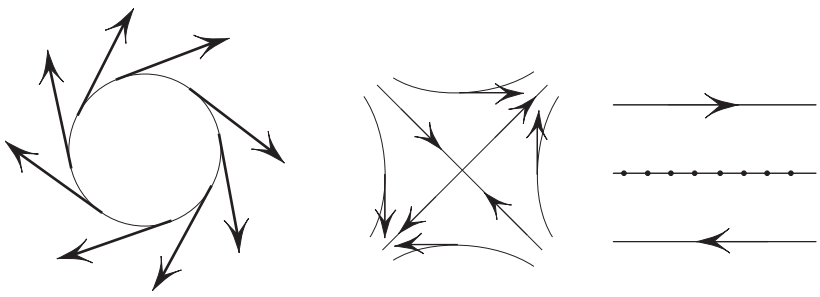

Figure 1.7: Some "building block" examples of vector fields.

Problem 1.24. Consider the vector field $A\mathbf{x}$, where $\mathbf{x} \in \mathbb{R}^2$ and where A is a 2×2 matrix. Describe in words and sketch the vector field in the following two cases: (i) A is a symmetric matrix, with the eigenvalues $\lambda_1,\ \lambda_2$ (consider cases of all sign combinations). (ii) A is skew–symmetric.

1.12 Conservative vector fields; potential

Any vector field, a mathematical object, can be thought intuitively either as a velocity field, or else as a force field. It is worth to keep both interpretations in mind, in this section in particular.

Definition. A vector field $\mathbf{F}$ in $\mathbb{R}^n$ is said to be *conservative* if for any closed contractible curve* γ

$$\oint_\gamma \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} = 0, \tag{1.38}$$

The integral in (1.38) is called the *circulation* around γ; it can be interpreted as the work done by $\mathbf{F}$ around γ.

Gravitational and electrostatic fields are conservative – otherwise the work done by such a field dragging a particle around a closed loop would have been nonzero, resulting in a perpetual motion machine.

Definition. Let $\mathbf{F}$ be a conservative vector field on $\mathbb{R}^n$. Fix a point $\mathbf{x}_0 \in \mathbb{R}^n$. The function

$$P(\mathbf{x}) = -\int_\gamma \mathbf{F}(\mathbf{r}) \cdot d\mathbf{r} \tag{1.39}$$

where γ is a curve connecting $\mathbf{x}_0$ with $\mathbf{x}$, is called a potential of $\mathbf{F}$. (Note the indefinite article: there are infinitely many potentials, but they all differ by a constant.)

Problem 1.25. Given that $\mathbf{F}$ is a conservative vector field, and that the domain of $\mathbf{F}$ is simply connected, show that:

1. $P(\mathbf{x})$ given by (1.39) does not depend on the choice of the curve connecting $\mathbf{x}_0$ and $\mathbf{x}$.
2. replacing a reference point $\mathbf{x}_0$ with a different point $\mathbf{x}_1$ changes the potential P by a constant.
3. $\mathbf{F} = -\nabla P$.

If the domain is not simply connected, then P can be multiple–valued, as Problem 1.26 illustrates.

Problem 1.26. Show that the vector field $\mathbf{F} = \langle -y, x\rangle/(x^2+y^2)$ is conservative, and that its potential P is a multiple–valued function. Can you find P?

Problem 1.27. Show that the vector field $\mathbf{F} = \langle x, y\rangle/(x^2+y^2)$ is conservative and find its potential.

*we call the curve contractible if it can be continuously deformed to a point without leaving the domain of $\mathbf{F}$. An example of a non–contractible curve is a circle centered at the origin and the vector field $\mathbf{F} = \mathbf{r}/r^2$.

Problem 1.28. Consider the constant vector field $\langle \cos\beta, \sin\beta \rangle$. Show that any constant vector field is conservative, i.e. satisfies (1.38). In particular, (1.38) holds for $\mathbf{F} = \langle \cos\beta, \sin\beta \rangle$ where β is some fixed angle, and for γ taken as the triangle with vertices $(0,0)$, $(\cos\alpha, 0)$ and $\langle \cos\alpha, \sin\alpha \rangle$. What familiar identity does (1.38) yield?

A physical interpretation of P. Think of $\mathbf{F}(x)$ as the force (gravitational, say) acting on a point particle. To keep the particle in place, or to move it infinitesimally slowly, I have to apply force $-\mathbf{F}(x)$ (to cancel $\mathbf{F}$). Thus the work I have to do to move along γ from $\mathbf{x}_0$ to $\mathbf{x}$ is precisely $\int_\gamma (-\mathbf{F}) \cdot d\mathbf{r} \stackrel{def}{=} P(\mathbf{x})$.

Theorem 1.3. *Any conservative vector field $\mathbf{F}$ is a gradient of some function, at least locally: for any point in the domain of $\mathbf{F}$ there exists a neighborhood $\mathcal{N}$ with a function $f : \mathcal{N} \to \mathbb{R}$ such that $\mathbf{F} = \nabla f$. Conversely, any gradient field $\mathbf{F} = \nabla f$ is conservative.*

Proof. Let $\mathcal{N}$ be a disk in the domain of $\mathbf{F}$, and define $f : \mathcal{N} \to \mathbb{R}$ by (1.39) where $\mathbf{x}_0$ is a chosen point in $\mathcal{N}$ (the center, say) and where γ is not allowed to leave $\mathcal{N}$. Since $\mathcal{N}$ is simply connected, f is well defined; and $\mathbf{F} = \nabla f$ according the

Since $\mathbf{F}$ is conservative on a simply connected set $\mathcal{N}$, the potential $P(\mathbf{x})$ of $\mathbf{F}$ is well defined on $\mathcal{N}$. And by Problem 1.25,

$$\mathbf{F} = -\nabla P$$

for all $\mathbf{x} \in \mathcal{N}$, so that $\mathbf{F}$ is indeed a gradient vector field. The converse is a consequence of the fundamental theorem of calculus: let $\mathbf{F} = \nabla f$, and let γ be any closed curve in the domain of $\mathbf{F}$ given by $\mathbf{r}(t)$, $t \in [0,1]$, with $\mathbf{r}(0) = \mathbf{r}(1)$:

$$\int_\gamma \mathbf{F} \cdot d\mathbf{r} = \int_0^1 \nabla f \cdot \dot{\mathbf{r}} dt = \int_0^1 \frac{d}{dt} f(\mathbf{r}(t) dt \stackrel{FTC}{=} \mathbf{r}(t) \Big|_0^1 = 0.$$

♢

Problem 1.29. Show that the gravitational field $\mathbf{F}(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|^3$, where $\mathbf{x} \in \mathbb{R}^3$, is conservative.

Problem 1.30. Show that the vector field $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ is conservative if and only if the matrix A is symmetric.

Problem 1.31. Find the potential of the vector field $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$, where A is a symmetric matrix.

1.13 Divergence

The divergence of a vector field quantifies precisely what the word suggests: think of $\mathbf{F}$ as the velocity field (rather than the force field); the divergence at a point is the *rate of change of an infinitesimal area around this point, per unit area*, as the area is carried by the flow $\mathbf{F}$. But since a mathematical expression of the last sentence requires making precise the words "carried by the flow", it is easier to refer to the *"net outflow", i.e. the flux of* $\mathbf{F}$ *through the boundary of the area*, a quantity expressed as a line integral. Here is a precise definition, referring to Figure 1.8.

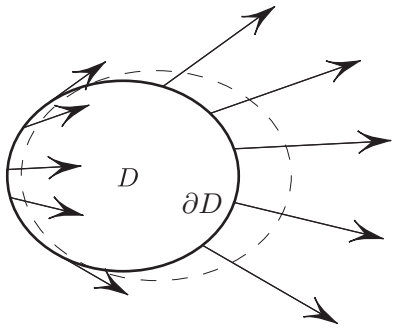

Figure 1.8: Definition of the divergence in $\mathbb{R}^2$. The dashed line is ∂D carried forward for a short time, showing an increased area and suggesting positive divergence in this illustration.

Definition. The divergence of the vector field $\mathbf{F}$ in $\mathbb{R}^2$ is the limit*

$$\mathrm{div}\mathbf{F}(\mathbf{x}) = \lim_{|D|\to 0} \frac{1}{|D|} \oint_{\partial D} \mathbf{F}\cdot\mathbf{N} ds, \tag{1.40}$$

where D denotes a domain bounded by a closed curve ∂D, $|D|$ is the area of D and $\mathbf{N}$ is the outward normal to γ.

For an intuitive feel of (1.40), think of $\mathbf{F}$ as the velocity field, and note that $\mathbf{F}\cdot\mathbf{N}$ is the velocity component normal to γ. Thus $\mathbf{F}\cdot\mathbf{N}\, ds$ is the area swept by the moving arc ds per second. The integral in (1.40) thus expresses the rate of expansion of area of the region carried by the flow.

Alternatively, we can think of the integral in (1.40) as the *flux* of $\mathbf{F}$ through ∂D, i.e. (speaking loosely) as the "area of gas" crossing the *fixed* curve.

*We leave aside the question of existence of this limit and of its independence of the choice of D.

Problem 1.32. For each of the basic vector fields from Section 1.11 on page 26 find the divergence directly from the definition, and compare to the result from the formula.

The following problem relates the concept of flux to the old calculus problem of whether it is better to run or to walk in the rain. The question boils down to computing the flux of a constant vector field through a flat area, the simplest setting involving flux (with an extra twist, however).

Problem 1.33. The rain is coming down vertically. A person wants to carry a sheet of perfectly flat and perfectly thin paper from A to B, holding the sheet vertically and facing it in direction of the walk (without this requirement she could carry it sideways, thus keeping it dry).

Is it better to walk or to run?

Hint: Take the reference frame moving with the sheet; write down the flux.

Problem 1.34. Using the definition, derive the formula $\mathrm{div}\mathbf{F} = P_x + Q_y$, where $P = P(x, y)$ and $Q(x, y)$ are the components of $\mathbf{F}$.

The following problem asks the same question but in polar coordinates.

Problem 1.35. The vector field in the plane is given in polar coordinates as follows: for a particle carried by the flow, its polar coordinates r, θ change at the prescribed rates $R(r, \theta)$, $T(r, \theta)$). Find the expression of the divergence at r, θ using the definition. Verify the formula on the two examples from problems 1.26 and 1.27.

Derivative as divergence in 1D. Let us interpret $f(x)$ as the vector field on $\mathbb{R}$ (one can think of a highway with the car finding itself at x obliged to move precisely with speed $f(x)$. Then $f(x+h) - f(x)$ is the speed at which the space between two cars distance h apart grows, and so

$$\frac{f(x+h) - f(x)}{h}$$

is the rate of growth of this distance per unit distance (essentially, the exponential rate of growth of distance between two cars). Thus f' is exactly the one dimensional divergence.

Interest rate as divergence. Consider a bank account that is compounded continuously at the annual rate r, which by the definition can be written as $\frac{\dot{x}}{x} = r$. Geometrically representing x as an interval $[0, x]$, we see that the rate of its elongation per unit length (which rate is the 1D divergence) is r.

Problem 1.36. Find $\mathrm{div}A\mathbf{x}$, where A is an $n \times n$ matrix. What is the divergence if A is anti–symmetric? What is a geometrical explanation of this in $\mathbb{R}^2$?

Problem 1.37. In this section we have been interpreting $\mathbf{F}$ as the velocity field. Instead, let $\mathbf{F}$ be a gravitational field created by some mass distribution – for example, the gravitational field inside the Earth. Is $\operatorname{div}\mathbf{F} = 0$ inside the Earth? And what is the physical meaning of $\operatorname{div}\mathbf{F}$ in this context? Can you answer the same question if $\mathbf{F}$ is an electrostatic field created by some charge smeared over a region in space?

The gist of the following is very revealing and simple: Newton's law of universal gravitation can be restated in a simpler way: *the divergence of the gravitational field is zero.* This sounds more fundamental than $F = k/r^2$. Besides, the inverse square law only applies to gravitational fields of point masses or spheres, while the zero divergence property is universal. The following problems asks to show that the latter property implies the inverse square law.

Problem 1.38. Show that if a vector field $\mathbf{F}$ in $\mathbb{R}^3$ has $\operatorname{div}\mathbf{F} = 0$, and if it is centrally symmetric then $\mathbf{F}$ satisfies the inverse square law. Hint: the flux through any two concentric spheres (centered at the origin) is the same if $\operatorname{div}\mathbf{F} = 0$. For a further hint, you may want to inspect the (provided) solution of the next problem.

Here is essentially the same problem, in a different dimension and with a different interpretation.

Problem 1.39. An infinitely large lake of small constant depth has a sinkhole into which the water disappears at a constant rate. Find the speed $v(r)$ of the water at the distance r from the sinkhole.

Solution. We assume that the speed depends only on r and not on the depth, and that the water moves radially. Consider the cylinder of radius r with the vertical axis through the sinkhole. The flux f of water through the surface of the cylinder, i.e. the volume of water crossing this surface per second) is independent of r (this is the key point), and equals the area times speed:

$$A(r)v = f = \text{const.},$$

where $A(r) = 2\pi rh$, h being the water depth. To summarize, $2\pi rhv = f$, which shows that v is proportional to r^{-1}:

$$v = k/r, \quad k = f/2\pi h.$$

Remark. Since $\operatorname{div}\mathbf{F}$ at a point depends only on the partial derivatives at that point, we conclude that $\operatorname{div}\mathbf{F}$ at a point depends only on the linearization of $\mathbf{F}$ at that point.

1.14 Curl in 2D

Let us think of a vector field $\mathbf{F}$ in $\mathbb{R}^2$ as the velocity field of imagined gas in the plane. The curl makes precise the vague concept of rotation, as follows.
Definition. (Figure 1.9) Given a vector field $\mathbf{F}$ in $\mathbb{R}^2$, the curl is defined as

$$\operatorname{curl} \mathbf{F}(\mathbf{x}) = \lim_{|D|\to 0} \frac{1}{|D|} \oint_\gamma \mathbf{F}(\mathbf{y}) \cdot d\mathbf{y}, \tag{1.41}$$

where D is a domain containing $\mathbf{x}$ and enclosed by a smooth closed curve γ and $|D|$ denotes the area of D. The integral of tangential velocity is called the *circulation* of $\mathbf{F}$ around γ.

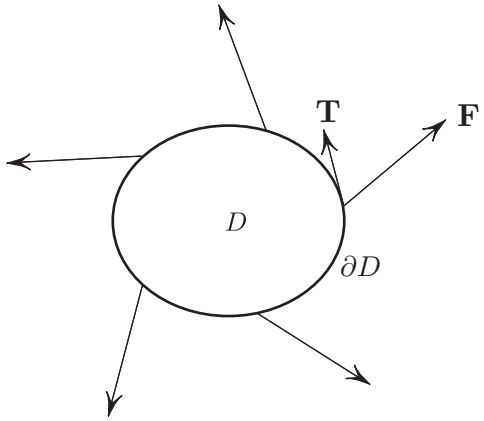

Figure 1.9: Definition of curl in $\mathbb{R}^2$. In this illustration $\mathbf{F} \cdot \mathbf{T} > 0$ on ∂D, suggesting positive curl.

If we think of $\mathbf{F}$ as the force field, rather than a velocity field, then the curl quantifies the non–conservativeness of the field.

A paradox. Interpret the vector field $\mathbf{F}$ from Problem 1.26 as a velocity field. The particles in this field move in circles around the origin. How to reconcile this with the fact that $\operatorname{curl} \mathbf{F} = 0$ which states, loosely speaking, that every small blob of fluid has zero spin?

Problem 1.40. 1. Show that the conservative fields have zero curl.

2. Let $A = (a_{ij})$ be any 2×2 matrix. Show that

$$\operatorname{curl} (A\mathbf{x}) = a_{21} - a_{12}, \quad \text{for any} \quad \mathbf{x} \in \mathbb{R}^2. \tag{1.42}$$

3. What is a necessary and sufficient condition on A for the vector field $A\mathbf{x}$ to be conservative?

Problem 1.41. Using the definition, find the curl of each of the vector fields in Section 1.11, page 26, and verify the result using the formula (1.44).

Problem 1.42. Let $\mathbf{F}$ and $\mathbf{G}$ be two vector fields. Show that curl is a linear operation, i.e. that for any $a,\ b \in \mathbb{R}$ we have

$$\operatorname{curl}(a\,\mathbf{F} + b\,\mathbf{G}) = a\operatorname{curl}\mathbf{F} + b\operatorname{curl}\mathbf{G}. \tag{1.43}$$

Problem 1.43. 1. Show that the curl of a constant vector field is zero.

2. Show that if $\mathbf{r} = \langle p, q\rangle$ satisfies $|\mathbf{r}(\mathbf{x})| \le c(x^2+y^2)$ for some $c > 0$ and for all (x, y) near $\mathbf{0}$, and if p, q have continuous partial derivatives in x, y up to order 2 in a neighborhood of $\mathbf{0}$, then $\operatorname{curl}\mathbf{r}\,(\mathbf{x})_{\mathbf{x}=\mathbf{0}} = 0$.

Theorem 1.4. *Assume that $\mathbf{F} = \langle P, Q\rangle$ is twice continuously differentiable at a point.* Then*

$$\operatorname{curl}\mathbf{F} = Q_x - P_y, \tag{1.44}$$

where the subscripts denote partial derivatives at the point in question.

Problem 1.44. If $\mathbf{F} = \langle P, Q\rangle$ is a velocity field, find an interpretation of Q_x and P_y in terms of certain angular velocities, obtaining thereby an interpretation of $\operatorname{curl}\mathbf{F}$.

Proof of Theorem 1.4. The gist of the proof is to decompose $\mathbf{F}$ into its linear part and the rest, and to show that *only* the linear part contributes to the curl; and for linear fields we already computed the curl in (1.42). Without the loss of generality, we take $\mathbf{x} = \mathbf{0}$. Now the linear part of $\mathbf{F}$ is given by the derivative matrix, according to the definition (1.31) of $\mathbf{F}'$ on page 24:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{0}) + \mathbf{F}'(\mathbf{0})\mathbf{x} + \mathbf{r}(|\mathbf{x}|), \tag{1.45}$$

where the remainder

$$|\mathbf{r}(\mathbf{x})| \le c(x^2+y^2)$$

for some $c > 0$ and for all $\mathbf{x}$ sufficiently close to $\mathbf{0}$, thanks to the assumption of continuity of the second partial derivatives of P, Q.† Taking curl of both sides, we obtain (using the additivity property (1.43)):

$$\operatorname{curl}\mathbf{F}(\mathbf{x}) = \underbrace{\operatorname{curl}\mathbf{F}(\mathbf{0})}_{A} + \underbrace{\operatorname{curl}(\mathbf{F}'(\mathbf{0})\mathbf{x})}_{B} + \underbrace{\operatorname{curl}(\mathbf{r}(\mathbf{x}))}_{C}. \tag{1.46}$$

*in the sense that all partial derivatives of P and Q up to order 2 are continuous.

†Even though $\mathbf{F}$ is thought of as a vector field here, it is still a function $\mathbb{R}^2 \to \mathbb{R}^2$ and it makes sense to speak of its derivative.

Now $A = 0$ and $C = 0$ according to Problem 1.43. And $B = a_{21} - a_{12}$ according to (1.42), where a_{21} and a_{12} are the entries of the matrix $\mathbf{F}'(\mathbf{0})$, and thus are equal to Q_y and P_x respectively, according to the expression for $\mathbf{F}'$, see (1.33). This completes the proof of the expression of the curl (1.44). ♢

The proof of the formula

$$\operatorname{div} \mathbf{F} = P_x + Q_y$$

is an almost verbatim repetition of the proof of (1.44), and is omitted. We only point out that *the divergence is the trace of the derivative matrix*:

$$\operatorname{div} \mathbf{F} = \operatorname{tr} \mathbf{F}'.$$

Remark. In fact, the divergence and the curl simply capture two different features of the derivative matrix $\mathbf{F}'(\mathbf{0})$, i.e. of the linearized vector field $\mathbf{F}'(\mathbf{0})\mathbf{x}$ (we took the point in question to be $\mathbf{0}$, WLOG.)

1.15 Green's and Stokes' theorems in $\mathbb{R}^2$

The theorems, if viewed properly, are scarcely more than restatements of the definitions of div and curl, as the proofs given below explain.

Theorem 1.5. *Given a smooth vector field* $\mathbf{F}$ *and a bounded region* D *enclosed by a piecewise smooth closed curve* ∂D*, one has*

$$\int\int_D \operatorname{div} \mathbf{F}\, dA = \int_{\partial D} \mathbf{F} \cdot \mathbf{N}\, ds, \tag{1.47}$$

in the same notations as used in the definition (1.40) of the divergence, and

$$\int\int_D \operatorname{curl} \mathbf{F}\, dA = \int_{\partial D} \mathbf{F} \cdot d\mathbf{r}. \tag{1.48}$$

(1.47) is called the *divergence theorem*, and it holds for any dimension; (1.48) is the 2D version of Stokes's theorem. There is a screaming similarity of these theorems to the definitions of div and curl, a fact that makes the proofs of both theorems straightforwards.
Here are the two theorems in scalar form, substituting $\mathbf{F} = \langle P, Q\rangle$, $\mathbf{r} = \langle x, y\rangle$ and $\mathbf{N}\, ds = J d\mathbf{r} = \langle -dy, dx\rangle$:

$$\int\int_D (P_x + Q_y)\, dA = \int_{\partial D} -P\, dx + Q\, dy, \tag{1.49}$$

$$\int\int_D (Q_y - P_x)\, dA = \int_{\partial D} P\, dx + Q\, dy. \tag{1.50}$$

Proof. The proofs of (1.47) and (1.48) are identical (or, one can derive one from the other by renaming P and Q), so we concentrate on the proof of the divergence theorem (1.47). Superimposing a square lattice with cell size

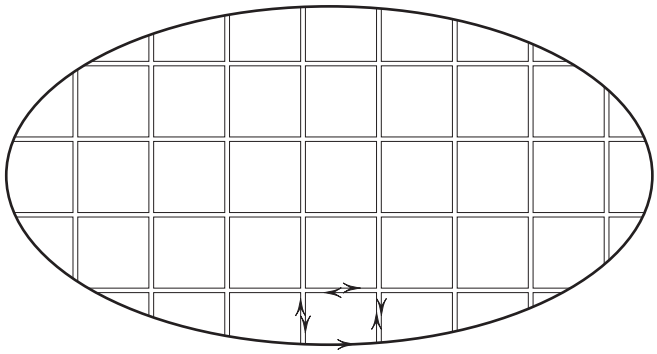

Figure 1.10: Cancellation over shared boundaries in the proof of the divergence theorem and of Stokes' theorem.

$h \times h$ (where h is arbitrary, to be sent to zero later) over the region D, we obtain a partition of D into subregions D_i as shown in the Figure 1.10. For each of these subregions, we have, by the definition of divergence (1.40):

$$\oint_{\partial D_i} \mathbf{F} \cdot \mathbf{N}\, ds = (\operatorname{div} \mathbf{F}(\mathbf{x}_i) + r_i)|D_i|, \tag{1.51}$$

where the remainder r_i is small uniformly in i in the sense that for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|r_i| < \varepsilon$ if $0 < h < \delta$ (the proof of this is omitted). We now sum (1.51) over all i. *The integrals over shared edges cancel* as Figure 1.10 suggests: indeed, the outward normals $\mathbf{N}$ at the shared boundary points of two neighboring subdomains are equal and opposite. Only the integrals over the segments of the outer boundary ∂D remain after summation – this is the main point of the proof. Summarizing (no pun intended), (1.51) results in

$$\oint_{\partial D} \mathbf{F} \cdot \mathbf{N}\, ds = \sum_i \operatorname{div} \mathbf{F}(\mathbf{x}_i)|D_i| + \sum_i r_i|D_i| \tag{1.52}$$

Now the last sum $\to 0$ as $h \to 0$, since for an arbitrary $\varepsilon > 0$ we have $|r_i| < \varepsilon$ for all sufficiently small h, so that

$$\sum_i r_i|D_i| \le \sum_i \varepsilon|D_i| = \varepsilon|D|.$$

And the first sum in (1.52) is a Riemann sum of $\int\int_D \operatorname{div}\mathbf{F}\, dA$. In summary, the limit as $h \to 0$ of (1.52) yields the claim (1.47) of the divergence theorem. $\diamondsuit$

Remark 1.2. *The above proof is the exact analog of the proof of the fundamental theorem of calculus, as summarized in Figure 1.4: there, the sum of the lengths $f(t_{i+1}) - f(t_k)$ telescopes to $f(b) - f(a)$. This telescoping is the exact counterpart of the cancellation of the internal contributions in the present case.*

Problem 1.45. In this problem, a vector $\mathbf{v} \in \mathbb{R}^2$ has been rotated by $\pi/2$ counterclockwise, with the resulting vector denoted by $\mathbf{v}^\perp$.

1. If $\mathbf{v} = \langle x, y\rangle$, what are the coordinates of $\mathbf{v}^\perp$?
2. Find the matrix J that turns (pardon the pun) $\mathbf{v}$ into $\mathbf{v}^\perp$, i.e. such that $\mathbf{v}^\perp = J\mathbf{v}$.

Problem 1.46. Find the matrix $R(\theta)$ that rotates vectors in $\mathbb{R}^2$ through angle θ counterclockwise.

Hint: Write $\mathbf{v} = \langle x, y\rangle$ as a combination of the coordinate vectors $\mathbf{e}_1$, $\mathbf{e}_2$; it is much easier to see what happens to these under rotation.

Problem 1.47. Sketch the following parametric curves (here a, b, λ denote some given constants):

1. $x = a\cos t,\ y = b\sin t$, $0 < a < b$. (Hint: use a linear transformation.)
2. $x = e^{\lambda t}\cos t + e^{\lambda t}\sin t$, where $\lambda < 0$.
3. $x = e^{\lambda t}(2\cos t + \sin t) + e^{\lambda t}(\cos t + \sin t)$, $\lambda < 0$.
4. $x = \cos t,\ y = \cos 2t$.
5. $x = \cos t,\ y = \cos\sqrt{2}t$.
6. How will the curves in the above examples be affected if we replace $\sin t,\ \cos t$ with $\sin 100t,\ \cos 100t$?
7. How does the size and sign of λ affect the curves in examples 2 and 4?

Problem 1.48. Give a complete description of the curve $x = 2\cos t + \sin t,\quad y = \cos t + \sin t$, or more generally, $x = a\cos t + b\sin t,\quad y = c\cos t + d\sin t$

Solution. for $x = 2\cos t + \sin t$, $y = \cos t + \sin t$: in vector form, the curve is

$$\mathbf{r}(t) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix},$$

the image of the unit circle $|\mathbf{u}| = 1$ under the map A. To find the shape of this image, let us write the equation of $\mathbf{r}$ without the parameter t. From $\mathbf{r} = A\mathbf{u}$ we have $\mathbf{u} = A^{-1}\mathbf{r} = B\mathbf{r}$, where $B = A^{-1}$. In short, $|B\mathbf{r}| = 1$ is the desired equation, which is better written without the square root:

$$(B\mathbf{r}, B\mathbf{r}) = 1,$$

or $(B^T B\mathbf{r}, \mathbf{r}) = 1$. The problem has been reduced to finding the shape of the quadratic curve $(S\mathbf{r}, \mathbf{r}) = 1$ for a symmetric matrix S (where $S = (AA^T)^{-1}$. In our case A is symmetric, so that $S = (A^2)^{-1}$. Thus the eigenVECTORS of S and A are the same. And the eigenVALUES of S are λ_k^{-2}, where λ_k are the eigenvalues of A, so that all we need is the eigen information on A. In addition, we see that *the eigenvalues of S are positive.*

The equation $(S\mathbf{r}, \mathbf{r}) = 1$ looks easy in the "right" coordinate system given by the eigenvectors of S (which are orthogonal b/c S is symmetric). We thus decompose $\mathbf{r}$:

$$\mathbf{r} = X_1\mathbf{v}_1 + X_2\mathbf{v}_2,$$

where X_k are just the coordinates in the new coordinate system. Substitute this into $(B\mathbf{r}, B\mathbf{r}) = 1$ and use $(\mathbf{v}_k, \mathbf{v}_k) = 1$, $(\mathbf{v}_1, \mathbf{v}_2) = 0$; the result becomes

$$\lambda_1^{-2} X_1^2 + \lambda_2^{-2} x_2^2 = 1.$$

This is an ellipse with semiaxes $|\lambda_k|$, in the direction of $\mathbf{v}_k$.

Problem 1.49. All we know about a function $f(x, y)$ of two variables is that $f(1, 2) = 3$ and that its partial derivatives $f_x(1, 2) = 4$, $f_y(1, 2) = -3$. Find the approximate value of $f(1.1, 1.9)$, and give an intuitive explanation based on the definition of the derivative.

Problem 1.50. x, y are given functions of time, and f is a given function of x, y. Write $\frac{d}{dt} f(x(t), y(t))$ in terms of the given functions and their derivatives.

Problem 1.51. A constant wind with velocity $\mathbf{v} = \langle v_1, v_2 \rangle$, same at all points, is blowing in the plane. What is the area of the air crossing the segment given by the vector* $\mathbf{g} = \langle g_1, g_2 \rangle$ per unit of time? (This is called the **flux** of $\mathbf{v}$ through the segment given by g, a special case of flux when $\mathbf{v}$ is constant and the "gate" is straight).

*g stands for "gate".

Problem 1.52. Write down the flux of the vector field $\mathbf{F}(x,y) = \langle P(x,y), Q(x,y)\rangle$ through the curve $\mathbf{r}(t) = \rangle x(t), y(t)\rangle$, $t \in [0,1]$ as a line integral, and also as a definite integral. The same question for the work done by the vector field along this curve.

Problem 1.53. Give a geometrical description of the curve in $\mathbb{R}^2$ given by the equation $(S\mathbf{x}, \mathbf{x}) = 1$ where S is a symmetric matrix one of whose eigenvectors is forms angle $\pi/6$ with the x–axis, with the corresponding eigenvalue $\lambda_1 = 1/4$, and the other eigenvalue $\lambda_2 = 1/9$.

Solution. The curve is the ellipse with the semiaxes of lengths 2 and 3 and aligned with the eigenvectors. Proof is in the preceding solution.

Problem 1.54. Define the directional derivative, the gradient, the divergence, and the curl (the latter two only in 2D).

1.16 Matrices viewed geometrically.

Square $n \times n$ matrix A often arise in two settings: (i) linear vector fields, given by $\mathbf{F}(\mathbf{x}) \mapsto A\mathbf{x}$ and (ii) linear transformations $T : \mathbf{x} \mapsto A\mathbf{x}$.

Linear vector fields. The following simple examples are "building blocks" for understanding the picture of any linear vector field in $\mathbb{R}^2$.

1. Velocity field $A\mathbf{x}$ of rigid rotation: $A = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix}$
2. Velocity field $A\mathbf{x}$ of expansion at the exponential rate α: $A = \operatorname{diag}(\alpha, \alpha)$
3. Hyperbolic flow $A\mathbf{x}$, with $A = \operatorname{diag}(\alpha, -\alpha)$
4. Shear flow $A\mathbf{x}$, with $A = \begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$

Problem 1.55. For each of the above examples, sketch the vector field and some representative trajectories (of particles moving with this vector field).

Solution. The sketch is given in Figure 1.7.

Linear transformations. Any linear transformation is a composition of the basic transformations $\mathbf{x} \mapsto A\mathbf{x}$ with the following matrices A.

1. $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, rotation through angle θ.
2. $A = \operatorname{diag}(\lambda, -\lambda)$, hyperbolic rotation.
3. $A = \lambda I = \lambda \operatorname{diag}(1, 1)$, dilation by factor λ.
4. $A = \operatorname{diag}(1, 0)$, projection onto the first coordinate axis.
5. $A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, shear along the first coordinate axis.

Problem 1.56. Show that the matrix in the first example is indeed the rotation, by the following method: first, find the rotated basis vectors $\mathbf{e}_k$, $k = 1, 2$; this will give the result of rotating any vector $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$; from this one can read off the matrix.

Problem 1.57. Prove that the rotation matrix indeed has the form given above by using a general observation: *the columns of any matrix A are the vectors $A\mathbf{e}_k$,* where $\mathbf{e}_k$ are the unit vectors of coordinate axes.

Problem 1.58. Draw the image of the unit square $0 \le x \le 1$, $0 \le y \le 1$ under each of the above transformations, and compare the area of the image in each case with $\det A$.

Solution. SeeFigure 1.11.

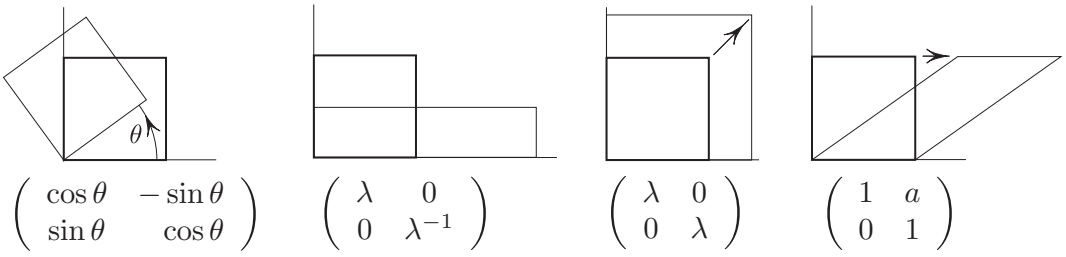

Figure 1.11: Basic linear transformations: rotation, hyperbolic rotation, dilation, shear.

Problem 1.59. Show that the image of the unit circle under any linear map is an ellipse and find the lengths and the directions of the axes in terms of the matrix of the transformation. Hint: See a solution to Problem 1.47.

1.17 The determinant as the n–volume.

The determinant of a matrix is usually defined algebraically. The algebraic definition, which may seem to be pulled out of a hat, is equivalent to a natural geometrical definition, namely the following:

the determinant of a $n \times n$ matrix is the signed volume of the parallelepiped formed by the n column vectors of A. For $n = 2$ "volume" means area, "parallelepiped" means the parallelogram, and the area comes with a sign which depends on whether the two column vectors form the right–handed frame or the left–handed one.

Permuting two columns in a determinant causes the change of handedness, and thus of sign. And if two columns are identical, their swapping does not change anything on the one hand, but the sign on the other, which means that the determinant with two equal columns is zero.

Problem 1.60. Let S$(\mathbf{a}, \mathbf{b})$ denote the (oriented) area of the parallelogram formed by vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^2$. Show *geometrically* that S is a bilinear anti–commutative function of 2 vectors, taking value 1 on the pair of unit coordinate vectors $\mathbf{e}_1$, $\mathbf{e}_2$. In other words, show that for any vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ the following hold:

1. $S(\mathbf{e}_1, \mathbf{e}_2) = 1$, where $\mathbf{e}_1$, $\mathbf{e}_2$ are the unit vectors of the coordinate axes.
2. $S(\mathbf{a}, \mathbf{b}) = -S(\mathbf{b}, \mathbf{a})$.
3. $S(k\mathbf{a}, \mathbf{b}) = kS(\mathbf{a}, \mathbf{b})$ for any real k,
4. $S(\mathbf{a} + \mathbf{b}, \mathbf{c}) = S(\mathbf{a}, \mathbf{c}) + S(\mathbf{b}, \mathbf{c})$.

Instead of defining the determinant using the "base times height" idea that works for $\mathbb{R}^2$ and $\mathbb{R}^3$, we simply define the determinant as a function with the characteristic properties of the volume.*

Defnition. The determinant of order n is the real–valued function defined on the set of $n \times n$ matrices, and satisfying the following properties:

1. $\det I = 1$, where I is the identity matrix.
2. $\det(a\mathbf{v}_1 + b\mathbf{v}_1', \mathbf{v}_2, \ldots) = a \det(a\mathbf{v}_1, \mathbf{v}_2, \ldots) + b \det(\mathbf{v}_1', \mathbf{v}_2, \ldots)$,
3. $\det(\ldots \mathbf{v}_i \ldots \mathbf{v}_j \ldots) = -\det(\ldots \mathbf{v}_j \ldots \mathbf{v}_i \ldots)$.

*One then must show that such function exists, and that it is unique.

To see that the a function det with properties $1-3$ indeed exists and is unique one simply has to expand each $\mathbf{v}_k$ in the coordinate basis and apply the rules $1-3$ to obtain the explicit formula (that is usually is given as the definition).

Problem 1.61. Consider a time–dependent matrix $A(t)$, and let $\mathbf{v}_k = \mathbf{v}_k(t) \in \mathbb{R}^n$ be the column vectors. Show that

$$\frac{d}{dt}\det A(t) = \det(\dot{\mathbf{v}}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) + \ldots + \det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \dot{\mathbf{v}}_n), \tag{1.53}$$

and give a geometrical explanation of the answer in $\mathbb{R}^3$.

Solution.

The proof of (1.53) is a copy of the proof of the usual calculus product rule

$$\frac{d}{dt}(v_1 v_2 \ldots v_n) = \dot{v}_1 v_2 \ldots v_n + \ldots + v_1 v_1 \ldots \dot{v}_n$$

for scalar functions $u_k = u_k(t)$, $k = 1, 2, \ldots, n$. The reason the same proof works is that the determinant is a kind of a product of n vectors $\mathbf{v}_k$, and that this product satisfies the distributive property 2 listed in the definition of the determinant. To proceed, let's abbreviate

$$\det(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = |\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n|.$$

For brevity's sake, write $\mathbf{v}_k(t) = \mathbf{v}_k$ and $\mathbf{v}_k(t+h) = \mathbf{v}_k^+$, and consider the increment due to change of t by h:

$$\Delta V \stackrel{def}{=} \det A(t+h) - \det A(t) = |\mathbf{v}_1^+, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| - |\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n|.$$

The trouble with this difference is that all the columns were changed simultaneously; we can make this change one-by-one, namely, by adding and subtracting a determinant differing from the one before only in one column:

$$\begin{aligned} &\Delta V = \\ &|\mathbf{v}_1^+, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| - |\mathbf{v}_1, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| + \\ &|\mathbf{v}_1, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| - |\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n^+| + \ldots + \\ &|\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n^+| - |\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n|. \end{aligned} \tag{1.54}$$

Using Property 2 of the determinant we combine each difference into one determinant, writing the above as

$$|\Delta\mathbf{v}_1, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| + \ldots + |\mathbf{v}_1, \mathbf{v}_2, \ldots, \Delta\mathbf{v}_n|, \tag{1.55}$$

where $\Delta\mathbf{v}_k = \mathbf{v}_k^+ - \mathbf{v}_k$. Thus

$$\frac{\Delta V}{h} = |h^{-1}\Delta\mathbf{v}_1, \mathbf{v}_2^+, \ldots, \mathbf{v}_n^+| + \ldots + |\mathbf{v}_1, \mathbf{v}_2, \ldots, h^{-1}\Delta\mathbf{v}_n|,$$

again using Property 2 to "import" $1/h$. But $h^{-1}\Delta\mathbf{v}_k \to \dot{\mathbf{v}}_k$ and $\mathbf{v}_k^+ \to \mathbf{v}_k$ $(k = 1, \ldots, n)$ as $h \to 0$; this proves Eq. (1.53).

1.18 The determinant as the volume stretch

In the previous section we interpreted the determinant of a matrix as the (signed) volume of the parallelepiped built on the column–vectors.

Here is an equivalent interpretation of the determinant: Let $\mathrm{Vol}(S)$ denote the n–volume of a set $S \in \mathbb{R}^n$.* Then

$$\det A = \frac{\mathrm{Vol}(A(S))}{\mathrm{Vol}(S)}, \tag{1.56}$$

Figure 1.12 in other words, $\det A$ is the factor by which the transformation A changes the volume.

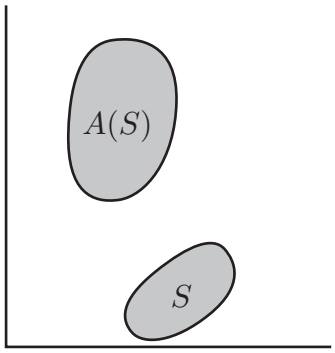

Figure 1.12: Determinant as the area stretching factor.

Proof. 1. Take S to be the unit cube built on the unit coordinate vectors $\mathbf{e}_i$. The parallelepiped $A(S)$ is built on vectors of $A\mathbf{e}_i$, so that $\mathrm{Vol}(A(S)) = \det(A\mathbf{e}_1, \ldots, A\mathbf{e}_n)$ (by the definition of the determinant as the volume). But $A\mathbf{e}_i = \mathrm{col}_i(A)$, so that $\mathrm{Vol}(A(S)) = \det A$. Substituting this into (1.56) and using $\mathrm{Vol}(S) = 1$ we see that (1.56) indeed holds true, at least in case when S is a unit cube.

*For the volume to be well defined, S has to be a measurable set; it suffices to think of a set bounded by a smooth or piecewise smooth surface, or even just a cube.

2. Let S be a set whose volume can be arbitrarily well approximated by the sum of volumes of cubes. (1.56) which we can write as

$$\text{Vol}(A(S)) = \det A \text{Vol}(S)$$

holds for an arbitrary dilation and translation of the unit cube, and therefore Finally, any measurable set can be approximated (as far as the volume goes) by a disjoint union of cubes, and by additivity of volume (1.56) holds for all such sets S. ♢

Problem 1.62. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of eigenvectors of matrix A and let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues. Use (1.56) to explain why $\det A = \lambda_1 \ldots \lambda_n$. Hint: pick the set S to be the parallelepiped built on the eigenvectors.

The volume–stretching interpretation (1.56) of $\det A$ gives an extremely simple explanation of the following property of the determinant.

Theorem 1.6. *For any $n \times n$ matrices A, B one has*

$$\det (AB) = \det A \det B. \tag{1.57}$$

Proof. Applying the transformation AB to a set S of unit volume in two steps: first applying B, then applying A, we multiply $\text{Vol}\ (S) = 1$ first by $\det B$, and then by $\det A$, with the resulting volume $= \det A \det B$. Summarizing, we showed that

$$\frac{\text{Vol}\ ((AB)(S))}{\text{Vol}\ S} = \det A \det B.$$

But the left–hand side is simply $\det AB$ according to (1.56). ♢

Note that the geometrical idea behind (1.57) is exactly the same as behind the chain rule, page 9.

1.19 Eigenvalues and eigenvectors – two geometrical interpretations.

Recall that a nonzero vector $\mathbf{v}$ is said to be an *eigenvector* of a matrix A, and a number λ an eigenvalue of A if

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{1.58}$$

The components of $\mathbf{v}$, as well as λ, may be complex numbers. The geometrical meaning of this will be explained separately; here we concentrate on the real case.

1. If a matrix A is treated as a transformation $\mathbf{x} \mapsto A\mathbf{x}$, then (1.58) says that A stretches the eigenvector $\mathbf{v}$ by the factor λ (the eigenvalue). Any other vector $a\mathbf{v}$ parallel to $\mathbf{v}$ undergoes the same stretching. Now *any* vector $\mathbf{x}$ is a combination $\mathbf{x} = X_1\mathbf{v}_1 + X_2\mathbf{v}_2$, and thus $A\mathbf{x} = \lambda_1 X_1\mathbf{v}_1 + \lambda_2 X_2\mathbf{v}_2$. In other words, the coordinates X_1, X_2 of $\mathbf{x}$ in the basis $\mathbf{v}_1$, $\mathbf{v}_2$ undergo stretching by factors λ_1, λ_2.
2. If a matrix A is treated as a vector field $\mathbf{x} \mapsto A\mathbf{x}$, then (1.58) says that the velocity at $\mathbf{v}$ is aligned with the direction $\mathbf{v}$; Figure 1.13 illustrates this alignment.

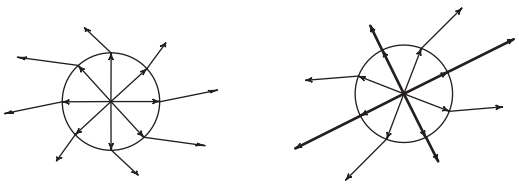

Figure 1.13: Eigenvectors illustrated.

Problem 1.63. Show that if $\mathbf{v}$ is an eigenvector of A, then so is $a\mathbf{v}$ for any nonzero $a \in \mathbb{R}$.

Solving linear ODEs $\dot{\mathbf{x}} = A\mathbf{x}$ Consider the vector field $A\mathbf{x}$, and the particles starting on the "eigenline" (the line of an eigenvector)* of A. These particles stay on that line, since their velocity $\dot{x} = A\mathbf{x} = \lambda\mathbf{x}$ is aligned with their position vector. In other words, their

$$\text{velocity} = \lambda \cdot \text{position},$$

a trademark of exponential motion, either away or towards the origin depending on the sign of λ. This explains why $e^{\lambda t}\mathbf{v}$ is a solution of the system of ODEs $\dot{\mathbf{x}} = A\mathbf{x}$. This also suggests that the system $\dot{\mathbf{x}} = A\mathbf{x}$ with real eigenvectors/eigenvalues is equivalent to a decoupled combination of one–dimensional ODEs of the form $\dot{x} = \lambda x$ (as shown in a later chapter).

1.20 Symmetric matrices.

A square matrix $A = (a_{ij})$ is said to be *symmetric* if

$$a_{ij} = a_{ji}. \tag{1.59}$$

*assuming it is real

Here are a few condensed insights into the meaning of the symmetry of a matrix. A is symmetric if an only if any of the following hold.

1. The *force field* $A\mathbf{x}$ is conservative.
2. The *velocity field* $A\mathbf{x}$ has zero curl for $n = 3$ and zero 2D curl for $n = 2$.
3. A has an orthogonal basis of (real) eigenvectors.

Here are five additional isights:

1. The elegant but also a bit antiseptic definition of a symmetric $n \times n$ real matrix A as the one satisfying the identity

$$(A\mathbf{x}) \cdot \mathbf{y} - \mathbf{x} \cdot (A\mathbf{y}) = 0 \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{1.60}$$

has a physical interpretation: the left–hand side is the work done by the linear force field $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ around the parallelogram generated by $\mathbf{x}$ and $\mathbf{y}$.

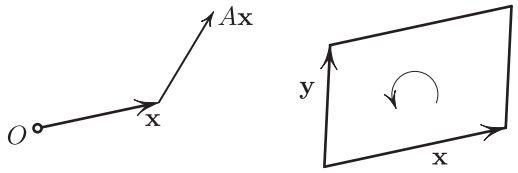

Figure 1.14: Vector field $A\mathbf{x}$ and the closed parallelogram path.

Figure illustrates the proof of this equivalence: the average forces on each of the sides of the parallelogram are equal to the forces $\mathbf{F}_i$ at the midpoints M_i; the total work W around the parallelogram, grouping parallel sides together, is

$$W = (\mathbf{F}_1 - \mathbf{F}_3) \cdot \mathbf{x} + (\mathbf{F}_2 - \mathbf{F}_4) \cdot \mathbf{y};$$

and since $\mathbf{F}_1 - \mathbf{F}_3 = -A\mathbf{y}$ and $\mathbf{F}_2 - \mathbf{F}_4 = A\mathbf{x}$, this gives

$$W = (A\mathbf{x}) \cdot \mathbf{y} - \mathbf{x} \cdot (A\mathbf{y}).$$

In particular, (1.60) expresses the conservativeness of the vector field $A\mathbf{x}$.

2. Here is a physical reason why eigenvalues of a symmetric matrix are real. Assuming for a moment that they are not, consider the plane spanned by

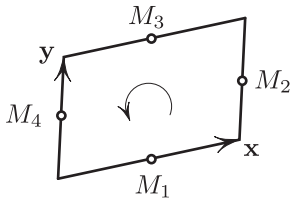

Figure 1.15: .

the real and the imaginary parts $\mathbf{u}$, $\mathbf{v}$ of the eigenvector $\mathbf{w} = \mathbf{u} + i\mathbf{v}$. At each point $\mathbf{x}$ in this plane the force $A\mathbf{x}$ lies in the plane (so that we can forget about the rest of $\mathbb{R}^n$). And since the work done by $A\mathbf{x}$ around a circle in this plane centered around the origin is zero, the tangential component of $A\mathbf{x}$ changes sign at some point(s) $\mathbf{x}_0$ on the circle – which is to say, $A\mathbf{x}_0$ is normal to the circle at $\mathbf{x}_0$, i.e. $\mathbf{x}_0$ is a (real) eigenvector.

3. Orthogonality of eigendirections of a real symmetric matrix can be seen by a "physical/geometrical" argument, where one can "feel" every step, not hidden by algebra. Let $\mathbf{u}$, $\mathbf{v}$ be two distinct eigenvectors of a symmetric $n \times n$ matrix A with the eigenvalues $\lambda \neq \mu = 0$ (the latter assumption involves no loss of generality since we can take $\mu = 0$ by replacing A with $A - \mu I$). Figure 1.16 shows the force field $A\mathbf{x}$ of such a matrix. Consider the work of $A\mathbf{x}$ around the triangle OQP. The only contribution comes from PO since $A\mathbf{x}$ vanishes along OQ and is normal to QP. And if $A\mathbf{x}$ is conservative, then $W_{PO} = 0$ and hence $P = O$, implying $\mathbf{u} \perp \mathbf{v}$. This completes a "physical" proof of orthogonality of the eigenvectors of symmetric matrices.

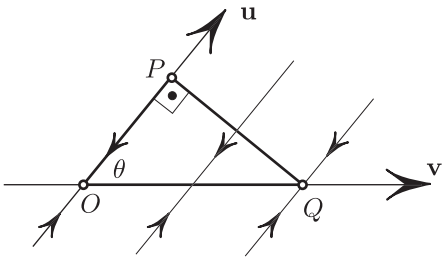

Figure 1.16: Orthogonality of the eigenvectors proven geometrically.

4. The off–diagonal entries a_{ij} of a square matrix $A = (a_{ij})$, interpreted dynamically, are the angular velocities, in the (ij)–plane, of $\mathbf{e}_i$ moving with

the vector field $A\mathbf{x}$.* Indeed, $a_{ij} = (A\mathbf{e}_i, \mathbf{e}_j)$ is the projection of the velocity $A\mathbf{e}_i$ onto $\mathbf{e}_j$, Figure 1.17, and thus measures "how fast $\mathbf{e}_i$ rotates" (see the last footnote for the explanation of quotes).

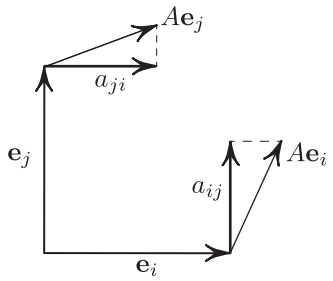

Figure 1.17: a_{ij} as the angular velocity.

And thus the symmetry condition $a_{ij} = a_{ji}$ illustrated in Figure amounts also to stating that the 2D curl in every ij–plane vanishes. For 3×3 matrices the symmetry is equivalent to $\operatorname{curl} A\mathbf{x} = \mathbf{0}$. In fact, decomposition of a general square matrix into its symmetric and antisymmetric parts amounts to decomposing the vector field $A\mathbf{x}$ into the sum of a curl free and of a divergence free fields, a special case of Helmholtz's theorem, itself a special case of the Hodge decomposition theorem.

5. The diagonal entries a_{ii} give the rate of elongation of $\mathbf{e}_i$; this explains geometrically why the cube formed by $\mathbf{e}_i$ at $t = 0$ and carried by the velocity field $A\mathbf{x}$ changes its volume at the rate $\operatorname{tr} A$ (at $t = 0$): at $t = 0$ the face perpendicular to side $\mathbf{e}_i$ moves with speed a_{ii}, causing the volume to grow at the same rate; with n contributions, the volume grows at the rate $\operatorname{tr} A$. This also gives (modulo some skipped details) a geometrical explanation of the matrix identity $\det e^A = e^{\operatorname{tr} A}$.

1.21 Geometrical meaning of complex eigenvalues and eigenvectors

Let A be a real 2×2 matrix, and let $\lambda = \alpha + i\omega$ be its complex eigenvalue, corresponding to an eigenvector $\mathbf{v}$, which must also therefore be complex: $\mathbf{v} = \mathbf{u} + i\mathbf{w}$, with $\mathbf{u}, \mathbf{w} \in \mathbb{R}^2$.

*to be more precise, the moving vector in question equals $\mathbf{e}_i$ at the instant in question only, since it moves.

To unearth the geometrical meaning of $A\mathbf{v} = \lambda\mathbf{v}$, we separate the real and the imaginary parts, obtaining

$$\begin{cases} A\mathbf{u} = \alpha\mathbf{u} - \omega\mathbf{w} \\ A\mathbf{w} = \omega\mathbf{u} + \alpha\mathbf{w}. \end{cases} \tag{1.61}$$

This can be written as a single matrix equation by introducing the matrix $T = (\mathbf{u}, \mathbf{w})$ with column–vectors $\mathbf{u}$, $\mathbf{w}$ and the matrix $\Lambda = \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}$ simply as

$$AT = T\Lambda, \quad \text{or} \quad A = T\Lambda T^{-1}. \tag{1.62}$$

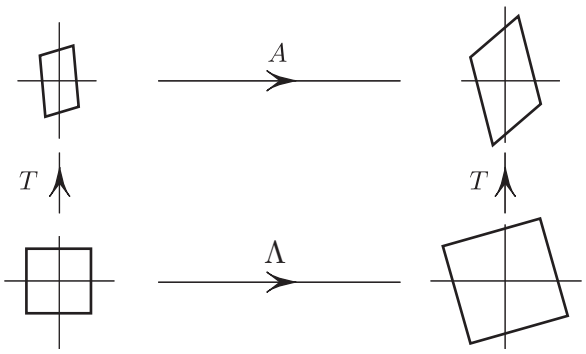

Figure 1.18: Geometry of $A = T^{-1}\Lambda T$: any 2×2 linear transformation with complex eigenvalues, "viewed through the lens" T, is a rotation with dilation

This means that A is conjugate to the action of the matrix Λ, i.e. to the composition of rotation through the angle $\arg(\alpha + i\omega)$ and dilation by the factor $\sqrt{\alpha^2 + \omega^2}$. In other words, the matrix A, when viewed through the linear "lens" T, is a composition of rotation and dilation.*

Alternatively, if $A\mathbf{x}$ is interpreted as a velocity field, then this field, viewed through the "lens" T, is a sum of rotational field with angular velocity ω and of the exponential spreading from the origin with the coefficient α.

*If $\det T < 0$, i.e. if the transformation T is orientation–reversing, then the rotations viewed through the "lens" T will appear to have the direction opposite to their true one. To avoid this, we note that the conjugate of λ, $\mathbf{v}$ is also an eigenvalue–eigenvector pair, and that for one of these pairs $\det T = \det(\mathbf{u}, \mathbf{w}) > 0$. And should have choosen that pair to start with.

Problem 1.64. Let $\Lambda = \begin{pmatrix} \alpha & -\omega \\ \omega & \alpha \end{pmatrix}$.

1. Find the curves in the plane which are invariant under the transformation $\mathbf{x} \mapsto \Lambda\mathbf{x}$.

2. Find the curves in the plane to which the vector field $\Lambda\mathbf{x}$ is tangent at every point of the curve.

1.22 Complex numbers

Definition. A *complex number* $z = x + iy$ is the point (x, y) in the plane, with the additional understanding that the points can be added by the parallelogram rule and multiplied by the following geometrical rule: in multiplying two complex numbers, the lengths multiply*, and the angles add.

This multiplication rule could have been discovered by observing that the familiar rules $1 \cdot 1 = 1,\ (-1) \cdot 1 = 1 \cdot (-1) = -1,\ (-1)(-1) = 1$ can be explained by one unifying principle of angle–addition. For the example $(-1)(-1) = 1$, since the angle addition rule gives the angle $\pi + \pi = 2\pi$ for the product; the product is therefore aligned the positive x–axis, and we get $+1$.

Incidentally, the angle–addition rule can also lead to the "discovery" of $i = (0, 1)$: if the product $i \cdot i = -1$ forms angle π with the positive x–axis, then i must form the angle $\pi/2$,† which leads to $i = (0, 1)$. Usually, the product $z_1 z_2$ of complex numbers is defined by the formula (amounting to the assumption $i^2 = -1$ and the distributive and commutative properties):

$$z_1 z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1), \tag{1.63}$$

and the angle–addition form is proven afterwards. Here we went the other way, giving geometry the upper hand.

Problem 1.65. Prove (1.63) using the geometrical definition of multiplication and addition.

---

*The length $|z|$, or the *absolute value* of z, is the distance to the origin, i.e. $|z| \stackrel{def}{=} \sqrt{x^2 + y^2}$.

†or $-\pi/2$, but let us discard that case; it was an accident of history that of the two possibilities ($(0, 1)$ or $(0, -1)$) the former was chosen as i.

1.23 Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$ – an intuitive derivation.

What is a reasonable definition of e^{it}? Since we defined e^t as the solution of the ODE $\dot{x} = x$ with $x(0) = 1$, let us similarly define e^{it} as the solution of

$$\dot{z} = iz, \quad z(0) = 1. \tag{1.64}$$

Here is a quick way to explain the famous Euler's formula

$$e^{it} = \cos t + i\sin t, \tag{1.65}$$

i.e. to see that the solution of (1.64) is given by $z = \cos t + i\sin t$. Indeed,

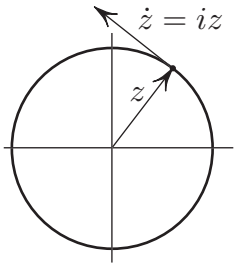

Figure 1.19: A geometrical explanation of Euler's formula.

1. $\dot{z} = iz \perp z$, according to (1.64): the velocity is perpendicular to the position vector, imply that z executes a circular motion.

2. since $z(0) = 1$, the circle in question has radius 1; in other words, $|z| = 1$ for all t.

3. and the speed $|\dot{z}| = |iz| = |z| = 1$. So in time t the point z travels the distance t along the circle from its starting point, which is 1, the point on the x–axis. This amounts to $z = (\cos t, \sin t)$, by the definition of cosine and sine,* or equivalently, $z = \cos t + i\sin t$, as we claimed (1.64) implies.

Remarkably, the complex equation (1.64) is easier to solve than the real one $\dot{x} = x$!

*Recall that $\cos t$ and $\sin t$ are defined as the coordinates of the point of the unit circle lying at the distance t from $(1, 0)$ (measured clockwise or counterclockwise according to the sign of t.)

Chapter 2

An overview of ODEs.

Ordinary differential equations describe a striking variety of things, both moving and stationary, including

1. vibration of interconnected mass–spring systems, oscillations of pendulums, of ships and bridges
2. the motion of projectiles, planets, comets, asteroids and artificial satellites
3. the tumbling of gymnasts
4. the motion of spinning tops, of rolling coins, etc.
5. shapes of hanging chains, cables, of sagging beams.
6. dynamics of chemical reactions and of biological processes
7. population dynamics
8. spread of infectious diseases
9. change of currents and voltages in electric circuits
10. dynamics of airplanes
11. motion of charged particles in electomagnetic fields
12. climate models, etc, etc.

In fact, the entire Newtonian mechanics is really a branch of the theory of ODEs, since Newton's second law usually amounts to an ODE.

Given this wide variety of applications, it is remarkable that *all* ODEs are equivalent to an object of one single kind: *a vector field,* hbv although often in space of dimension other than 2 or 3, as will be explained shortly. And solving an ODE amounts, as we shall see, to finding paths of particles carried by a vector field.

The above motivates our nearest plan, which is to (i) practice formulating some "real life" problems as ODEs (without this skill we would be learning about the hammer without ever hitting the nails) – this is relegated to the Problems section; (ii) explaining how any ODE reduces to studying paths of particles in vector fields, and (iii) giving a bird's eye view of main classes of ODEs.

2.1 Definition and reduction to vector fields

Definition 2.1. *An ordinary differential equation for the unknown function $x = x(t)$ of the independent variable t is the relationship*

$$F(t, x, \dot{x}, \ddot{x}, \dots x^{(n)}) = 0, \tag{2.1}$$

where F is a given function of $n + 1$ variables. One says that (2.1) is an ODE of order n, according to the order of the highest derivative.

Since the highest derivative is the "most important" one, it is customary to express it in terms of the rest:

$$x^{(n)} = f(t, x, \dots x^{(n-1)}). \tag{2.2}$$

As a side remark, for $x^{(n)}$ to be thus expressible, it suffices to require that F be "sensitive" to its last argument:

$$F(t, x_0, \dots x_0^{(n)}) = 0, \text{ and } \partial_{n+1} F(t, x_0, \dots x_0^{(n)}) \neq 0, \tag{2.3}$$

the condition of the implicit function theorem. Here ∂_{n+1} stands for the partial derivative with respect to the last argument of F.

Reducing any ODE to a vector field

The ODE (2.2) can be rewritten as a first order ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$ in $\mathbf{x} \in \mathbb{R}^n$, n being the order of the highest derivative. This is done as follows. Introduce $x_1 = x$, $x_2 = \dot{x}_1$, $x_3 = \dot{x}_2, \dots, x_n = \dot{x}_{n-1}$; in other words, we

treat higher derivatives as new coordinates in a higher dimensional space. Now $\dot{x}_n = x_1^{(n)} \overset{(2.2)}{=} f(t, x_1, \dots, x_{n-1})$. In summary, Eq. becomes

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dots \\ \dot{x}_n = f(t, x_1,\ x_2, \dots x_{n-1}), \end{cases} \tag{2.4}$$

or

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}),, \tag{2.5}$$

where $\mathbf{x} = (x_1, \dots x_n)$, $\mathbf{f}(t, \mathbf{x}) = (x_2, \dots, x_{n-2}, f(t, x_2, \dots, x_{n-1}))$.

Autonomous vs. non–autonomous ODEs

Definition. The ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ with the vector field $\mathbf{f}$ independent of time is referred to as *autonomous*.

The term "autonomous" suggests the absence of external influence upon the system. For example, the ODE describing the exponential grows $\dot{x} = ax$ is autonomous, but if the "interest rate" a depends on time: $a = a(t)$, then the equation is non–autonomous.

The non-autonomous ODE Eq. (2.5) is equivalent to an autonomous ODE in dimension $n+1$. Indeed, introduce the new dimension $x_0 = t$; then $\dot{x}_0 = 1$, combined with $\dot{\mathbf{x}} = \mathbf{f}(x_0, \mathbf{x})$ can be written as a single ODE for the vector $\mathbf{X} = (x_0, \mathbf{x}) \in \mathbb{R}^{n+1}$:

$$\dot{\mathbf{X}} = \mathbf{F}(\mathbf{X}), \text{ where } \mathbf{F} = (1, \mathbf{f}). \tag{2.6}$$

We rewrote Eq. (2.5) as an autonomous system at the cost or raising the number of dependent variables to $n+1$. Note that the velocity field in the x_0–direction has velocity 1.

The imaginary gas interpretation of ODEs. As mentioned before, the vector field $\mathbf{f}$ can be interpreted as velocity field of an imagined gas in $\mathbb{R}^n$; a gas particle carried by this flow then satisfies our ODE $\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x})$. To solve the ODE means to determine the motion $\mathbf{x}(t)$ (as a function of time and of initial position). It is remarkable that this gas interpretation works for any ODE, whatever it's origin: an RLC circuit, a pendulum, an ecosystem with competing species, a planetary motion, etc.

Definitions.

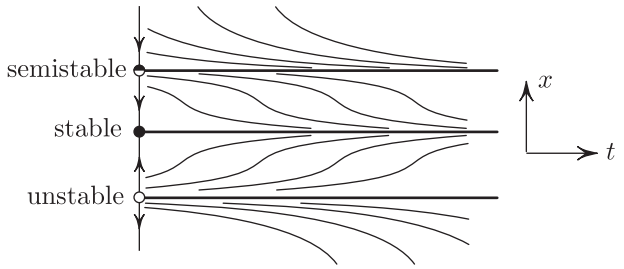

Figure 2.1: Phase space $\mathbb{R}^2$ versus the extended phase space $\mathbb{R}^3$ for the harmonic oscillator $\ddot{x} + x = 0$.

1. A *solution* of the differential equation is any function $\mathbf{x} = \mathbf{x}(t)$ which satisfies the equation.

2. The space $\{\mathbf{x}\}$ is called the *phase space* of the system.

3. The space $\{(t, \mathbf{x})\}$ is called an *extended phase space.*

4. The curve $\mathbf{x}(t)$ (where $\mathbf{x}$ is a solution of the ODE) is called the *trajectory.* The trajectory is thus a projection of the solution curve $\langle t, \mathbf{x}(t)\rangle$ from the extended phase space onto the phase space; Figure shows an example.

Problem 2.1. Sketch the vector field of the ODE $\dot{x} = x(x-1)(x-2)$ in the extended phase space and sketch the trajectories in the extended phase space.

Problem 2.2. Write the ODE $\ddot{x} = x$ as a first order system, and sketch trajectories in $\mathbb{R}^2$ and the solution curves in the extended phase space $\mathbb{R}^3$.

2.2 The flow of an autonomous ODE

Historically, the most basic problem in the theory of ODEs is the initial value problem (IVP):

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}) \\ \mathbf{x}(t_0) &= \mathbf{x}_0, \end{aligned} \tag{2.7}$$

also called the *Cauchy problem.* The notation $\mathbf{x}(t)$ for the solution is inadequate since it hides the initial data. The notation should include $\mathbf{x}_0$ someplace. A common notation for the solution of the IVP (2.7) is $\boldsymbol{\phi}(t; t_0, \mathbf{x}_0)$.

This shows (i) the starting time, (ii) the starting position and (iii) "the current time". Formally,

$$\begin{aligned} &\tfrac{\partial}{\partial t}\boldsymbol{\phi}(t;t_0,\mathbf{x}_0) = \mathbf{f}(t,\boldsymbol{\phi}(t;t_0,\mathbf{x}_0)) \\ &\boldsymbol{\phi}(t_0;t_0,\mathbf{x}_0) = \mathbf{x}_0. \end{aligned} \tag{2.8}$$

The *theorem on existence and uniqueness* states that $\boldsymbol{\phi}$ is a well defined function (for a certain range of arguments), under some mild assumptions on the vector field $\mathbf{f}$; these will be stated when we discuss the existenc/uniqueness theorem, and here I only mention that differentiability of $\mathbf{f}$ is more than enough, but mere continuity is not.

2.3 Properties of the flow of autonomous ODEs

For particles in autonomous flows the timing of the trip does not matter:

$$\boldsymbol{\phi}(t;0,\mathbf{x}_0) = \boldsymbol{\phi}(t+t_0;t_0,\mathbf{x}_0), \tag{2.9}$$

Loosely speaking, only the starting point $\mathbf{x}_0$ and the duration of the trip

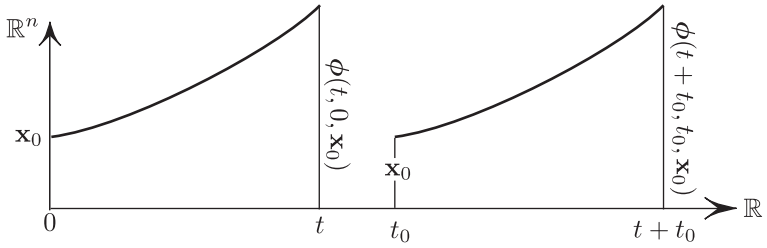

Figure 2.2: Translation invariance (2.9) of the solution operator is inherited from the invariance of the direction field field in $\mathbb{R}\times\mathbb{R}^n$ under t–translations $(t,\mathbf{x})\to(t+t_0,\mathbf{x})$.

determine the destination; the starting time t_0 does not matter.

Problem 2.3. Prove (2.9).

Solution. Both sides of (2.9) satisfy the same ODE (the right one because the time shifted solution is still a solution because $\mathbf{f}$ does not depend on t). For $t=0$ both sides coincide, and thus they coincide for all t by the uniqueness theorem.

According to (2.9) no generality is lost by always choosing choosing the starting time $t_0=0$. So let us not speak of t_0 again and use a shorter notation

$$\boldsymbol{\phi}(t;0,\mathbf{x}_0) \stackrel{def}{=} \boldsymbol{\phi}^t\,\mathbf{x}_0. \tag{2.10}$$

In short, $\boldsymbol{\phi}^t \mathbf{x}_0$ denotes the solution of the IVP $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, $\mathbf{x}(0) = \mathbf{x}_0$. $\boldsymbol{\phi}^t$ is referred to as the t–advance map, or as the *flow* associated with the ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$.

So far we thought of trajectories, i.e. of the path emanating from a fixed $\mathbf{x}_0$. Instead, let us fix t; then for each starting point $\mathbf{x}_0$ we have the destination point $\boldsymbol{\phi}^t \mathbf{x}_0$. That is, we have the time t map $\boldsymbol{\phi}^t : \mathbb{R}^n \mapsto \mathbb{R}^n$, for any value of t. We thus have a one parameter family of maps.

Problem 2.4. Find the maps $\boldsymbol{\phi}^t$ associated with (i) $\dot{x} = x$; (ii) the harmonic oscillator $\dot{x} = y,\ \dot{y} = -x$.

Answer. (i): the dilation $\phi^t x = e^t x$; (ii) $\boldsymbol{\phi}^t \mathbf{x} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \mathbf{x}$, the rotation around the origin through angle $-t$.

Theorem 2.1. *Assume that the autonomous vectorfield* $\mathbf{f}$ *on* $\mathbb{R}^n$ *is such that the flow* ϕ^t *is well defined for all* $t \in \mathbb{R}$ *and for all* $\mathbf{x}_0 \in \mathbb{R}^n$. *Then*

$$\boldsymbol{\phi}^t \circ \boldsymbol{\phi}^s = \boldsymbol{\phi}^{t+s} \quad \textit{for all} \quad t, s \in \mathbb{R}, \tag{2.11}$$

and

$$\boldsymbol{\phi}^0 = id. \tag{2.12}$$

In other words, the one-parameter family of maps $\boldsymbol{\phi}^t$ *is a group under composition.*

Proof. It suffices to prove that $\mathbf{x}(t) = \boldsymbol{\phi}^t(\boldsymbol{\phi}^s \mathbf{x}_0)$ and $\mathbf{y}(t) = \boldsymbol{\phi}^{t+s} \mathbf{x}_0$ are equal for all t and all $\mathbf{x}_0$. Now $\mathbf{x}(t)$ is a solution of $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ by the definition of $\boldsymbol{\phi}$, while $\mathbf{y}(t)$ is the solution of the same ODE because it is a time–shift of the solution $\boldsymbol{\phi}^t \mathbf{x}_0$. Moreover, these solutions coincide for $t = 0$ (with $\boldsymbol{\phi}^s \mathbf{x}_0$), and thus for all t by the uniqueness of solutions. ♢

2.4 More on applications of ODEs.

A major reason why the ODEs are so common is that Newton's second law $\mathbf{F} = m\mathbf{a}$, which governs classical mechanics, usually gives rise to an ODE. Hence the entire subject of classical mechanics is a branch of the theory of ordinary differential equations. Here are a few examples.

The Kepler problem

Kepler's problem deals with the motion of a planet in the gravitational field of the Sun. Effects of other celestial bodies are ignored, as are the tidal effect,

solar wind, relativistic effects, etc. The position vector $\mathbf{r}$ of the planet with the Sun at the origin* satisfies a vector ODE

$$\ddot{\mathbf{r}} = -k\,\frac{\mathbf{r}}{r^3}, \quad r = |\mathbf{r}|, \tag{2.13}$$

where k is a constant.

Problem 2.5. Derive (2.13) from two basic principles: (i) Newton's second law and (ii) Newton's law of gravitational attraction, according to which two point masses attract with the force inverse proportional to their distance from each other.

Newton showed that all trajectories of (2.13) equation are conic sections (ellipses, parabolas or hyperbolas, depending on the initial data), and thereby explained Kepler's laws from two very simple principles, the inverse square gravitational law and the law $F = ma$, both rolled into one differential equation (2.13).† This was probably the greatest success of science since the antiquity up to that time.

Later developments

Since Newton's time, the theory of ODEs underwent great developments at the hands of some of the greatest mathematicians, including Laplace, Legendre, Hamilton, Jacobi, Lyapunov, Poincaré, and more recently Kolmogorov, Arnold and Moser. In the late 1800s Poincaré discovered "chaos" (although the term "chaos" came into use almost 100 years later).

Most developments in the theory of ODEs were stimulated by applications – first by celestial mechanics, and later by the problems of space exploration, by electric circuits, and, more recently, by problems in biology, climate and population dynamics. New questions were asked and new phenomena were discovered. The term of *dynamical systems* – another name for an ODE – came into increasing use starting in the late 1930s.

The theory of ODEs uses methods from many different areas of mathematics, including analysis, differential geometry, topology, number theory, functional analysis, algebra.

*Or, one can choose the common center of mass as the origin–the equation is affected only by a constant factor which can be eliminated by a change of units.

†Actually, Newton showed the converse: if the motion satisfies Kepler's laws then the attraction force varies as inverse square of the distance. But Newton's argument is reversible.

2.5 A birds' eye view.

A first ODE course emphasizes explicit solutions of basic ODEs. For basic ODEs this may work, but there are limitations with looking for explicit solutions, including the following:

1. Fragility: with the tiniest change of the equation the explicit solution may no longer exist. For example, $\dot{x} = x$ can be solved by a formula (namely, $x = ce^t$), but $\dot{x} = x + .00001(\sin x + \sin t)$ can not.

2. If a formula is messy, it may be not that easy to get a geometrical picture of all solutions at a glance.

3. Explicitly solvable equations are extremely rare.

The volume of literature on ODEs may give an illusion that we understand a lot. In reality only a sliver of a huge unexplored universe of ODEs is understood qualitatively. And, as mentioned before, only a tiny subset of these are solvable analytically. And since the majority of systems is not discussed-simply because there is nothing to say about them - we get a somewhat distorted view of reality.

In the retrospect it is not surprising at all that most equations are not solvable explicitly: it would be presumptuous to think that the small collection of elementary functions is sufficient to model the enormous complexity of the physical world. Even a simple–looking pendulum equation exhibits chaotic behavior, as next described, and thus has no hope of being explicitly solvable.

2.6 Chaos and the lack of explicit solutions.

Consider a pendulum consisting of a mass on a weightless rod pivoting on a hinge, and subject to a sinusoidally varying torque at the hinge. With proper scaling, the angle x of the pendulum with the vertical satisfies

$$\ddot{x} + \sin x = \sin t. \tag{2.14}$$

This equation has no solutions expressible by any elementary functions. This may seem surprising until one tries to solve it. What is much more surprising is that this simple-looking equation has "chaotic" solutions: It can be proven to any given **any** infinite sequence of positive integers, for example

$$\mathbf{m} = (9, 28, 2017, 333, 4444, \ldots)$$

there is an associated initial condition $x(0) = x_0, \dot{x}(0) = v_0$ such that the pendulum will tumble 9 times clockwise, then, after hesitating near the top equilibrium, 28 times counterclockwise, 2017 times clockwise, etc, forever – with no further interference on our part. Our choice of the initial conditions encodes the whole sequence!

Note that each digit can be prescribed independently of the preceding one – we could make the choices by tossing a coin and record the number of heads we get after any number of tosses we choose. All this despite uniqueness theorem (which says that the initial data determine the solution uniquely).

What kills the hope that a formula for solving Eq. (2.14) exists? Such a formula would have to exhibit the chaotic behavior just described. None of the formulas we have ever seen shows such behavior. The analytic formulas are too "nice" to behave in such a chaotic fashion. But behind this chaos there IS a simple picture that explains what is really going on, better and more directly than any formula could. I will explain this picture later on. Geometry with its pictures, and not algebra with its formulas, turns out to be the right tool for understanding (2.14).

2.7 The Cauchy Problem and the phase flow.

The *Cauchy problem*, a.k.a. the *inital value problem*, asks for the solution of (2.7) where $t_0 \in \mathbb{R}$ is the prescribed initial time and $\mathbf{x}_0 \in \mathbb{R}^n$ is the prescribed initial position. In a later chapter we will show that if the vector field $\mathbf{f}$ is sufficiently "nice" (e.g. is twice differentiable), then the solution $\mathbf{x}(t) = \phi^t \mathbf{x}_0$ exists for an interval of t–values around t_0, is uniquely defined, and is a smooth function of $\mathbf{x}_0$. Later we will prove this nice dependence, but for now we assume it. To avoid annoying technicalities, we assume that a solution exists for all $t \in \mathbb{R}$. *

We denoted the solution of Eq. (2.7) by $\phi^t \mathbf{x}_0$; according to our assumption, the map $\phi^t : \mathbb{R}^n \mapsto \mathbb{R}^n$ is smooth for any t. By the definition, $\phi^t \mathbf{x}$ is differentiable in t as well.

Definition 2.2. *The family of maps $\phi^t : \mathbb{R}^n \mapsto \mathbb{R}^n$ parametrized by t is called the* **phase flow**. *The map ϕ^t is referred to as a t-***advance map**.

*See problem **??** where this assumption fails.

2.8 Limitations of the theory.

Solvable equations are exceptional. About the simplest example of the differential equation is $\dot{x} = t$, which is solved by simple antidifferentiation: $x(t) = \frac{t^2}{2} + c$, where c is an arbitrary constant. A slightly more complicated example is $\dot{x} = 2x$, with exponential solutions $x = ce^{2t}$.

The vast majority of equations do not admit an explicit solution. This is to be expected; it would be naive to expect that a small supply of functions we learn in school can describe the complexity observed in life. Even a simple-looking equation

$$\ddot{x} + \sin x = \sin t, \tag{2.15}$$

is not solvable by any known functions. This equation describes the angle x of the frictionless pendulum – a unit mass on a stick of unit length, in gravitational field $g = 1$, subject to a sinusoidally varying torque.

In fact, this simple equation exhibits "chaos"! – given any infinite sequence of integers $\mathbf{m} = (m_1, m_2, \ldots, m_k)$ (take the digits of π, for example), there exists an initial condition such that the pendulum will execute m_1 tumbles clockwise, then (after hesitating near the top equilibrium) it will tumble m_2 turns clockwise, etc., ad infinitum. This happens with no external interference, once the initial conditions were imparted! In other words, by the careful choice of initial data we can control future sequence $\mathbf{m}$. The sequence $\mathbf{m}$ is encoded in the initial data, and as the time goes on, further and further "digits" in the initial data "come to the surface" in the form of integers m_k. We term the motion chaotic because the consecutive integers m_k can be prescribed completely independently of one another. Of course, the solution is deterministic: initial data determine the solution uniquely.

Why there is no hope for an explicit solution. With such chaotic behavior there is no hope for a formula for the general solution of Eq. (2.15): elementary functions or anything we can build out of them (by a finite number of steps) don't behave so chaotically.

2.9 Problems

Problem 2.6. Find the mistake in the following "solution" of the differential equation $\dot{x} = x$: integrating, we obtain $x = x^2/2 + c$. Solving this algebraic equation, we get x.

Problem 2.7. A mass of algae in a jar grows at the rate which is proportional to both the volume of the algae mass and to the volume that is

still algae-free. Write down the equation governing the time-evolution of the volume of algae.

Problem 2.8. Consider the one-dimesional ODE $\dot{x} = x(x-1)(x-2)$.

1. Sketch the vector field on $\mathbb{R}$ corresponding to of this ODE, so that the pattern is clear.

2. Consider the solution with $x(0) = \frac{1}{2}$. Find $\lim_{t\to\infty} x(t)$ without solving the equation explicitly.

Problem 2.9. Consider the ODE $\dot{x} = f(x)$, $x \in \mathbb{R}$. Let $a < b$ be two isolated zeros of $f(x)$, with $f(x) > 0$ for all $a < x < b$. Prove that for any solution with $x(0) \in (a, b)$ we have $\lim_{t\to\infty} x(t) = b$, $\lim_{t\to-\infty} x(t) = a$.

Problem 2.10. Write the following equations as first-order systems and sketch the corresponding vectorfields.

1. $\ddot{x} + x = 0$.

2. $\ddot{x} - x = 0$.

3. $\ddot{x} = 0$.

4. $\ddot{x} = -x^{-2}$.

Solution. The systems corresponding to the equations are, respectively:

1. $\dot{x} = y,\ \dot{y} = -x$

2. $\dot{x} = y,\ \dot{y} = x$

3. $\dot{x} = y,\ \dot{y} = 0$

4. $\dot{x} = y,\ \dot{y} = -x^{-2}$

The vectorfields corresponding to these systems are sketched in figure 2.3.

Problem 2.11. The temperature inside the house changes at the rate proportional to the difference in temperatures between the outside and the inside. The outside temperature is changing sinusoidally with time: $T_o(t) = \sin t$. Write the differential equation for the temperature T inside the house, and sketch the vectorfield in the extended phase space.

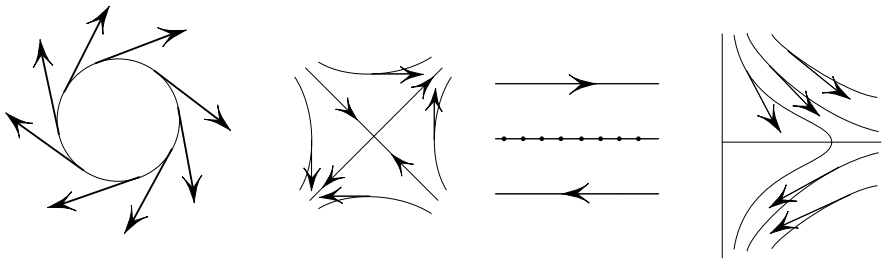

Figure 2.3: For Problem 2.10

Problem 2.12. As in the preceding problem, the inside temperature changes at the rate directly proportional to the temperature difference between inside temperature $x = x(t)$ and the outside temperature $p = p(t)$, which is assumed to be periodic: $p(t+1) = p(t)$ (think of time measured in days, so $t = 1$ corresponds to 24 hours).

1. Write the ODE for the inside temperature $x = x(t)$ (answer in the footnote, don't look to soon)*

2. Show that if the inside temperature is periodic (i.e. $x(t+1) = x(t)$), then its average equals the average outside temperature.

3. Show that the two temperatures are equal at the moment when the inside is the warmest (or the coldest).

4. Show that the moment of maximal temperature difference† is also a moment when the graph of the house temperature has an inflection point.

Problem 2.13. Find the flows ϕ^t for the following ODEs: $\dot{x} = x$, $\dot{x} = x^2$, $\dot{x} = x(1-x)$. Is φ^t defined for all t?

Problem 2.14.

1. Show that any solution of $\dot{x} = 1 + x^2$ blows up in finite time.

2. Show that some solutions of the logistic equation $\dot{x} = x(1-x)$ blow up in finite time when followed backwards in time.

---

*$\dot{x} = -k(x-p)$.

†assume that the maximum is non–degenerate, i.e. that the second derivative is strictly negative.

3. Show that all nontrivial solutions of the second-order equation $\ddot{x} = x^2$ blow up in finite time. Physically, a particle subject to a quadratic repelling force escapes to infinity in finite time*.

4. Show that solutions cannot reach infinity in finite time if $|\mathbf{f}(\mathbf{x})| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^n$. Here C is a constant independent of $\mathbf{x}$.

   Hint: note that $\frac{d}{dt}\left(\frac{\dot{x}^2}{2} - \frac{x^3}{3}\right) = \text{const.}$ for any solution.

Solution. Blow-up of solutions of $\dot{x} = 1 + x^2$. Two methods:

1. the equation is equivalent to $\frac{\dot{x}}{1+x^2} = 1$, i.e. $\frac{d}{dt}\tan^{-1} x = 1$, i.e. $\tan^{-1} x - \tan^{-1} x_0 = t$. Thus $x = \tan(t + c)$ ($c = \tan^{-1} x_0$.) This function is defined on the maximal interval of length π and blows up to $\pm\infty$ as t approaches the ends of the interval.

2. The solution of $\dot{y} = y^2$ with $y(t_0) = 1$ blows up in finite time (for any t_0, as we showed earlier (by explicitly writing $x(t)$). But since $\dot{x} = x^2 + 1 \geq 1$, for any solution $x(t)$ we have $x(t_0) > 1$ for some t_0. By comparison theorem, $x(t) \geq y(t)$ for all $t > t_0$ for as long as the two solutions exist. Since y escapes to infinity in finite time, so does x, and no later than y.

Problem 2.15. Derive the differential equation for the shape of the hanging chain. This curve is referred to as the *catenary*. Hint: the net force acting on an infinitesimal segment of the chain is zero, both in the horizontal and in the vertical directions.

Solution. Let $T(x)$ be the tension of the chain, see figure. The sum of all forces on the arc corresponding to the interval $[x, x + dx]$ is zero both in the x-direction:

$$T(x + \Delta x)\cos\theta(x + \Delta x) = T(x)\cos\theta(x) \tag{2.16}$$

and in the y -direction:

$$T(x + \Delta x)\sin\theta(x + \Delta x) - T(x)\sin\theta(x) = \rho g ds, \tag{2.17}$$

where ρ the linear density of the chain and $ds = \sqrt{1 + y'^2}dx$ is the length of the infinitesimal arc. Dividing the second equation by $T(x)\cos\theta(x) = T(x+\Delta x)\cos\theta(x+\Delta x) \equiv T_0$, we obtain

$$\frac{\tan\theta(x + \Delta x) - \tan\theta(x)}{dx} = \frac{\rho g}{T_0}\frac{ds}{dx},$$

*This idealized situation is, of course, non-realistic.

or, since $\tan\theta(x) = y'$ and $ds/dx = \sqrt{1+y'^2}$:

$$y'' = k\sqrt{1+y'^2},\ k = \frac{\rho g}{T_0}. \tag{2.18}$$

Notes:

1. T_0 is the tension of the chain at the point where the chain bottoms out, and that this is equal to the horizontal component of the tension holding the ends of the chain.
2. By rescaling x and y it suffices to consider the case of $k = 1$.
3. The hanging chain has the shape of the hyperbolic cosine.

Problem 2.16. Consider the pendulum - a point mass on a weightless rod of length ℓ attached to a pivot point by a frictionless hinge. Derive the equation governing the time-evolution of the angle θ between the rod and the vertical.

Answer. $\ddot{\theta} = -\frac{g}{L}\sin\theta$ comes from Newton's second law: $ma = F$ projected on the direction tangent to the circle. Note that $L\theta$ is the distance along the circle to the equilibrium point, so that the acceleration in the tangential direction is $a = \frac{d^2}{dt^2}(L\theta)$. The sum of all forces acting on the mass is $-mg\sin\theta$ (one must draw a figure to see all this). Substituting the above expressions for a and F gives the equation above. Solving this equation is a different matter addressed later.

Problem 2.17. A concrete column with a statue of weight W on top is tapered so that the weight per unit area of a horizontal cross–section is the same for all cross–sections (this way all parts of the column are equally stressed). Find the dependence of the radius on height.

Problem 2.18. A ping-pong ball is subject to the air resistance force linearly proportional to the ball's airspeed. A ball is dropped from height H with zero initial speed. Derive the equation governing the time-evolution of the ball's height off the ground during its fall.

Problem 2.19. * An object is tossed directly upwards from the ground level. Which is greater: the time of ascent or the time of descent? What if the initial velocity is not vertical? Assume that the air resistance is an increasing function of the speed.

Problem 2.20. A toy boat's "wave engine" extracts energy from water waves. The force produced by the "wave engine" is linearly proportional to

the speed of the waves relative to the boat. The drag on the boat is linearly proportional to the square of the speed of the boat relative to the water. Write the differential equation for the speed of the boat. The speed w of the waves relative to the water is given.

Solution. $m\ddot{x} = a(\dot{x} - w) - b|\dot{x}|\dot{x}$.

Problem 2.21. A bullet is fired from the surface of the Moon vertically upwards. Write down the differential equation governing the distance of the bullet to the Moon's surface. The gravitational acceleration on Moon's surface is g_m, and the radius of the Moon is R.

Solution. $\ddot{x} = -\frac{k}{(x+R)^2}$. To find k note that at the surface ($x = 0$) we have $\frac{k}{(R)^2} = g_m$, so that $k = g_m R^2$, and the ODE becomes $\ddot{x} = -g_m \frac{R^2}{(x+R)^2}$.

Problem 2.22. If a plane flying straight up with speed of Max 3 (roughly $1\,km/sec$) suddenly turned off its engine, and if the air resistance vanished, how high would it get? Take $g = 10m/sec^2$.

Problem 2.23. A capacitor* starts out with charge q_0 is then shorted by a resistor R. Try to guess the formula for the charge $q(t)$ on the capacitor at time t. Then write the differential equation for $q(t)$ and compare with the guess.

Solution. The guessing process: it feels like the charge should be decaying exponentially, i.e. $q(t) = ce^{-at}$; what are c and a? First, $q(0) = q_0$ gives $c = q(0)$. Now a large R makes the decay of $q(t)$ slow, so R should be in the denominator of a. Similarly, a large capacitance C makes for a slower discharge as well, and so C also belongs in the denominator, so the guess is $c = \frac{1}{RC}$.

Here is a rigorous solution: $V_C + V_R = 0$ (Kirkhoff's First Law, see Chapter **??** for all the necessary circuits background). Substituting $V_C = q/C$ (definition of capacitance) and $V_R = IR$ (Ohm's law), where $I = \dot{q}$ (definition of the current) we get $q/C + \dot{q}R = 0$, or

$$\dot{q} = -kq,\ k = \frac{1}{RC}. \tag{2.19}$$

We conclude that the charge on the capacitor $q(t) = q_0 e^{-\frac{t}{CR}}$ decays exponentially. This agrees with the guess.

Problem 2.24. Derive the differential equation for the time-evolution of the current in an RLC circuit.

*For the necessary background on electric circuits see Chapter **??**

Problem 2.25. A bank account grows with the instantaneous speed proportional to the amount present, doubling after a year. (a) What happens to the amount after half a year? (b) Prove that the length of time it takes for the money to (say) triple does not depend on the starting date.

Problem 2.26. An island has a population of rabbits and foxes. From one day to the next, the population of rabbits increases by the amount proportional to the number of rabbits present and decreases by the amount proportional to the number of foxes present. The population of foxes increases in proportion of the number present, and also increases in proportion to the population of rabbits. Write the system of differential equations approximating the evolution of the two populations with time, treating these populations as continuous functions of time.

Problem 2.27. Consider the IVP

$$\dot{y} = a(t)y + b(t), \quad y(0) = 0, \tag{2.20}$$

where the function a is such that the solution of the IVP $\dot{x} = a(t)x$, $x(0) = 1$ satisfies $0 \leq x \leq e^{-t}$ for all $t \geq 0$, and where b satisfies $0 \leq b(t) \leq e^{-t}$. *True or false:* the solution of (2.20) is bounded for all $t \geq 0$.

Problem 2.28. Two identical glasses A and B of water are at different temperatures: $0°C$ and $100°C$ Figure **??**. Cold water is clean; the hot water is dirty. We want to heat the clean water using dirty water without mixing them. To that end, we pump cold water from A through a thin tube into into the third glass C. Part of the tube is submerged into the hot and dirty glass B, so that the water picks up some heat from B and emerges from the tube at the same temperature as B. The dirty glass B is gradually cooling, as cold water passing through the pipe picks up some of the heat. Find the eventual temperature of the clean water after all all of it ends up in the third glass (and mixed, so that the water temperature is the same throughout the glass. Assume that no heat is lost to the surrounding medium, that the vessels are perfectly insulating, that the water in each glass has the same temperature throughout, and also assume that the pipe conducts the heat perfectly.

Problem 2.29. A tick jumps on a passing person. Assume that he/she jumps the distance $1m$, landing with speed of $.1m/sec$. What is his approximate takeoff speed, given that his free fall speed is $.5m/sec$. Neglect the gravity, i.e. assume that the tick moves horizontally.

Problem 2.30. I am twirling a stone on a rope in a circle. Both ends of the rope go in concentric circles of radii $r < R$. The rope forms angle θ with the radius–vector of the fingertips holding the rope. Explain why such concentric motion implies accelerating spin. Write the ODE for the speed v of the stone. Find $v(1)$ given that $v(0) = 1$, $\theta = \pi/4$, and $R = 1$. (all quantities are measured in the units of the same system.

Problem 2.31. A bicycle is guided by hand so that its front wheel goes straight along the curb with speed v. The frame of the bike forms angle θ with the direction of the curb. Write the differential equation satisfied by θ. In this problem the bike is a segment RF, with R and F representing the points of contact of the wheels with the ground, with $|RF| = \text{const.}$, and with the velocity of R pointing at F at all times.

2.10 English-to-Math Translation Problems

The following problems exercise translation skills. There is no physics required beyond what's in calculus books (i.e.understanding the precise meaning of the speed, acceleration, and Newton's second law (which says that the acceleration of a particle is proportional to the vector sum of all forces acting on it and inverse proportional to the particle's mass).

Problem 2.32. Let $x(t)$ denote the amount of money (in dollars, say) in an account at time t. The amount grows with the speed (measured in dollars per year) equal to 5% of the amount present. Express this as an ODE for $x(t)$.

Problem 2.33. A bullet is shot into water vertically down. Let $x(t)$ be the penetration distance at time t. The bullet is subject to two forces: water resistance, directly proportional to the square of the velocity, and gravity. Express this as an ODE for $x(t)$.

Problem 2.34. Let $x(t)$ be the angle formed by the string of the pendulum with the downward vertical direction. The angular acceleration of the pendulum is in direct proportion to the torque of the gravity relative to the pivot.* Express this as an ODE for $x(t)$.

Problem 2.35. Let $x(t)$ be the position at time t of a bug crawling on the x–axis. The bug's velocity is in direct proportion to his distance to the point $x = 1$. Express this as an ODE for $x(t)$.

*Recall the definition of torque from Math 230 or 231: it is the "intensity of turning", given by the product of the lever and the component of the force perpendicular to the lever.

Problem 2.36. Let $x(t)$ denote the position at time t of a point mass m on the x–axis. The mass is subject to the force directly proportional to the distance of the mass to the point $x = 2$, and pointing towards that point. Express this as an ODE for $x(t)$.

Problem 2.37. A point mass m in the (x, y)–plane is subject to a constant force pulling it directly down in the direction of the negative y–axis (and no other forces). Express this as a pair of ODEs for x and for y.

Problem 2.38. A point mass m in the (x, y)–plane is subject to the force pulling it directly into the origin, and of magnitude proportional to the distance to the origin. Express this as an ODE for the vector $\mathbf{r}(t) = \langle x(t), y(t)\rangle$, or equivalently (i.e. if you prefer) as a pair of ODEs for $x(t)$, $y(t)$.

Problem 2.39. A point mass m in the (x, y)–plane is subject to the gravitational force of the star located at the origin. Express this as an ODE for the vector $\mathbf{r}(t) = \langle x(t), y(t)\rangle$, or equivalently (i.e. if you prefer) as a pair of ODEs for $x(t)$, $y(t)$.

Problem 2.40. Let $x(t)$ be the temperature of a cup of coffee, and let T_0 be the room temperature. The temperature of coffee approaches that of the room at the rate directly proportional to the mismatch between the two temperatures. Express this as an ODE for $x(t)$.

Problem 2.41. A full pot with hot water is standing in the sink. Water is pouring into the pot at the rate of a gallons per minute and spills over the edge (at the same rate a, of course) after being thoroughly mixed. Denoting by $x(t)$ the temperature of the water in the pot at time t, write the ODE for $x(t)$.

Problem 2.42. A rock, thrown upwards, is subject to two forces: gravity and air resistance, assumed to be directly proportional to the speed. Write the ODE for $x(t)$, the height of the rock at time t.

Problem 2.43. 1. Consider a smooth hill with the given elevation function $h(x, y)$, where x and y are the latitude and longitude of the point above which the elevation is measured. A hiker is descending the hill, always choosing the steepest direction down, and so that his speed *projected onto the*(x, y)*–plane* equals the slope at his location. Write the above as the differential equation for the coordinates $(x(t), y(t))$ of the hiker.

2. The same question as above, with the only difference that now he follows the level line, and travels in the direction pointing to the right of the downhill direction.

The following problem explains, in a remarkably transparent way, Euler's formula $e^{it} = \cos t + i \sin t$.

Problem 2.44. A bug travels in the (x, y)–plane, keeping his velocity perpendicular to his position vector at all times and his speed equal to his distance to the origin. Write the ODE for the bug's position vector, using the complex notation, i.e. by writing $z(t) = (x(t), y(t)) = x(t) + iy(t)$ to denote the bug's position.*

Problem 2.45. A projectile in the vertical (x, y)—plane is subject to two forces: gravitational, and the air drag, the latter directly proportional to the speed. Write the ODE for the in the projectile's coordinates. Gravity points down along the y–axis.

Problem 2.46. Translate the following sentences into formulas.

1. Vector **a** is parallel to vector **b** and is twice as long.
2. Vector **c** is a linear combination of the vectors **a** and **b** (and draw a sketch).
3. Vectors **a**, **b** and **c** are linearly dependent.
4. The matrix M applied to a vector **v** rotates **v** by the angle π and doubles the length of **v**.

*Please keep in mind that there is nothing imaginary about $x + iy$; it is simply another way to write (x, y).

Chapter 3

First Order Systems

3.1 Classification

The most general first order ODE has the form

$$\dot{x} = f(t, x). \tag{3.1}$$

Even such a simple–looking form allows too much variety to be captured by elementary functions. Therefore we list important special cases and leave the vast sea of all other possibilities unexplored.[†]

Here the list of the special types of (**??**) we will discuss:

Here they are:

1. *Autonomous:* $\dot{x} = f(x)$.

2. *Linear:* f is linear in x: $\dot{x} = a(t)x + b(t)$.[‡]

3. *Separable:* f separates into a product, each depending only on one variable: $\dot{x} = a(t)b(x)$.

4. *Homogeneous:* f is a homogeneous function, i.e. it depends on the ratio of its variables only: $\dot{x} = f(\frac{x}{t})$.

5. *Riccati's equation:* $\dot{x} = a(t) + b(t)x + c(t)x^2$.

[†]Of course, any *individual* ODE can be solved numerically with some degree of approximation – but we are looking for something meaningful to say about a class of ODEs, not about a single one.

[‡]nonlinearity in t is allowed; what really matters is the linearity in x.

A note on autonomous systems. Autonomous systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in $\mathbb{R}^n$ can be classified by the dimension n. The scalar case $n = 1$ is solvable analytically (the solution boils down to integration). For $n = 2$ analytic solution is impossible except in special cases, but still, the *qualitative* behavior is well understood. And the case $n \geq 3$ is so rich that only special cases have been analyzed, and it is safe to say that this case will never be fully understood.*

In the next four sections we deal with each of the above types.

3.2 Autonomous ODEs

Autonomous first order ODEs are of the form $\dot{x} = f(x)$ remind of an even simpler case $\dot{x} = f(t)$. The latter is solved by taking the antiderivative with respect to t, resulting in the general solution $x = \int f(t) + \text{const.}$. This method does not apply *directly* to the former ODE, since we cannot find $\int f(x)dt$, not knowing the function $x = x(t)$. However, let us switch the roles of x and t: take x to be the *independent variable*, making $t = t(x)$ the unknown function of x; the ODE $dx/dt = f(x)$ then becomes

$$\frac{dt}{dx} = \frac{1}{f(x)},$$

which we now solve by the old method of taking the antiderivative w.r.t. x:

$$t = \int \frac{dx}{f(x)} + \text{const.}$$

If we are given initial data $x(t_0) = x_0$ then the solution is given by

$$t - t_0 = \int_{x_0}^{x} \frac{dx}{f(x)}. \tag{3.2}$$

If x is such that f has no zeros between x and x_0, then t is a monotone function of x. Thus x is uniquely defined by t, and hence (3.2) defines a solution to the IVP. And there is no other solution, because (3.2) is a consequence of $\dot{x} = f(x)$ and $x(t_0) = x_0$.

Problem 3.1. Show that every solution of $\dot{x} = x(1 - x)\cos x$ with $x(0) = x_0 \in (0, 1)$ approaches 1 as $t \to \infty$.

*This reminds of Stanislaw Lem's parody [**?**] where he classifies the types of genius: the first type is "run of the mill", recognized during his lifetime; the second type of genius is too far ahead of other and is recognized only posthumously; and the pinnacle type #3 is so far advanced that he is never recognized.

Solution. For any $t > 0$ there is a unique $x = x(t) \in (x_0, 1)$ such that

$$t = \int_{x_0}^{x} \frac{dx}{f(x)}, \quad \text{where} \quad f(x) = x(1-x)\cos x.$$

As $x \to 1$, $t \to \infty$ since the integral above diverges: $1/f(x) \geq c/x$ on $(0,1)$, where $c = \cos 1$. The last sentence treated x as the independent variable. But since also t determines x uniquely from the above formula, we conclude that as $t \to \infty$, $x(t) \to 1$.

3.3 Linear ODEs

Linear first order ODEs are of the form

$$\dot{x} = a(t)x + b(t); \tag{3.3}$$

recall that *linear* refers to the x–dependence in f; the t–dependence can be nonlinear, with no restrictions on the functions a and b (apart from the properties needed as we go).

Lagrange's method of variation of constant. To solve this ODE, we first solve the homogeneous case

$$\dot{y} = a(t)y$$

by separation of variables. Rewriting this as

$$\frac{dy}{y} = a(t)\,dt$$

we integrate, obtaining $\ln y = \int a(t)\,dt + c$, or

$$y = Ce^{A(t)}, \quad \text{where} \quad A(t) = \int a(t)\,dt.$$

Lagrange's wonderful idea* was to make C vary with t (hence the name "the method of variation of constant") so as to make $C(t)e^{A(t)}$ satisfy (3.3). This dictates substituting the guess into the ODE (from now on we write $C(t) = C, a(t) = a$, etc.):

$$\frac{d}{dt}(Ce^A) = aCe^A + b, \quad \text{or} \quad \dot{C}e^A + \cancel{Cae^A} = \cancel{aCe^A} + b,$$

*among the many others he had; one of these is the second greatest contribution to classical mechanics since Newton.

using $\dot{A} = a$. The reduces to $\dot{C}e^A = b$, forcing the choice of $C = \int e^{-A} b\, dt + c$. We obtained a solution of (3.3)

$$x = Ce^A = e^A \left(c + \int e^{-A} b\, dt \right), \tag{3.4}$$

for any constant c. It is more interesting to know the solution in terms of the initial condition, rather than in terms of some c. Note that so far A was also defined up to a constant, so to eliminate this ambiguity, let us make a specific choice:

$$A(t) = \int_0^t a(s)\, ds.$$

With this choice of A, we find c from (3.4) as $c = x(t_0)$, and obtain the solution

$$x(t) = x(0)e^{A(t)} + e^{A(t)} \int_0^t e^{-A(\tau)} b(\tau)\, d\tau \tag{3.5}$$

in terms of its initial condition.

This completes the description of Lagrange's method of variation of constants.

A heuristic derivation of (3.5). It helps to think in concrete terms: let us interpret

1. $x(t)$ in (3.3) as the amount of money in an account at time t;
2. $a(t)$ as the variable interest rate $a(t)$, and finally
3. $b(t)$ as the rate of continuous deposit: during an infinitesimal time $[\tau, \tau + d\tau]$ the amount $b(\tau)\, d\tau$ is deposited.

Here is a derivation/explanation of (3.5) with "bare hands", using only one fact:

D_0 **dollars at** $t = 0$ **turn into** $D_\tau = D_0 e^{A(\tau)}$ **dollars at time** τ ,

assuming no deposits are made in the interim.

Returning to our ODE (3.3), note first that the first term in its solution (3.5) is the result of growth of the initial amount $x(0)$ without accounting for the deposits. Now to find the contribution of the continuously made deposit, let us divide $[0, t]$ into small intervals $[\tau, \tau + d\tau]$; deposited during $[\tau, \tau + d\tau]$ is $b(\tau)\, d\tau$; this amount would at time $t = 0$ would have been $e^{-A(\tau)} b(\tau)\, d\tau$, and this, left to grow to time t turns into

$$e^{A(t)} e^{-A(\tau)} b(\tau)\, d\tau; \tag{3.6}$$

integrating gives precisely the last term in (3.5)!* This competes the heuristic explanation of the variations of constant formula (3.5).

3.4 Separable ODEs

Here is another subclass of the general class $\dot{x} = f(t, x)$: separable ODEs

$$\dot{x} = a(t)b(x). \tag{3.7}$$

This includes two even more special subclasses: $\dot{x} = a(t)$ and $\dot{x} = b(x)$, so special that they were discussed in calculus courses. The term "separable" not only stands for the main feature of the right–hand side of the ODE, but also suggests the solution: (1) separate the variables to the opposite sides of the equation:

$$\frac{dx}{b(x)} = a(t)\ dt \tag{3.8}$$

and (2) integrate, getting $\int \frac{\dot{x}}{b(x)} = \int a(t)\ dt$ – the desired solution, implicit in x and containing an arbitrary constant which can be adjusted to satisfy the initial condition, if given.

Now the meaning of (3.8) is a bit vague, since we did not clarify what is meant by, say, $a(t)\,dt$. To make a more honest solution of the above we separate the variables in (3.7):

$$\frac{\dot{x}}{b(x)} = a(t) \tag{3.9}$$

and note that the left–hand side is the t–derivative of

$$B(x) = \int_0^x \frac{ds}{b(s)},$$

so that (3.9) turns into

$$\frac{d}{dt}B(x(t)) = a(t).$$

Integrating and using FTC we get

$$B(x(t)) - B(x(0)) = \int_0^t a(\tau)\,d\tau. \tag{3.10}$$

This essentially completes the solution: all that's left is to solve for $x(t)$.

*The term $e^{A(t)}$ was/can be factored out since it is independent of τ.

Example. Solve $\dot{x} = \frac{1}{3}x^{-2}$ with the initial condition $x(0) = 1$.
Solution. Rewrite the equation as $3x^2\dot{x} = 1$ (this is what one calls separation of variables), and notice that the left-hand side is the time-derivative of x^3. Integrating both sides from $t = 0$ to t we get

$$\int_0^t 3x^2(\tau)\dot{x}(\tau)d\tau = \int_0^t d\tau,$$

or $x^3(t) - x^3(0) = t$. Substituting the initial condition $x(0) = 1$, we obtain $x^3 = t + 1$, and solving for x we obtain the solution explicitly:

$$x(t) = (t + 1)^{\frac{1}{3}}.$$

This is the solution, given implicitly.

Problem 3.2. Solve the IVP $\dot{x} = x^2t$, $x(0) = 1$.

Problem 3.3. Under what conditions on b does (3.10) determine $x(t)$ uniquely?

Problem 3.4.
1. Show that the solution of the IVP $\dot{x} = f(x)$, $x(0) = x_0$ exists (for t in a certain interval) and is unique provided that f is merely continuous and $f(x) \neq 0$ for all x.
2. Show that without the assumption $f(x) \neq 0$ the solution may not be unique.

3.5 Homogeneous ODEs

Here is another subclass of the general case. The ODE reducible to the form

$$\dot{x} = f\left(\frac{x}{t}\right) \tag{3.11}$$

is called *homogeneous*. Homogeneity may not be immediately obvious, as in the example

$$\dot{x} = \frac{at + bx}{ct + dx}.$$

Geometrically, homogeneity amounts to the following statement: the slope of the direction field along each ray $x = kt$ is fixed; in other words, each such ray is an *isocline* (unfortunately no one uses the term *isoslope* instead of isocline).

Geometrically, homogeneity amounts to the invariance of the direction field under dilations $(t, x) \mapsto (kt, kx)$; under such dilations x/t doesn't change, which suggests using $s = x/t$ as the new unknown. To rewrite (3.11) with the new unknown s we note that $\dot{x} = \frac{d}{dt}(st) = \dot{s}t + s$, and (3.11) becomes

$$\dot{s}t + s = f(s),$$

which is separable in a disguise, which we remove by rewriting it as

$$\frac{\dot{s}}{f(s) - s} = \frac{1}{t},$$

which we know how to solve from the preceding section.

3.6 Riccati's equation

Riccati's equation is quadratic in the unknown function:

$$\dot{x} = a(t) + b(t)x + c(t)x^2. \tag{3.12}$$

If $c(t) \equiv 0$ then this is just the linear inhomogeneous ODE we solved completely. This one is not generally solvable by using integrals and elementary functions.

Perhaps the main reason this equation is of interest is that it describes the time–evolution of the slope of vectors $\mathbf{x}(t)$ satisfying linear ODEs $\dot{\mathbf{x}} = A(t)\mathbf{x}$ in $\mathbb{R}^2$, where $A(t)$ is a 2×2 matrix possibly depending on t. This equation cannot in general be solved in elementary functions. The reason I bring this equation up is that it describes linear systems in $\mathbb{R}^2$; this is described in the following theorem.

Theorem 3.1. *Consider a linear ODE $\dot{\mathbf{x}} = A(t)\mathbf{x}$ in $\mathbb{R}^2$:*

$$\begin{cases} \dot{x}_1 = a_{11}x_1 + a_{12}x_2 \\ \dot{x}_2 = a_{21}x_1 + a_{22}x_2 \end{cases} \tag{3.13}$$

The slope of any solution (x_1, x_2) of the above system, i.e. the ratio $x = x_2/x_1$, satisfies the Riccati equation (3.12) with

$$a = a_{21},\ b = a_{22} - a_{11},\quad c = -a_{12}.$$

Problem 3.5. Prove the above theorem. Hint: differentiate x and use the fact that (x_1, x_2) is a solution.

Problem 3.6. Find a linear system in $\mathbb{R}^2$ whose solutions' slopes satisfy the ODE $\dot{x} = x^2$.

Problem 3.7. Can you explain geometrically the blow–up of solutions of $\dot{x} = x^2$ in finite time? Can you actually determine the moment of the blow–up time geometrically without using the solution formula?

Problem 3.8. Can you solve the IVP $\dot{x} = x^2$, $x(0) = x_0$ by treating x as the slope of a solution of a linear system?

3.7 Qualitative theory of first order autonomous ODEs.

The ODE $\dot{x} = f(x)$ can be interpreted as follows: the x–axis is a "highway" with velocity $f(x)$ prescribed at every single location x, Figure. The zero velocity points, i.e. the zeros of f, organize the behavior of the system. If $a < b$ are two neighboring zeros of f then f has a fixed sign on the interval (a, b), and thus every "car" on (a, b) moves either to the right or to the left. Figure **??** shows the flow on a line; the graph of $f(x)$ is plotted in the same

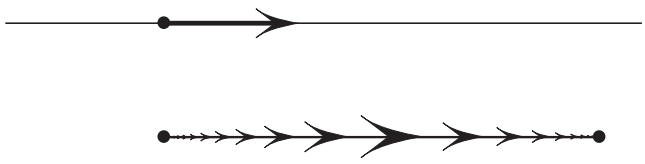

Figure 3.1: Interpreting the ODE $\dot{x} = f(x)$ as a vector field on the line.

figure; the points move right along the x–axis where $f(x) > 0$.

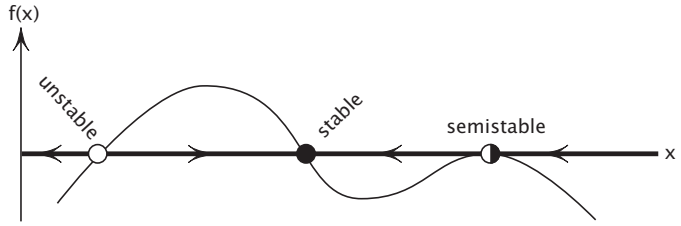

Figure 3.2: Qualitative picture of the flow on the line.

Figure **??**

Definition 3.1. *A constant solution $x(t) = a = const.$ of an ODE is called an equilibrium, an equilibrium solution, or a rest point. An equilibrium*

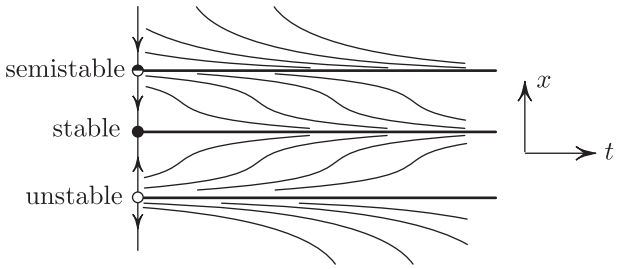

Figure 3.3: Solution curves in the (t, x)–plane.

$x = a$ is said to be stable if all solutions starting near it approach it as $t \to \infty$, or formally, if there exists $\delta > 0$ such that $\varphi^t x \to a$ as $t \to \infty$ for all $x \in (a - \delta, a + \delta)$.

If $x(t) = a$ is an equilibrium solution of $\dot{x} = f(x)$, then $f(a) = 0$. Indeed, since a is a solution, we have $\dot{a} = f(a)$, implying $f(a) = 0$. Conversely, if $f(a) = 0$, then $x(t) = a$ is a solution – indeed, it satisfies the ODE $\dot{a} = f(a)$, since both sides vanish.

The following theorem gives a complete qualitative description of any system $\dot{x} = f(x)$.

Theorem 3.2. *Assume that $f : [a, b] \to \mathbb{R}$ is differentiable, with $f(a) = f(b) = 0$ for $a < b$ and that $f(x) > 0$ for all $x \in (a, b)$. Then any solution $x(t)$ of $\dot{x} = f(x)$ with $x(0) \in (a, b)$ approaches b as $t \to \infty$ and approaches a as $t \to -\infty$. $\lim_{t\to\infty} \phi^t x \to b$ and $\lim_{t\to-\infty} \phi^t x \to a$.*

Proof. Fix any $x \in (a, b)$. By the uniqueness theorem (which applies since f is differentiable), $a < \phi^t x < b$ for all $t \in \mathbb{R}$. But then $f(\phi^t x) = \frac{d}{dt}\phi^t x > 0$ for all t, so that $\phi^t x$ is a monotone increasing function. Since it is also bounded from above (by b), the limit $\lim_{t\to\infty} \phi^t x_0 = b^* \leq b$ exists. We must only show that $b^* = b$, which amounts to proving that $f(b^*) = 0$. Assuming the contrary: $b^* < b$, we have $f(b^*) > 0$. By the continuity assumption f is positive also on an interval surrounding b^*: there exist $\delta > 0$, $\varepsilon > 0$ such that $f(x) \geq \varepsilon > 0$ for $(b^* - \delta, b^* + \delta)$. Hence in time less than $2\delta/\varepsilon$ the solution would pass through the interval, never to return to it because of monotonicity. Thus b^* cannot be the limit. The contradiction completes the proof. ♢

Corollary 3.1. *The equilibrium $x = a$ is stable if $f'(a) < 0$ and backwards stable (i.e. it attracts nearby solutions as $t \to -\infty$) if $f'(a) > 0$.*

Proof. Assume $f'(a) < 0$. Then $f(x) > 0$ for $x < a$, assuming x is also sufficiently close to a. By the above proof $\phi^t x \to a$ as $t \to \infty$. The same holds for x on the other side of a. The case of $f'(a) > 0$ is treated similarly.

Remark 3.1. *In the higher dimensional system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ the above stability condition of an equilibrium is replaced by the condition that the matrix $\mathbf{f}'(\mathbf{a})$ has eigenvalues in the left half plane.*

3.8 Comparison Theorems for $\dot{x} = f(t, x)$

Sometimes the only way to get information on an otherwise intractable ODE $\dot{x} = \mathbf{f}(t, x)$ is to compare it to a simpler one. The following theorem is the tool for such comparison.

Theorem 3.3. *Consider two ODEs $\dot{x} = f(t, x)$ and $\dot{y} = g(t, y)$, with $g(t, x) \geq f(t, x)$ for all t, x and with the initial conditions $y(0) \geq x(0)$. Assume that the solutions depend continuously on initial conditions*[*]*. Then $y(t) \geq x(t)$ for all $t \geq 0$ for which both solutions are defined.*

Proof. Consider at first the strict inequality case: $g(t, x) > f(t, x)$, $y(0) > x(0)$. We claim that

$$y(t) > x(t) \text{ for all } t \geq 0.^{\dagger} \tag{3.14}$$

This is clear from Figure: the graph of y starts above that of x (since $y(0) > x(0)$), and must remain above, since otherwise at the first time $t = t^*$ of crossing we would have had $\dot{y}(t^*) \leq \dot{x}(t^*)$, i.e. $g(t^*, y(t^*) \leq f(t^*, x(t^*))$, where $y(t^*) = x(t^*)$, contradicting the assumption $g(t, y) > f(t, y)$.

It remains to remove the strictness assumptions in (3.14), which we do by a perturbation argument, making the inequality strict and then taking a limit. Let us perturb one ODE: $\dot{y}_\varepsilon = g(t, y_\varepsilon) + \varepsilon$ and its initial condition: $y_\varepsilon(0) = y_0 + \varepsilon$. The inequalities between this perturbed IVP and that for x now strict, and the result of the preceding paragraph applies:

$$y_\varepsilon(t) > x(t) \text{ for all } t \geq 0 \tag{3.15}$$

for all $\varepsilon > 0$. Since $y_\varepsilon(t)$ depends continuously on the parameter ε, $\lim_{\varepsilon \to 0} y_\varepsilon(t) = y(t)$. Taking limit as $\varepsilon \downarrow 0$ in Eq. (3.15) gives $y(t) \geq x(t)$ for $t \geq 0$. ♢

[*] it suffices to require that f and g be continuously differentiable in x and measurable in t.

[†] We assume that the solutions are defined for all $t \geq 0$; otherwise t must be additionally restricted to the common interval of existence of the two solutions.

Problem 3.9. Give a rigorous proof of the fact that if $y(t) > x(t)$ for $t \in [0, t^*)$ and if $y(t^*) = x(t^*)$, then $\dot{y}(t^*) \leq \dot{x}(t^*)$. Must the last inequality be strict?

3.9 Numerical solutions of $\dot{x} = f(t, x)$

I will describe two methods for solving the IVP

$$\begin{aligned} \dot{x} &= f(t, x) \\ x(t_0) &= x_0. \end{aligned} \tag{3.16}$$

The first method (Euler's) is very simple but not very accurate; the second one (Runge–Kutta's) is the other way around: very accurate but not very simple.

Euler's method

Discrete compounding is an example of Euler's method when applied to the simplest ODE $\dot{x} = ax$. The idea is exactly the same for a general ODE $\dot{x} = f(t, x)$. Figure **??** explains the method: We compute the slope $f(t_0, x_0)$

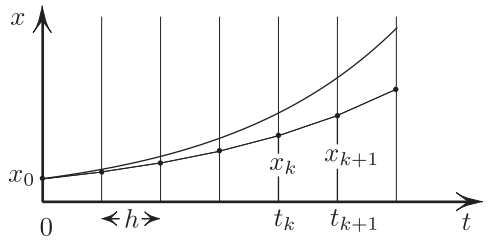

Figure 3.4: Euler's method.

at the starting point (t_0, x_0), and follow the straight line with this slope to the intersection with the line $t = t_0 + h$, where h is a small step size we choose, and repeat the process, treating the point of intersection as the new starting point. In other words, we define the iteration process

$$\begin{aligned} t_{n+1} &= t_n + h \\ x_{n+1} &= x_n + f(t_n, x_n)h. \end{aligned}$$

Example. Consider the IVP $\dot{x} = x, \quad x(0) = 1$; find an approximation of $x(1)$ using Euler's method. Dividing the unit time interval into N pieces,

we get the step size $h = \frac{1}{N}$ (for an integer n). Iteration step is $x_{n+1} = x_n + f(t_n, x_n)h = \big(1 + \frac{1}{N}\big)x_n$. Using $x_0 = 1$ we obtain, after N steps:

$$x_N = \left(1 + \frac{1}{N}\right)^N,$$

the approximation to e, as expected.

Accuracy of Euler's method. The error of Euler's step is, by the definition, the difference between the true solution and its estimate after one step h. Since a smooth curve deviates from its tangent by an amount at most quadratic in the distance from the tangency point, we expect $O(h^2)$ for the error. More formally, we must show that there exists a constant $C > 0$ such that the mismatch between $x(a + h)$ and the linear approximation $x(a) + \dot{x}(a)h$ is quadratically small:

$$|x(a + h) - x(a) - f(a, x(a))h| \le Ch^2. \tag{3.17}$$

But

$$x(a + h) - x(a) = \dot{x}(a)h + \ddot{x}(a + \hat{h})\frac{h^2}{2}, \tag{3.18}$$

for some $0 < \hat{h} < h$, according to Taylor's formula with the remainder in Lagrange's form. Now $\dot{x}(a) = f(a, x(a))$; and $\ddot{x}$ in (3.18) is bounded, since

$$\ddot{x}(t) = \frac{d}{dt}f(t, x(t)) = f_t + f_x\dot{x} = f_t + f_x f,$$

with the derivatives of f bounded by the assumption. In short, the last term in (3.18) is bounded by Ch^2 for some C independent of h. This turns (3.18) into the desired estimate (3.17). ♢

Global truncation error We just estimated an error on one step, but in numerical computation we care about the error after a fixed time, say $t = 1$, i.e. after many small steps. On the hand each error is small; on the other, there are many steps, and the question is who wins: the large number of steps, or the smallness of each error? A rough estimate gives hope: multiplying the each error of size $O(h^2)$ by the number of steps $1/h$ gives the cumulative error $O(h)$.

This is indeed the case, provided f has continuous partial derivatives in t and x. It must be noted that the errors made in the early steps may grow, and it is not immediately clear that they do not accumulate to give a resulting error greater than $O(h)$ – but fortunately they do not; we omit the proof.

3.10 Existence, uniqueness and regularity.

What conditions must $f(t,x)$ satisfy for the Cauchy problem $\dot{x} = f(t,x)$, $x(0) = x_0$ to have a solution, and under what further conditions is the solution unique? And, assuming it exists and is unique, under what conditions does it depend on x_0 continuously? To jump ahead, the answer turns out to be the following. (i) For the existence, mere continuity of f in both variables is (more than) enough. (ii) For uniqueness, the Lipschitz condition in x and mere summability in t is (more than) enough. (iii) The continuous dependence on initial conditions follows from the uniqueness automatically.

The one-dimensional autonomous case $\dot{x} = f(x)$

Assume that $x = 0$ is an equilibrium, i.e. $f(0) = 0$, with no other equilibria in some surrounding interval $[-a, a]$. The constant function $x(t) \equiv 0$ is a solution of the IVP $\dot{x} = f(x)$, $x(0) = 0$. This equilibrium solution is unique if both integrals $\int_0^a \frac{dx}{f(x)}$ and $\int_{-a}^0 \frac{dx}{f(x)}$ diverge. Indeed, the first integral expresses the time it takes for a solution to reach from $-a$ to 0, and to say that this time is infinite is saying that the solution starting at 0 is stuck at 0 forever, since leaving 0 means reaching a in finite time. As an example, for $f(x) = x$ the integral diverges, and indeed the equilibrium solution of $\dot{x} = x$ is unique. On the other hand, for $f(x) = x^{1/3}$ the integral converges, and in fact the uniqueness fails in this case.

Dependence on initial data. Consider the Cauchy problem

$$\dot{x} = f(x,\lambda),\ x(0) = x_0,\ x,\ \lambda \in \mathbb{R} \tag{3.19}$$

where the right-hand side is allowed to depend on a parameter. We already showed by explicit method that the solution exists if $f(x) \neq 0$. We observe that the same method shows that the solution depends continuously on x_0 and on the parameter λ.

Theorem 3.4. *Assume that $f(x,\lambda) \neq 0$ for all $x,\ \lambda$ and that f is continuous in both x and λ. Then Eq. (3.19) has a unique solution which, moreover, depends continuously both on x_0 and λ.*

Proof of Theorem 3.4 Following the separations of variables procedure, we rewrite the Cauchy problem Eq. (3.19) in an equivalent form:

$$F(x,\lambda) = t, \tag{3.20}$$

where

$$F(x,\lambda) = \int_{x_0}^{x} \frac{dy}{f(y,\lambda)}$$

Now, F is monotone in x, and therefore (3.20) defines $x = x(t,\lambda,x_0)$ as a function of t, λ, x_0. Furthermore, F is continuous in λ and x_0. By the implicit function theorem, $x(t,\lambda,x_0)$ is a a continous function of λ and x_0 (and t). ♢

Remark. The autonomous case $\dot{x} = f(x)$ is exceptional in that mere continuity of f together with $f(x) \neq 0$ imply existence and uniqueness (see Problem 3.25). In higher dimension this is no longer true: loosely speaking, there is more room for two solutions to split apart. There are even examples of continuous vectorfields $\mathbf{f}(\mathbf{x})$ in $\mathbb{R}^2$ with infinitely many integral curves passing through every point! This is quite striking: the velocity at each point is strictly prescribed but there are still infinitely many "legal" paths through every point in $\mathbb{R}^2$.

3.11 Linearizing transformation.

In this section we point our microscope at neighborhoods of equilibrium solutions of

$$\dot{x} = f(x) \tag{3.21}$$

.

The following theorem says that there exists a "lens" through which the nonlinear flow $\dot{x} = f(x)$ becomes simply $\dot{y} = -y$, provided a mild assumption is met.

Theorem 3.5. *Assume that $x = 0$ is an equilibirum of (3.21) where f is continuously differentiable in a neighborhood of $x = 0$, with $f'(0) < 0$. There exists a continuous one-to-one mapping h of a neighborhood of the equilibrium $x = 0$ with $h(0) = 0$ which converts (3.21) to the linear ODE $\dot{y} = -y$, in a vicinity of the equilibrium. More precisely, the image $y = h(x)$ of any solution of (3.21) satisfies $\dot{y} = -y$ for as long as $x(t)$ is sufficiently close to 0. Equivalently, the flow map ϕ^t is conjugate via h to the linear contraction $y \mapsto e^{-t}y$.*

Problem 3.10. Find the conjugacy homeomorphism h establishing equivalence of $\dot{x} = -2x$ and $\dot{y} = -y$. Note that h is not differentiable at 0.

Proof. Since f is positive on the left of $x = 0$ and negative on the right in some neighborhood $\mathcal{N} = [-\varepsilon, \varepsilon]$, every solution $\phi^t x$ with $x \in N$ approaches 0 monotonically. Figure 3.5 outlines the construction of the conjugacy map

$y = h(x)$: Starting with $x \in [-\varepsilon, \varepsilon]$, we flow backwards against ϕ until we reach the end of the interval; the flowing time $T = T(x)$ depends on x. Then we flow with $\dot{y} = -y$ for the same time. In short,

$$h(x) = e^{-T(x)}\, \phi^{-T(x)} x, \;\; x \neq 0, \tag{3.22}$$

where $T(x)$ is defined by

$$\psi^{T(x)}\varepsilon = x \;\; \text{if} \;\; x > 0 \;\; \text{or} \;\; \psi^{T(x)}(-\varepsilon) = x \;\; \text{if} \;\; x < 0.$$

Now $h(x) \to 0$ as $x \to 0$, and thus setting $h(0) = 0$ makes h continuous on $[-\varepsilon, \varepsilon]$. Now h is a homeomorphism (one-to-one and continuous), and it satisfies $h \circ \phi^t = e^{-t} h(x)$, as claimed. ♢

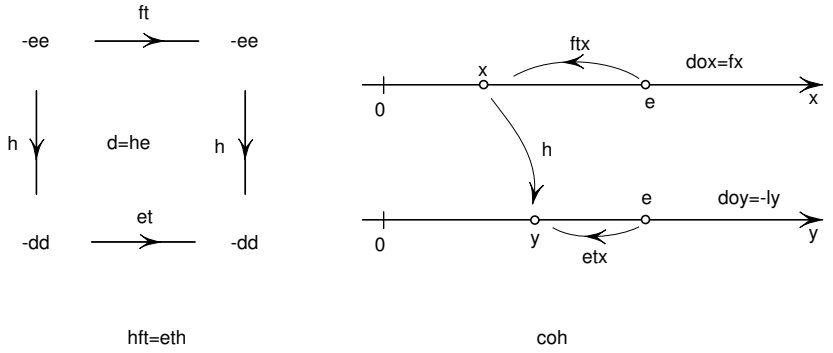

Figure 3.5: The linearizing map h.

3.12 Bifurcations

Bifurcation is a change of qualitative behavior of an ODE when a parameter in the ODE changes. Rather than giving a precise definition of this vague statement, I will simply list three key examples of bifurcations, and only for first order autonomous ODEs. The simplest example of a bifurcation arises in the ODE

$$\dot{x} = x^2 + \lambda, \tag{3.23}$$

where λ is a parameter. For $\lambda < 0$ there are two equilibria at $x = \pm\sqrt{-\lambda}$, one stable and one unstable; for $\lambda = 0$ these coalesce into one semistable equilibrium at $x = 0$; and for $\lambda > 0$ the equilibiria disappear (from the real axis; one can say they become complex, if we decide to allow complex x). In

this case we say that $\lambda = 0$ is the bifurcation value. And the change in the qualitative behavior, namely the passage from a pair of equilibria to none, is called a bifurcation. This particular bifurcation is called the saddle–node bifurcation – the name inherited from its 2D analog, where the a saddle and a node can collide and disappear.*

Here is the list of three "normal forms" of ODEs showing the key types of bifurcations, shown in Figure 3.6

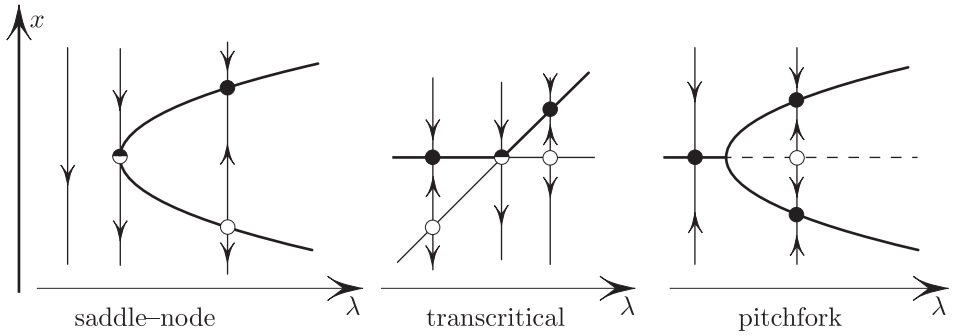

Figure 3.6: Three types of bifurcations.

1. $\dot{x} = -x^2 - \lambda$, the saddle-node bifurcation
2. $\dot{x} = -x^2 - \lambda x$, the transcritical bifurcation.
3. $\dot{x} = -x^3 + \lambda x$, the pitchfork bifurcation.

3.13 Some paradoxes

The surprising fact regarding even the simplest IVP, such as $\dot{x} = \sqrt{x}$, $x(0) = 0$ is that the solution is not unique, as Figure illustrates. Indeed, $x \equiv 0$ is a solution, and so is $t^2/4$. In fact, any function vanishing on the interval $t \in [0, c]$ and equal $(t - c)^2/4$ for $t \geq c$ is a solution.

More than a curiosity, this examples teaches two lessons, as explained next:

1. The non–uniqueness is caused by super–exponential growth
2. Newtonian mechanics is not always deterministic!

*This bifurcation is also called the "blue sky" bifurcation, since if we decrease λ two equilibria appear where none were before, "out of the blue ".

A heuristic explanation of solutions splitting. What is it about $f(x) = \sqrt{x}$ that causes two solutions $x_1(t) = 0$ and $x_2(t) = t^2/4$ to split apart? In other words, how can the difference $D(t) = x_2(t) - x_1(t)$ become nonzero? The answer, as it turns out, is that the logarithmic rate of growth* of D is infinitely large at $t = 0$. Indeed, differentiating D we get $\dot{D} = \dot{x}_2 - \dot{x}_1 = f(x_2) - f(x_1)$ (we write $x_1 = x_1(t)$ for brevity); by the intermediate value theorem $f(x_2) - f(x_1) = f'(\overline{x}(t))D$, for some $\overline{x} \in [x_1, x_2]$. Denoting $f'(\overline{x}(t)) = 1/\sqrt{\overline{x}(t)} = a(t)$, we get

$$\dot{D} = a(t)D; \tag{3.24}$$

note that $a(t) \to \infty$, as $t \downarrow 0$. Had a been bounded, D would have grown no faster than an exponential, and would have remained $= 0$ since an exponential cannot grow from zero to nonzero. But the unboundedness of a destroys this property in this example.

Problem 3.11. What is the logarithmic rate of growth of the length of the interval $[0, 0.001]$ according to the ODE $\dot{x} = \sqrt{x}$?

Another way to think of uniqueness is given in the following paragraph.

Osgood's uniqueness criterion. Let us consider two solutions $x_1(t) = 0$, $x_2(t) = t^2/4$ from the previous example. As Figure **??** shows, it takes finite time for the two solution curves followed backwards from $t > 0$ to "stick together". It is this finiteness of time that is equivalent to non–uniqueness. This is the whole idea behind Osgood's criterion. Let us formulate this criterion precisely, and then prove it rigorously.
Consider the function f with $f(0) = 0$ and with $f(x) \neq 0$ for $x \neq 0$, and consider the IVP

$$\dot{x} = f(x), \quad x(0) = 0.$$

Claim: The equilibrium solution $x(t) \equiv 0$ is unique iff for any $A > 0$ the travel time from 0 to A

$$\int_0^A \frac{dx}{f(x)} = \infty. \tag{3.25}$$

Proof. Assume the contrary: there exists the second solution y different from x, i.e. satisfying with $y(T) \neq 0$, say $y(T) > 0$, for some $T > 0$. The particle representing this solution starts at $x = 0$ at $t = 0$ and reaches

*in other words, $\dot{D}/D$, the rate of change of D as the proportion of D.

$y(T) = A > 0$, and therefore passing every $0 < \varepsilon < A$ on the way: there exists the time $0 < t_\varepsilon < T$ such that $y(t_\varepsilon) = \varepsilon$. Now the duration of the trip from ε to A is $\int dy/\dot{y}$, i.e.

$$T - t_\varepsilon = \int_\varepsilon^A \frac{dy}{f(y)}.$$

But this is a contradiction: the left–hand side does not exceed T, while the right–hand side approaches ∞ as $\varepsilon \downarrow 0$, according to (3.25). To repeat, (3.25) says that it takes infinitely long for two solutions to coalesce, thus guaranteeing uniqueness.

♢

3.14 Problems

Problem 3.12. Find the mistake in the following "solution" of the differential equation $\dot{x} = x$: integrating, we obtain $x = x^2/2 + c$ and solve this quadratic equation for x.

Problem 3.13. *Solve the following initial value problems. Also, describe the asymptotic behavior of* **all** *solutions. In other words, specify what is the limit of a solution depending on the initial condition.*

1. $\dot{x} = x$, $x(0) = 1$
2. $\dot{x} = x(1 - x)$, $x(0) = \frac{1}{2}$
3. $\dot{x} = \sin x$, $x(0) = \pi/2$
4. $\dot{x} = \sin^2 x$, $x(0) = \pi/2$

Problem 3.14. Find all equilibria and classify their stability for the differential equation $\dot{x} = \sin x$. Sketch the phase portrait of the above differential equation and sketch enough solutions in the x, t–plane to fully understand the behavior of all solutions. Draw the bifurcation diagram for the above system.

Problem 3.15. A particle's position x on the line evolves according to $\dot{x} = f(x)$, where x is a given function. Find the particle's acceleration as a function of its position.

Problem 3.16. A particle moves according to $\dot{x} = f(x)$, where $f(x) > 0$ for all $x \in \mathbb{R}$. Find the time of travel from a to $b > a$.

Problem 3.17. Show that no solution of $\dot{x} = f(t, x)$ blows up in finite time if f satisfies $|f(x)| \leq a(t)|x| + b(t)$ for all $t,\ x \in \mathbb{R}$, where $a, b \in C(\mathbb{R})$ (here $C(\mathbb{R})$ denotes the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$).

Hint. Use the comparison theorem.

Problem 3.18. Consider the system

$$\begin{cases} \dot{x} = y \\ \dot{y} = -\omega^2 x \end{cases}, \tag{3.26}$$

where ω is a constant. Find the first order ODE for the angle $\theta = \tan^{-1}(y/x)$ formed by the solution vector (x, y) with the x-axis. What becomes of this ODE if $\omega = 1$? Note that $\dot{\theta}$ does not depend on the distance of the point (x, y) to the origin, but only on θ. What special property of the ODE is responsible for this fact?

Solution. We have:

$$\frac{d}{dt}\theta = \frac{d}{dt}\tan^{-1}(y/x) = \frac{\dot{y}x - \dot{x}y}{x^2 + y^2}$$

(a small algebraic step was skipped). Substituting $x = r\cos\theta,\ y = r\sin\theta$, where $r = \sqrt{x^2 + y^2}$, $\dot{x} = y$ and $\dot{y} = -\omega^2 x$, we get

$$\dot{\theta} = -\omega^2 \cos^2\theta - \sin^2\theta. \tag{3.27}$$

If $\omega = 1$ the ODE for θ reduces to $\dot{\theta} = -1$. This fits with the geometric analysis we did directly the first day of class. Note that r does not appear in Eq. (3.27). This is due to the fact that (3.26) is a linear system, and so the straight lines through the origin remain straight lines through the origin under the evolution under (3.26).

Problem 3.19. Use the idea of Problem 3.18 to solve the ODEs $\dot{\theta} = \sin^2\theta$, $\dot{\theta} = \sin 2\theta$. Hint: for the first ODE, set $\omega = 0$ in (3.26). For the second ODE, consider the system $\dot{x} = -x,\ \ \dot{y} = y$.

The following problem sketches a simple proof of the Sturm Comparison Theorem.

Problem 3.20. (See Problem 3.18) Consider two differential equations $\ddot{x} + q(t)x = 0$ and $\ddot{y} + p(t)y = 0$ with $p(t) > q(t)$.

1. Write each equation as a system of two ODEs, and consider the angle formed with the positive x-axis by the solution vector of each system.

2. Show that these angles satisfy $\dot{\alpha} = f(t, \alpha)$ and $\dot{\beta} = g(t, \beta)$ with $f(t, \alpha) \geq g(t, \alpha)$. In other words, whenever the vectors $(y, \dot{y})$ and $(x, \dot{x})$ are aligned, one rotates faster than the other.

3. Let $x(t)$ and $y(t)$ be two solutions of the above equations. Show that between any two consecutive zeros of $x(t)$ there is a zero of $y(t)$ (a zero of $x(t)$ is, by the definition, a value of t for which $x(t) = 0$).

Problem 3.21. The logistic equation. In this problem we consider a simple model of the propagation of a rumor, or a joke, or the flu in a population. Let $x(t) \in [0, 1]$ be the proportion of the people in a population who have heard a particular joke. The population is so large that we can treat x as a continuous variable. Think of an enormous party where people mingle randomly, with the average person tells the joke to everyone they meet. Write a simple differential equation for the evolution of x in time.

Problem 3.22. Consider the ODE $\dot{x} = f(x)$ where $-2x \leq f(x) \leq -x$ for all $x > 0$. Prove that the solution with $x(0) = 1$ satisfies $e^{-2t} \leq x(t) \leq e^{-t}$.

Solution. Use the comparison theorem, comparing with $\dot{x} = -x$ and $\dot{x} = -2x$.

In the preceding problem, the solution never becomes zero. The next problem looks at this question closer.

Problem 3.23. Can the solution of $\dot{x} = f(x)$ with $x(0) = 1$, where f is a continuous function with $f(0) = 0$ and $f(x) < 0$ for $x > 0$, reach zero in finite time?

Hint: If speed $f(x)$ is of the same order of magnitude as x, i.e. the speed towards the origin is on the order of the distance to the origin – think of $\dot{x} = -kx$, then x decreases no faster than some exponential and thus will never become zero. The only hope is to make the speed $f(x)$ be much larger than x.
Solution: The time to reach the origin (assuming it is reached) is $\int_1^0 dx/f(x)$. If this is finite, then the $x = 0$ is reached in the finite time. An example of f is $f(x) = \sqrt{x}$, or more generally any power of x between 0 and 1.

The following problem amounts to the well known **Osgood uniqueness theorem**; behind this theorem is a tautologically sounding idea: for two solutions to avoid a meeting the coalescence time must be infinite!

Problem 3.24 (Osgood uniqueness theorem)**.** Consider the Cauchy problem $\dot{x} = f(x)$, $x(0) = 0$, where $f(0) = 0$, and $f(x) \neq 0$ for $\neq 0$. Prove that if the improper integral $\int_0^x \frac{dy}{f(y)} = \infty$ diverges, then the solution $x \equiv 0$ is unique.

Remark 3.2. *If $|f'(0)| < \infty$, then the zero solution is unique (indeed, the above integral diverges). Thus finite divergence $f'(0)$ guarantees uniqueness; this makes intuitive sense. One can say, roughly speaking, that that non-uniqueness is caused by infinite divergence. For example, $f(x) = x^{\frac{1}{3}}$ has infinite divergence: $f'(0) = \infty$ and exhibits non-uniqueness.*

The following is an interesting fact:

Problem 3.25. Show that uniqueness implies continuous dependence for the ODEs $\dot{x} = f(t, x)$, $x \in \mathbb{R}$. In other words, show that if the solution of a Cauchy problem is unique for all initial data, then the solution depends on the initial data continuously.

Chapter 4

Linear systems in the plane

The preview of the chapter

The simplest nontrivial subclass of systems $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ consists of the linear ones, i.e. the ones for which $\mathbf{f}$ is a linear: $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, where A is a constant matrix:

$$\dot{\mathbf{x}} = A\mathbf{x}. \tag{4.1}$$

We limit ourselves to $\mathbf{x} \in \mathbb{R}^2$, although all we do extends to any dimension almost verbatim.

This chapter's main result is this: *any system (4.1) reduces to one of the following* (see Figure):

1. Two decoupled equations $\dot{y}_1 = \lambda_1 y_1$, $\dot{y}_2 = \lambda_2 y_2$.

2. Rotation in the plane coupled with exponential growth/decay: $\mathbf{y} = e^{\alpha t} R(\omega t)\mathbf{y}(0)$, were $R(\theta)$ is the rotation through angle θ.

3. Shear flow coupled with dilation: $\mathbf{y} = e^{\alpha t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mathbf{y}_0$.

Each of these systems is solved trivially, and its trajectories in $\mathbb{R}^2$ are easy to draw. And any solution of the general system (4.1) is simply a linear transformation of one of these simple solutions. All this is explained below in detail.

4.1 The general method

The main idea for solving (4.1) is to first seek special solutions that "slide" along the eigendirections of A, namely, the solutions of the form

$$\mathbf{x}(t) = a(t)\mathbf{v}, \tag{4.2}$$

where $\mathbf{v}$ is an eigenvector* and where $a(t)$ is a scalar function to be determined so as to force (4.2) to satisfy $\dot{\mathbf{x}} = A\mathbf{x}$. Substituting the guess (4.2) into (4.1) we obtain $\dot{a}\mathbf{v} = A(a\mathbf{v})$; using $A\mathbf{v} = \lambda\mathbf{v}$ this turns into

$$\dot{a}\mathbf{v} = \lambda a\mathbf{v}; \tag{4.3}$$

For this to hold, it suffices to set $\dot{a} = \lambda a$, i.e. $a = ce^{\lambda t}$.
To summarize, we proved that

$$\mathbf{x}(t) = ce^{\lambda t}\mathbf{v}, \tag{4.4}$$

is a solution of Eq. (4.1).

Some questions still remain: (1) how to find the most general solution? (2) what if λ and $\mathbf{v}$ are complex? (3) How to plot solution curves? These question are addressed next.

A question. Can you motivate the guess (4.2) geometrically? (This question tests the full understanding of this method.)

4.2 Real eigenvalues.

In this section we consider the simplest case: A has a real eigenbasis $\mathbf{v}_1$, $\mathbf{v}_2$. Before reading this section, make sure you can solve the two background–testing problems at at its end.

Any linear combination

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2 \tag{4.5}$$

is a solution of the system.† Moreover, it is a general solution, meaning that any solution of the system is of the form (4.5). Indeed, let $\mathbf{y}(t)$ be any solution of (4.1). If we can pick c_1, c_2 in (4.5) so as to match the initial value of $\mathbf{x}$ to that of $\mathbf{y}$:

$$\mathbf{x}(0) = \mathbf{y}(0); \tag{4.6}$$

*i.e. $A\mathbf{v} = \lambda\mathbf{v}$ for $\lambda \in \mathbb{R}$ and where $\mathbf{v} \neq \mathbf{0}$

†see Problem 4.1

by the uniqueness theorem, two solutions with identical initial data must be identical. So if we will manage to satisfy (4.6), then we will have $\mathbf{y}(t) = \mathbf{x}(t)$ for all t, thus proving that $\mathbf{y}$ is indeed of the form (4.5). Now (4.6) written out in more details amounts to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{y}(0). \tag{4.7}$$

There is indeed a solution $c_1,\ c_2$ (a unique one, in fact) since $\mathbf{v}_1,\ \mathbf{v}_2$ are linearly independent*. This proves that (4.5) is indeed the general solution. ♢

Problem 4.1. [Superposition principle] Prove that the sum of two solutions of a linear system $\dot{\mathbf{x}} = A\mathbf{x}$ is also a solution, as is the product of a solution with a scalar.

Problem 4.2. Given three vectors $\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{w}$ of which the first two are linearly independent, consider the system $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{w}$ for the unknowns c_1, c_2. Explain geometrically why the solution is unique, and in particular give a geometrical interpretation of c_1, c_2.

Test question. What is meant by the general solution of a system of ODEs?

4.3 Phase portrait in the real case

Having found the general solution (4.5) of $\dot{\mathbf{x}} = A\mathbf{x}$, we now show how to plot these parametric curves.

The main idea is to first understand the very simple special case when $\mathbf{v}_1,\ \mathbf{v}_2$ in (4.5) are replaced by $\mathbf{e}_1,\ \mathbf{e}_2$, the unit vectors of the coordinate axes; and with that done, transform linearly the whole picture so that $\mathbf{e}_1$, $\mathbf{e}_2$ map to $\mathbf{v}_1,\ \mathbf{v}_2$. With such replacement, the simpler curves

$$\mathbf{y} = c_1e^{\lambda_1 t}\mathbf{e}_1 + c_2e^{\lambda_2 t}\mathbf{e}_2 = \begin{pmatrix} c_1e^{\lambda_1 t} \\ c_2e^{\lambda_2 t} \end{pmatrix} \tag{4.8}$$

are easy to plot: their parametric equations are $x = c_1e^{\lambda_1 t},\ y = c_2e^{\lambda_2 t}$; eliminating t this gives $x_2 = cx_1^{\alpha}$, where $\alpha = \lambda_2/\lambda_1$, Figure 4.2(left). Now let T be the linear transformation such that

$$T\mathbf{e}_1 = \mathbf{v}_1,\ \ T\mathbf{e}_2 = \mathbf{v}_2.$$

*see Problem 4.2

Then

$$\mathbf{x}(t) = T\mathbf{y}(t),$$

i.e. the trajectories of (4.5) are the T–images of the simple curves $x_2 = cx_1^\alpha$, as Figure 4.2 illustrates.

Problem 4.3. [An important observation] Show that T is the matrix built out of column–vectors $\mathbf{v}_1$ and $\mathbf{v}_2$.

Problem 4.4. Plot the phase portrait of the linear system whose coefficient matrix has the eigenvectors $\mathbf{e}_1$, $\mathbf{e}_1 + \mathbf{e}_2$ in the following three cases: the corresponding eigenvalues are (i) $-1, -2$; (ii) $-1, -1$; (iii) $-1, 2$.

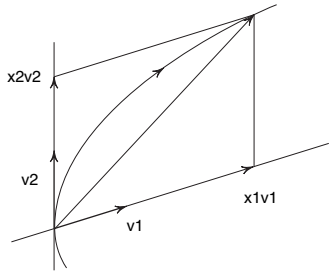

Figure 4.1: Parallel projections of the solution onto eigendirections change exponentially.

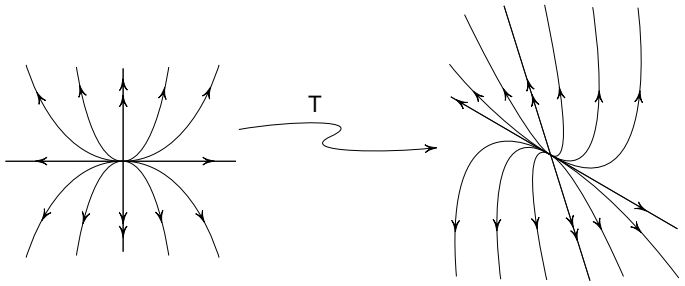

Figure 4.2: Phase portrait of the system.

4.4 Complex Eigenvalues.

Assume that $\lambda = \alpha + i\omega$ is complex eigenvalue of A, with the corresponding eigenvector $\mathbf{v}$. We showed that $\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$ is a solution; the fact that it is

not real is actually a blessing in disguise. Indeed, the fact that a complex number is a pair of real ones gives hope that $\mathbf{x}$ is simply a pair of real solutions. Indeed, let us separate $\mathbf{x}$ into its real and the imaginary parts: $\mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_1$ and substitute this solution into the ODE:

$$\frac{d}{dt}(\mathbf{x}_1 + i\mathbf{x}_2) = A(\mathbf{x}_1 + i\mathbf{x}_2) = A\mathbf{x}_1 + iA\mathbf{x}_2. \tag{4.9}$$

Since A is a real matrix, the vectors $A\mathbf{x}_1$ and $A\mathbf{x}_2$ are real. Matching real to real and imaginary to imaginary in Eq. (4.9) gives $\dot{\mathbf{x}}_1 = A\mathbf{x}_1$ and $\dot{\mathbf{x}}_2 = A\mathbf{x}_2$, as claimed.

Extracting the real and the imaginary parts from

$$\mathbf{x}(t) = e^{(\alpha+i\omega)t}(\mathbf{u} + i\mathbf{w}) = e^{\alpha t}(\cos\omega t + i\sin\omega t)(\mathbf{u} + i\mathbf{w}),$$

we obtain

$$\begin{aligned} \mathbf{x}_1 &= e^{\alpha t}(\cos\omega t\ \mathbf{u} - \sin\omega t\ \mathbf{w}) \\ \mathbf{x}_2 &= e^{\alpha t}(\sin\omega t\ \mathbf{u} + \cos\omega t\ \mathbf{w}). \end{aligned} \tag{4.10}$$

4.5 Phase portrait in the complex case

We use the same approach as in the real case: let T be the linear map taking $\mathbf{e}_1$ to $\mathbf{u}$ and $\mathbf{e}_2$ to $\mathbf{w}$, i.e. the matrix T is built out of columns $\mathbf{u}$, $\mathbf{w}$. Then

$$\mathbf{x}_1 = Te^{\alpha t}\begin{pmatrix} \cos\omega t \\ -\sin\omega t \end{pmatrix},$$

We conclude: $\mathbf{x}_1$ sweeps the image under T of a logarithmic spiral if $\alpha \neq 0$, towards or away from the origin depending on the sign of $\alpha = \operatorname{Re}\lambda$. And if $\alpha = 0$ then $\mathbf{x}_1$ describes an ellipse. The same holds for $\mathbf{x}_2$ with a phase shift.

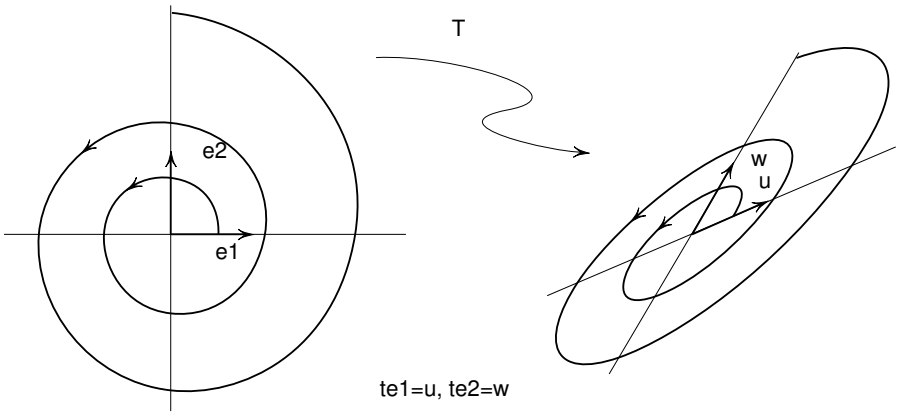

Figure 4.3: Phase portrait in the complex case.

4.6 Multiple eigenvalues.

It remains to consider one more case, when the matrix A does not have a full set of eigenvectors. That is, A has a single eigenvalue with a one–dimensional eigenspace.

By the Jordan Normal Form Theorem, there exists a basis of vectors such that

$$\begin{aligned} A\mathbf{v}_1 &= \lambda\mathbf{v}_1 + \mathbf{v}_2, \\ A\mathbf{v}_2 &= \lambda\mathbf{v}_2 \end{aligned} \tag{4.11}$$

Just like in the preceding two cases, we seek the solution in the form

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2. \tag{4.12}$$

Substituting this into $\dot{\mathbf{x}} = A\mathbf{x}$ and using Eq. (4.11) we obtain

$$\dot{a}_1\mathbf{v}_1 + \dot{a}_2\mathbf{v}_2 = a_1(\lambda\mathbf{v}_1 + \mathbf{v}_2) + a_2\lambda\mathbf{v}_2$$

Equating the coefficients of $\mathbf{v}_1,\ \mathbf{v}_2$, we get

$$\begin{aligned} \dot{a}_1 &= \lambda a_1 \\ \dot{a}_2 &= \lambda a_2 + a_1 \end{aligned} \tag{4.13}$$

We have $a_1 = e^{\lambda t}c_1$,

It remains to solve the last system. One way is to substitute $a_1 = ce^{\lambda t}$ into the second equation and solve the resulting inhomogeneous linear

equation. A nicer way is the following. The form of (4.13) suggests that the exponential growth of the type $e^{\lambda t}$ is present. This gives the idea of setting $b_k = e^{-\lambda t} a_k$ in the hope that the system for b_k is simpler. Indeed, substitute $a_k = e^{\lambda t} b_k$ into Eq. (4.13); after dividing both sides by $e^{\lambda t}$ we get

$$\begin{aligned} \dot{b}_1 &= 0 \\ \dot{b}_2 &= b_1 \end{aligned} \tag{4.14}$$

The solution of this system is

$$b_1 = c_1,\ b_2 = c_1 t + c_2,$$

so that the general solution of Eq. (4.13) is

$$\begin{aligned} a_1 &= e^{\lambda t} c_1 \\ a_2 &= (c_1 t + c_2) e^{\lambda t}; \end{aligned}$$

these are the coordinates of the solution $\mathbf{x}$ in the basis $\{\mathbf{v}_k\}$. Substituting these into the expansion Eq. (4.12) we obtain the general solution (grouping by constants):

$$\mathbf{x}(t) = c_1 e^{\lambda t}(\mathbf{v}_1 + t\mathbf{v}_2) + c_2 e^{\lambda t}\mathbf{v}_2.$$

Problem 4.5. Sketch the phase portrait of the system with the matrices $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and $A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$.

Problem 4.6. Sketch the phase portrait of the system with $A = \begin{pmatrix} -1 & 1 \\ \varepsilon & -1 \end{pmatrix}$ for two small values of ε of opposite signs. How does one portrait transition to the other?

4.7 Summary in the matrix notation.

It is instructive to summarize all three cases of solution of $\dot{\mathbf{x}} = A\mathbf{x}$ in matrix notation.

Substituting $\mathbf{x} = T\mathbf{y}$, where T is a matrix to be specified shortly, into the equation gives $\frac{d}{dt}(T\mathbf{y}) = A(Ty)$, or the transformed equation

$$\dot{\mathbf{y}} = T^{-1} A T \mathbf{y}. \tag{4.15}$$

We now consider three cases: real, complex, and degenerate.

1. If A is diagonalizable with real eigenvalues, let $T = (\mathbf{v}_1\mathbf{v}_2)$ consist of column-eigenvectors of A. Then $T^{-1}AT = D = \text{diag}\,(\lambda_1, \lambda_2)$ and (4.15) turns into
$$\dot{\mathbf{y}} = D\mathbf{y},$$
thus decoupling into two simple equations $\dot{y}_k = \lambda_k y$, $k = 1, 2$. Plotting of solutions was explained earlier.

2. In the complex case $\lambda_{1,2} = \alpha + i\omega$ we build matrix $T = (\mathbf{v}, \mathbf{w})$ out of column-vectors of the real and the imaginary parts of the eigenvector $\mathbf{v}_1 = \mathbf{u} + i\mathbf{w}$. With this choice of T the matrix in (4.15) becomes
$$T^{-1}AT = \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix},$$
as follows from the relation $A\mathbf{v} = \lambda\mathbf{v}$. Now (4.15) becomes
$$\dot{\mathbf{y}} = \begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}\mathbf{y}. \tag{4.16}$$
The solutions curves
$$\mathbf{y} = e^{\alpha t}\begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix}\mathbf{y}_0$$
are logarithmic spirals unless $\alpha = 0$, in which case the are circles. The $\mathbf{x}$-trajectories are the T-images of these spirals.

3. Non-diagonalizable case: $T^{-1}AT = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. The transformed equation $\dot{\mathbf{y}} = (D+N)\mathbf{y}$ has solutions $\mathbf{y} = e^{\lambda t}\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}\mathbf{y}_0$, or in coordinates: $y_1 = (a+bt)e^{\lambda t}$, $y_2 = ce^{\lambda t}$, see figure*. Again, the trajectories of $\mathbf{x}$ are the T–images of the trajectories of $\mathbf{y}$.

*Here is an intuitive explanation of the shape of the curves: first, imagine $\lambda = 0$; in that particular case we have a shear flow: each solution follows a horizontal straight path as shown in figure. Now the effect of $\lambda < 0$ (say) is to attract solutions to the origin. In a nutshell, this flow is a simultaneous action of shear and contraction (if $\lambda < 0$.

4.8 Problems.

Problem 4.7. Consider four *identical* balls connected by heat-conducting rods, so as to form a tetrahedron. Denote by T_k the temperature of the kth ball, and let a_{kl} be the heat conductivity of the rod connecting the kth and lth balls. The heat flows through the rods at the rate proportional to the temperature difference.

1. Write the system $\dot{\mathbf{x}} = A\mathbf{x}$ for the evolution of the temperature vector $\mathbf{x} = \mathrm{col}(T_1, \ldots, T_4)$ by specifying the matrix A. No heat is lost to the surrounding medium.

2. Show that $\sum T_k = \mathrm{const}$.

3. Show that $A^T\mathbf{1} = \mathbf{0}$, where $\mathbf{1} = \mathrm{col}(1, \ldots, 1)$.

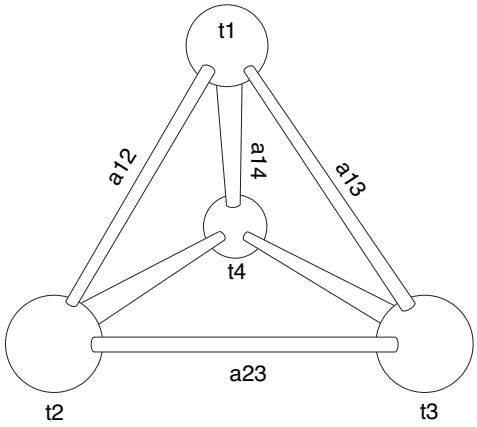

Figure 4.4: The heat flows through the rods

Problem 4.8. Consider the system consisting of three balls linked by heat-conducting channels, figure 4.5. The balls at the two ends are connected to heat sinks kept at zero temperature.

1. Write the ODE for temperatures vector $\mathbf{x} = (T_1, T_2, T_3)$ in vector form $\dot{\mathbf{x}} = A\mathbf{x}$.

2. Find the eigensolutions $\mathbf{x}_k(t) = e^{\lambda_k t}\mathbf{v}_k$. Verify that the eigenvectors are mutually orthogonal.

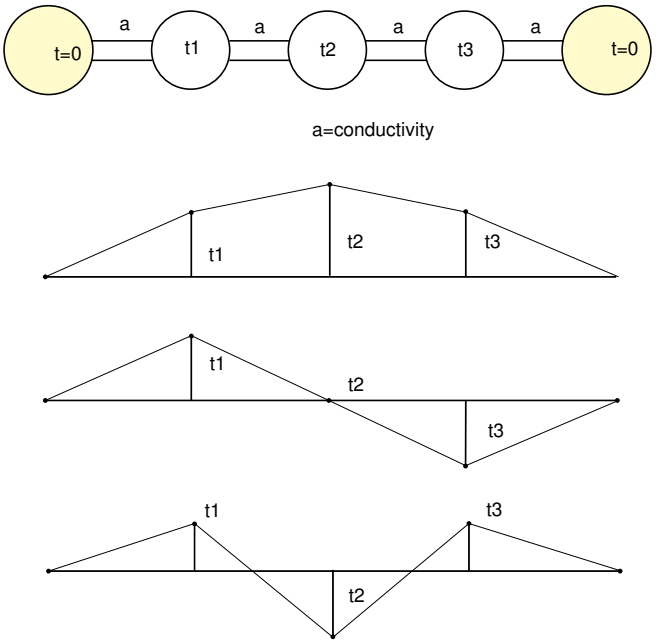

Figure 4.5: A simple first order system and it eigensolutions.

3. Show that, as a consequence of orthogonality, no two eigenvectors can have the same sign patter of its coordinates. See figure 4.5.

If you don't know what a capacitor is and how it works, please read the relevant paragraph in Section **??**.

Problem 4.9. Consider a chain of resistors R and capacitors C shown in figure 4.6. The capacitors start with some initial charge. the evolution of the vector $\mathbf{x} = (q_1, q_2, q_3)$ As the charge trickles through the resistors, $\mathbf{x}$ changes. Derive the system of ODEs governing this change. Show that if $q + q_1 + q_2 + q_3 = 0$ at $t = 0$, then the same is true for all t. In other words, the plane $q + q_1 + q_2 + q_3 = 0$ is invariant under the flow of $\dot{x} = Ax$. What property of the coefficient matrix A implies this property?

Hint. (1) Only three charges are independent, since $q_1 + q_2 + q_3 + q = 0$. (2) The general solution $\mathbf{x}(t)$ is a combination of vectors with exponentially decaying coefficients $e^{\lambda_k t}$. For large t, the one whose λ is closest to 0 dominates.

Problem 4.10. Solve the equation $\ddot{x} + a\dot{x} + bx = 0$ by writing it as a system and using eigenvalues and the eigenvectors.

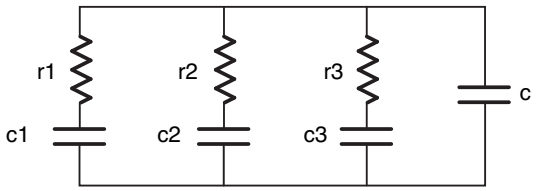

Figure 4.6: The capacitors discharge through the resistors.

Problem 4.11. Let $x_1(t),\ x_2(t)$ be two solutions of $\ddot{x} + a\dot{x} + bx = 0$. The Wronskian of these two solutions is defined as $W(x_1, x_2) = \det \begin{pmatrix} x_1 & x_2 \\ \dot{x}_1 & \dot{x}_2 \end{pmatrix} = \dot{x}_1 x_2 - x_1 \dot{x}_2$.

1. Prove: $\frac{d}{dt}W = -aW$.
2. Give a geometrical interpretation of the Wronskian.

Problem 4.12. Find the general solution of the system

$$\begin{cases} \dot{x} = -x + y \\ \dot{y} = -y + z \\ \dot{z} = -z \end{cases}, \tag{4.17}$$

and find the solution with the initial condition $\mathbf{x}(0) = (1, 1, 1)$.

Problem 4.13. Write Eq. (4.5) in matrix form - more precisely, write the solution as a product of a fixed matrix and a time-dependent vector.

Solution. $\mathbf{x}(t) = \left(e^{\lambda_1 t}\mathbf{v}_1, e^{\lambda_2 t}\mathbf{v}_2\right)\mathbf{c} = \left(\mathbf{v}_1, \mathbf{v}_2\right)\begin{pmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \end{pmatrix}$.

Problem 4.14. Prove the superposition principle: the sum of solutions of a linear system is a solution, as is the product of a solution with a scalar. In short, the set of solutions of a linear system is a linear space.

Problem 4.15. Prove that if the eigenvectors $\mathbf{v}_1,\ \mathbf{v}_2$ of A are linearly independent, then *any* solution is expressible in the form (4.5). In other words, show that any initial condition can be satisfied by a proper choice of constants in Eq. (4.5).

Problem 4.16. Find the general solution of the system

$$\dot{x} = 3x + y, \quad \dot{y} = x + 3y \tag{4.18}$$

and sketch the phase portrait.

Problem 4.17. Find

$$e^{\begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix}}.$$

Hint: Observe that $\begin{pmatrix} \alpha & \omega \\ -\omega & \alpha \end{pmatrix} = \alpha I + \omega J$, where $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, and that I and J commute.

Problem 4.18. Consider the system of three masses on springs, as shown in figure. Can all three normal modes have the property that all coordinates x_i have the same sign? Hint: the eigenvectors of A form an orthogonal frame in $\mathbf{R}^3$.

Hint: use orthogonality of the eigenvectors.

Problem 4.19. A matrix A is called **positive definite** if for any two nonzero vectors the dot product $(A\mathbf{x}, \mathbf{x}) > 0$ for any nonzero vector $\mathbf{x}$, and if A is symmetric. Prove:

1. all eigenvalues of a positive definite matrix are positive. What is the corresponding statement for a negative definite matrix?

2. the eigenvectors of a positive definite matrix are orthogonal, provided the eigenvalues corresponding to the two eigenvectors are distinct. Hint. Use the fact that if A is symmetric, then for any vectors $\mathbf{x}$ and $\mathbf{y}$ the following dot products are equal: (A $\mathbf{x}$, $\mathbf{y}$)=$(A\mathbf{y}, \mathbf{x})$.

Problem 4.20. Assume that both eigenvectors of a 2×2 matrix A are negative. Let $x = x(t)$ be a solution of $\dot{x} = Ax$. True or false: (a) $|x(t)|$ is monotone decreasing function of t. (b) There exists a constant $a > 0$ such that $|x(t)| \leq e^{-at}$ for all t.

Problem 4.21. Assume that $\det A = 1$. Show:

1. If A has real independent eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ with the eigenvalues λ_1, λ_2, then A leave a family of hyperbolas invariant, and express the equations of these hyperbolas in the frame of the eigenvectors.

2. If A has complex eigenvectors $\mathbf{v}_{1,2} = \mathbf{u} \pm i\mathbf{w}$, then A leaves invariant a family of ellipses, and find the equation of these ellipses in the frame $\mathbf{u}, \mathbf{w}$.

Problem 4.22. Assume that the spectrum of the matrix A lies in the left half-plane*. Consider the function $V(x) = \int_0^\infty (e^{At}x, e^{At}x)dt$, where $(\cdot,\cdot)$ stands for the dot product.

1. Prove that $V(x)$ decreases monotonically along the solutions of $\dot{x} = Ax$.

2. Show that V is a quadratic form in x: $V(x) = (Qx, x)$ and find the matrix Q.

3. Show that Q solves the matrix equation $A^TQ + QA = -I$.

Problem 4.23. Assume that all eigenvalues of A are in the left half-plane. Let $L(x)$ be the length of the curve $\{e^{At}\}$, $t \geq 0$. Write $L(x)$ as an integral and prove that W is a Lyapunov function.

Solution. 1. L well -defined (i.e. finite): $L(x) = \int_0^\infty |\frac{d}{ds}e^{As}x|ds = \int_0^\infty |Ae^{As}x|ds$. This integral is finite since all the entries of the matrix e^{As} contain exponentials with negative exponents, according to the earlier analysis of solutions. 2. We have

$$\frac{d}{dt}L(e^{At})| = \frac{d}{dt}\int_0^\infty |Ae^{As}e^{At}x|ds = \frac{d}{dt}\int_0^\infty |Ae^{A(s+t)}x|ds =$$

$$\frac{d}{dt}\int_t^\infty |Ae^{A\tau}x|d\tau = -|Ax|.$$

Problem 4.24. Assume that all the eigenvalues of the $n \times n$ matrix A have negative real parts (this does not preclude real negative eigenvalues). Prove that all solutions of $\dot{\mathbf{x}} = A\mathbf{x}$ approach zero as $t \to \infty$.

Problem 4.25. Under the assumptions of the preceding problem, is it true that $|\mathbf{x}(t)|$ is monotonically decreasing? Give a geometrical explanation of the answer.

Problem 4.26. Let A be a 2×2 matrix with eigenvalues $\lambda_1 \neq \lambda_2$. Show that

$$e^{tA} = e^{\lambda_1 t}C_1 + e^{\lambda_2 t}C_2 \tag{4.19}$$

with some constant matrices $C_1,\ C_2$.

*in other words, every eigenvalue is either negative or complex with a negative real part.

Hint. Method 1. Let T be the diagonalizing matrix: $A = T\Lambda T^{-1}$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$, and observe that $e^{tA} = Te^{tL}T^{-1}$.

Method 2 (works even if A is not diagonalizable; in that case instead of the exponentials in (4.19) one must put two independent solutions of the ODE given by the characteristic polynomial of A, namely, $e^{\lambda t}$ and $te^{\lambda t}$). Let $p(t) = \det(A - \lambda I)$ be the characteristic polynomial of A. By Hamilton-Cayley's theorem, $p(A) = 0$. Note that $\frac{d}{dt}e^{tA} = Ae^{tA}$, i.e. applying the derivative by amounts to mulitplying by A. Thus applying a polynomial in the derivative amounts to multiplying by that polynomial in A, i.e.

$$p\left(\frac{d}{dt}\right)e^{tA} = p(A)e^{tA} = 0.$$

This means that $p\left(\frac{d}{dt}\right)b_{ij}(t) = 0$ for all the entries $b_{ij}(t)$ of e^{tA}. In other words, these entries are solutions of the linear ODE. Since the characteristic roots of this ODE are $\lambda_1,\ \lambda_2$, we conclude that each $b_{ij}(t) = c_{ij}e^{\lambda_1 t} + c'_{ij}e^{\lambda_2 t}$.

Problem 4.27. Find e^{tA}, where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, in terms of A and its eigenvalues $\lambda_1 \neq \lambda_2$.

Hint: use (4.19).

Problem 4.28. Using method 2 in Problem 4.26, adopted to the case of non–diagonalizable matrix, find e^{tA} with

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Chapter 5

Nonlinear Dynamical Systems in the Plane

This section discusses autonomous systems in the plane:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^2, \tag{5.1}$$

a class much richer than the 1D systems $\dot{x} = f(t, x)$. The salient points of this chapter are the following:

1. Qualitative picture near equilibrium.

2. Key representative examples.

3. Two opposite ends of the spectrum: Hamiltonian systems arising in frictionless mechanics and gradient systems, arising in systems with strong dissipation.

Systems (5.1) in 2D exhibit phenomena, such as self–sustained oscillations, or limit cycles, that don't come up in 1D. As we mentioned in the introduction, the class (5.1) also includes scalar second order ODEs, such as the mass–spring systems, again too rich to be described by 1D systems.

We do not discuss the non–autonomous systems $\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x})$ in this section: allowing $\mathbf{f}$ to depend on t introduces new complexity not seen in (5.1) in principle, namely the deterministic chaos.

As mentioned before, (5.1) cannot be solved in elementary functions in general, and even if it could be, an page–long solution could look worse than the problem. Such solution could tell us more than we want to know. Instead, a *qualitative* understanding of the behavior could be : do solutions

approach an equilibrium? or an oscillatory motion (whatever that means for a vector $\mathbf{x}(t)$)? or do they split into different classes, each doing a different thing? The explicit solutions are overrated: there is more variety in (5.1) than these solutions can capture.

5.1 Linearization at an equilibrium point

5.2 Linearized Equations.

Consider a system of ODEs

$$\begin{cases} \dot{x} = f(x, y) \\ \dot{y} = g(x, y) \end{cases} \tag{5.2}$$

where f, g need not be linear functions. Imagine now traveling along some solution (x, y) of (5.2) and watching nearby solutions. In other words, *let's understand the evolution of the difference between two nearby solutions with time.* This difference, which we denote by (u, v), is governed (approximately) by a linear system of ODEs (5.6) given below.

Let $(x(t), y(t))$ and $(x_1(t), y_1(t))$ be two solutions of the system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$:

$$\begin{cases} \dot{x} = f(x, y), \quad \dot{x}_1 = f(x_1, y_1) \\ \dot{y} = g(x, y), \quad \dot{y}_1 = g(x_1, y_1) \end{cases} \tag{5.3}$$

We are interested in the differences

$$u = x_1 - x, \ v = y_1 - y.$$

Subtracting the first equation from the second in each row in (5.3) we obtain

$$\begin{cases} \dot{u} = f(x_1, y_1) - f(x, y) \\ \dot{v} = g(x_1, y_1) - g(x, y), \end{cases} \tag{5.4}$$

By Taylor's formula

$$\begin{cases} f(x_1, y_1) - f(x, y) = f_x u + f_y v + \dots, \\ g(x_1, y_1) - g(x, y) = g_x u + g_y v + \dots, \end{cases} \tag{5.5}$$

where $\ldots = O(u^2 + v^2)$ are higher order terms in u, v, and where $f_x = \frac{\partial f}{\partial x}(x(t), y(t))$, etc. Substituting Eq. (5.5) into Eq. (5.4) and dropping higher-order terms, we obtain

$$\boxed{\begin{cases} \dot{u} = f_x u + f_y v \\ \dot{v} = g_x u + g_y v \end{cases},} \tag{5.6}$$

with the abbreviation $f_x = f_x(x(t), y(t))$, etc. This system (5.6) is referred to as the *linearization of the system (5.2) near the solution* (x, y).

Vector form of (5.6) is

$$\dot{\mathbf{w}} = A(t)\mathbf{w}, \quad A(t) = \mathbf{f}'(\mathbf{x}(\mathbf{t})), \tag{5.7}$$

where

$$\mathbf{w} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{f}'(\mathbf{x}) = \begin{pmatrix} f_x & f_y \\ g_x & g_y \end{pmatrix}.$$

Example. Find the linearization of the system

$$\begin{cases} \dot{x} = -y(1 - x^2 - y^2) \\ \dot{y} = x(1 - x^2 - y^2) \end{cases} \tag{5.8}$$

near (i) the equilibrium solution $(x(t), y(t)) = (0, 0)$, and (ii) near the periodic solution $(x(t), y(t)) = (\cos t, \sin t)$.

Solution. (i) at the equiibrium solution, we find $f_x = 0$, $f_y = -1$, $g_x = 1$, $g_y = 0$, and the linearized system is

$$\dot{u} = -v, \quad \dot{v} = u.$$

(ii) at the periodic solution, the linearized system has form (5.6) with the coefficient matrix

$$A(t) = \begin{pmatrix} 2\cos t \sin t & 2\sin^2 t \\ 2\cos^2 t & -2\cos t \sin t \end{pmatrix} = \begin{pmatrix} \sin 2t & 1 - \cos 2t \\ -1 - \cos 2t & -\sin 2t \end{pmatrix}$$

5.3 Linearization near an equilibrium.

Linearization is a perfect tool for understanding phase portraits near an equilibrium of (5.2). This is because (1) the linearization (5.6) at an equilibrium solution of (5.2) has constant coefficients, and thus is simple to

analyze and (2) the linearized system "looks like" the nonlinear system near an equilibrium solution. The precise sense of "looks like" is this: there exists a smooth transformation which converts the trajectories of the nonlinear system in the vicinity of an equilibrium into the trajectories of the linear system near the origin.

Example 1. Van der Pol's equation $\ddot{x} + (x^2 - 1)\dot{x} + x = 0$. The equivalent system

$$\begin{cases} \dot{x} = y \\ \dot{y} = -x - (1 - x^2)y. \end{cases} \tag{5.9}$$

has a single equilibrium, and this equilibrium is at the origin. Linearization of (5.9) at $(x, y) = (0, 0)$ is

$$\begin{cases} \dot{u} = v \\ \dot{v} = -u - v \end{cases} \tag{5.10}$$

The eigenvalues of the matrix of this linear system are complex with positive real parts. The trajectories of Eq. (5.10) are therefore spirals, spiralling outward. They spiral clockwise. Indeed, according to the equation $\dot{u} = v$ the trajectories cross the positive v-axis to the right.

To summarize, the trajectories of the actual van der Pol equation (5.9) spiral out of the origin clockwise. The equilibrium at the origin is, in other words, an unstable focus.

We leave aside the question of how to analyze the rest of the phase portrait of the van der Pol equation.

Example 2. The damped pendulum $\ddot{x} + \alpha\dot{x} + \sin x = 0$, or written as a system

$$\begin{cases} \dot{x} = y \\ \dot{y} = -\sin x - \alpha y \end{cases} \tag{5.11}$$

This system has two equilibria (modulo 2π): the origin (pendulum hanging) and the point $(\pi, 0)$ (pendulum standing upside-down). We consider the two equilibria separately. Linearized system near $(0, 0)$ is

$$\begin{cases} \dot{u} = v \\ \dot{v} = -u - \alpha v \end{cases} \tag{5.12}$$

This system has a stable focus if $0 < \alpha < 2$, or a stable node if $\alpha \geq 2$. Therefore the equilibrium $(0, 0)$ of the nonlinear system (5.11) is the stable focus (if $0 < \alpha < 2$) or a node (if $\alpha > 2$) as well.

Linearization around the other equilibrium is

$$\begin{cases} \dot{u} = v \\ \dot{v} = u - \alpha v \end{cases} \tag{5.13}$$

The eigenvalues of the corresponding matrix are real and the equilibrium is therefore a saddle.

5.4 Linearization in vector form.

Let φ^t be the flow of the system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}). \tag{5.14}$$

In other words, the solution $\mathbf{x}(t)$ with $\mathbf{x}(0) = \mathbf{x}_0$ is denoted by $\mathbf{x}(t) = \varphi^t \mathbf{x}_0$, i.e.

$$\frac{\partial}{\partial t}\varphi^t \mathbf{x}_0 = f(\varphi^t \mathbf{x}_0), \quad \varphi^0 \mathbf{x}_0 = \mathbf{x}_0. \tag{5.15}$$

We want to derive the (approximate) equation for the propagation of the

x

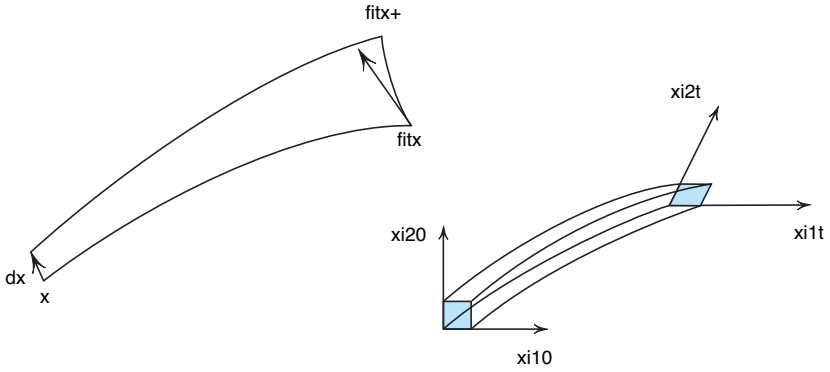

Figure 5.1: Evolution of the difference; evolution of an infinitesimal region.

small difference* $\phi^t(\mathbf{x} + \Delta\mathbf{x}) - \phi^t(\mathbf{x})$ We have by the Taylor approximation

$$\phi^t(\mathbf{x} + \Delta\mathbf{x}) - \phi^t(\mathbf{x}) = \frac{\partial \phi^t(\mathbf{x})}{\partial \mathbf{x}} \Delta\mathbf{x} + o(\Delta\mathbf{x}); \tag{5.16}$$

*We drop the subscript in $\mathbf{x}_0$, writing $\mathbf{x}$ instead.

this shows that the Jacobian matrix $\Phi(t) = \frac{\partial \phi^t(\mathbf{x})}{\partial \mathbf{x}}$ propagates the intinitesimal initial difference from $t = 0$ to t. To determine the evolution of $\Phi(t)$ with time, we just differentiate Eq. (5.15) with respect to the initial condition $\mathbf{x}$:

$$\frac{d}{dt}\frac{\partial \phi^t(\mathbf{x})}{\partial \mathbf{x}} = D\mathbf{f}(\varphi^t\mathbf{x})\frac{\partial \phi^t(\mathbf{x})}{\partial \mathbf{x}}, \tag{5.17}$$

where $D\mathbf{f} \equiv \frac{\partial}{\partial \mathbf{x}}\mathbf{f}$ is the Jacobi derivative matrix of $\mathbf{f}$. We proved

Theorem 5.1. *The matrix $\Phi = \frac{\partial \phi^t(\mathbf{x})}{\partial \mathbf{x}}$ of derivative of solution with respect to initial condition satisfies the linear system*

$$\dot{\Phi} = D\mathbf{f}(\varphi^t\mathbf{x})\Phi. \tag{5.18}$$

According to Liouville's theorem, the determinant of Φ satisfies Abel's equation

$$\frac{d}{dt}(\det \Phi) = (\operatorname{div} \mathbf{f})\ (\det \Phi). \tag{5.19}$$

Note also that $\Phi(0) = \frac{\partial \varphi^0\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{x}} = I$. Thus $\Phi(t)$ is the fundamental solution matrix of the linear ODE

$$\dot{\mathbf{y}} = A(t)\mathbf{y} \quad \text{where} \quad A(t) = D\mathbf{f}(\varphi^t\mathbf{x}) \tag{5.20}$$

This linear system is called the **linearized system**, or the **linearization** of Eq. (5.14) along the solution $\varphi^t\mathbf{x}$. Any solution $\mathbf{y}$ of the linearized equation is given by

$$\mathbf{y}(t) = \Phi(t)\mathbf{y}(0).$$

One can think of $\mathbf{y}(t)$ as the infinitesimal difference between two nearby solutions, as follows from Eq. (5.16).

5.5 Linearization near a periodic solution

. Consider again an *autonomous* system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{5.21}$$

and assume that the system possesses a periodic solution $\mathbf{x} = \mathbf{p}(t)$ of period T: $\mathbf{p}(t+T) = \mathbf{p}(t)$ for all $t \in \mathbb{R}$. Linearization of this system around $\mathbf{p}$ is a system with a periodic coefficient matrix:

$$\dot{\mathbf{u}} = A(t)\mathbf{u}, \quad A(t) = \mathbf{f}'(\mathbf{p}(t)), \tag{5.22}$$

where $\mathbf{f}'$ denotes the Jacobi derivative matrix of $\mathbf{f}$. It turns out that $\lambda = 1$ is one of the Floquet multipliers of this system:

Theorem 5.2. *Consider the linearized system (5.22), obtained by linearizing an autonomous system (5.21) around a periodic solution* $\mathbf{x} = \mathbf{p}(t)$. *One of the Floquet multipliers of the linearized system is* 1, *and the corresponding eigenvalue of this system is* $\mathbf{v} = \dot{\mathbf{p}}(0)$.

Proof. Differentiating the identity

$$\dot{\mathbf{p}}(t) = \mathbf{f}(\mathbf{p}(t))$$

by t, and denoting the velocity $\dot{\mathbf{p}}(t) = \mathbf{v}(t)$, we obtain

$$\dot{\mathbf{v}}(t) = A(t)\mathbf{v}(t), \quad \text{where} \quad A(t) = \mathbf{f}'(\mathbf{p}(t)),$$

concluding that $\mathbf{v}(t)$ is a solution of the linearized system; therefore the fundamental matrix $X(t)$ of (5.22) propagates $\mathbf{v}$ from $t = 0$ to $t = T$:

$$\mathbf{v}(T) = X(T)\mathbf{v}(0). \tag{5.23}$$

But since $\mathbf{p}$ is periodic, so is $\mathbf{v} = \dot{\mathbf{p}}$, and in particular $\mathbf{v}(T) = \mathbf{v}(0)$; then (5.23) becomes

$$X(T)\mathbf{v}(0) = \mathbf{v}(0).$$

This shows that 1 is an eigenvalue of the Floquet matrix, and that $\dot{\mathbf{p}}(0)$ is the corresponding eigenector. ♢

5.6 Twist in the phase plane.

The idea of linearization can be used to answer the following question:

Is the period of oscillations of the pendulum $\ddot{x} + \sin x = 0$ a monotone function of the energy E? This problem was of interest historically, when pendulum clocks were the best means to keep time. Makers of pendulum clocks would have preferred that the period of the pendulum be independent of the amplitude. As we shall see shortly, the period actually grows with amplitude.

In fact, we will deal with a slightly more general equation

$$\ddot{x} + V'(x) = 0, \tag{5.24}$$

which for $V = -\cos x$ reduces to the pendulum equation. For $V = x^4/4 - x^2/2$ Eq. (5.24) is the Duffing eqution $\ddot{x} - x + x^3 = 0$; for $V = x^4/4$ we get $\ddot{x} + x^3 = 0$. Quadratic potential $V = x^2/2$ gives the harmonic oscillator $\ddot{x} + x = 0$. In summary, Eq. (5.24) governs the motion of a particle in a potential $V(x)$. The above question is answered by

Figure 5.2: Some potentials and phase portraits.

figure missing

Theorem 5.3. *If the potential V satisfies*

$$\frac{V'}{x} < V'', \tag{5.25}$$

for all $x \neq 0$ in some interval $[-a, a]$, then the period of solutions of Eq. (5.24) is a monotone decreasing function of the amplitude, for those solutions satisfying $-a < x(t) < a$ for all t. If the opposite of the inequality Eq. (5.25) holds, then the period is a monotone increasing function of the amplitude.

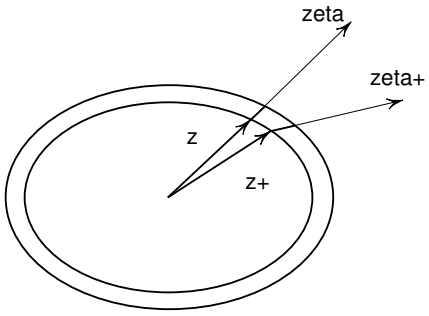

Figure 5.3

For the pendulum we have $V' = \sin x$, and the opposite of Eq. (5.25) holds: $0 < \sin x/x > \cos x$, or $\tan x > x$. We conclude that the period of the pendulum is a monotone decreasing function of the amplitude, for the class of oscillatory solutions. In the cases of superquadratic potentials $V(x) = x^4$, or $V(x) = x^2 + x^4$, or more generally, $V(x) = \sum_{k=0}^{n} a_k x^{2k}, a_k \geq 0$, our criterion Eq. (5.25) shows that the period is a decreasing function of the energy, while for the subquadratic potential $V(x) = x^\alpha$ with $1 < \alpha < 2$ the period increases with the energy.

Proof of Theorem 5.3. Let us rewrite Eq. (5.24) as an equivalent system

$$\begin{cases} \dot{x} = y \\ \dot{y} = -V'(x) \end{cases}, \text{ or } \dot{z} = f(z), \tag{5.26}$$

along with the linearized equation along a solution of Eq. (5.26):

$$\begin{cases} \dot{\xi} = \eta \\ \dot{\eta} = -V''(x)\xi \end{cases}, \text{ or } \dot{\zeta} = f'(z)\zeta. \tag{5.27}$$

We assume that all solutions of Eq. (5.26) in question are periodic and contain the origin, as in the case of all of the examples in the preceding paragraph, with the exception of the Dufing equation.

The key to what comes next is the earlier observation that the linearized solution $\eta(t)$ describes the infinitesimal difference between two nearby solutions of the system Eq. (5.26). Consider now any solution $z(t)$ of Eq. (5.26) along with the associated solution $\zeta(t)$ of the linearized system, with $\zeta(0) \parallel z(0)$. If the vector $\zeta(t)$ turns clockwise faster than $z(t)$ at $t = 0$, then the solutions on larger curves travel with higher angular velocity and thus the period of a solution $z(t)$ is a decreasing function of the amplitude, or of the area enclosed by the orbit in the phase plane. The condition on the angular velocities of z and ζ can be written as

$$\frac{d}{dt}\left(\frac{y}{x}\right) > \frac{d}{dt}\left(\frac{\eta}{\xi}\right), \quad \textbf{whenever} \quad \frac{y}{x} = \frac{\eta}{\xi}.$$

This condition simplifies drastically to give an elegant criterion, as follows. Carrying out the differentiation we obtain

$$\frac{x\dot{y} - \dot{x}y}{x^2} > \frac{\dot{\eta}\xi - \dot{\xi}\eta}{\xi^2} \text{ when } \frac{y}{x} = \frac{\eta}{\xi}.$$

Using Eq. (5.26) and Eq. (5.27) to eliminate the derivatives, and using the collinearity: $\frac{y}{x} = \frac{\eta}{\xi}$, we obtain Eq. (5.25). ♢

5.7 Twist in planar Hamiltonian systems

Consider a Hamiltonian vector field in the plane:

$$\begin{cases} \dot{x} = \ \ H_y(x,y) \\ \dot{y} = -H_x(x,y), \end{cases} \tag{5.28}$$

or $\dot{z} = J\nabla H(z)$, where $z = \mathrm{col}\,(x,y)$, $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $\nabla \equiv$ grad. Assume that the level curves of H are closed, convex and contain the origin, which is the rest point ($\nabla H = 0$).

Then all solutions of (1) are periodic. Denote $T(E)$ the period of the solution on the level curve $H(z) = E$. It is often important to know whether $T(E)$ is monotone.

Theorem 5.4. *In the assumptions of the last paragraph, if* $(H''(z)z, z) < (H'(z), z)$, *i.e.*

$$H_{xx}x^2 + 2H_{xy}xy + H_{yy}y^2 < H_x x + H_y y, \tag{5.29}$$

then $T(E)$ *is a monotone increasing function of* E. *If the opposite of Eq. (5.29) holds, then* $T(E)$ *is a monotone decreasing function.*

The proof is exactly the same as for the preceding theorem.

5.8 Problems

Problem 5.1. Consider the period T of a tumbling solution of the pendulum $\ddot{x} + \sin x = 0$ as a function of the energy $E = \frac{\dot{x}^2}{2} - \cos x$. State and prove the monotonicity property of $T = T(E)$.

Solution. Hint for Method 1. Use the idea of the proof of Theorem 5.3. Hint for Method 2. Observe that for the tumbling solutions $T(E) = \int_0^{2\pi} \frac{dx}{\sqrt{2(E-V(x))}}$, where $V = -\cos x$.

Problem 5.2. Determine the types of all equilibria.

$$\begin{cases} \dot{x} = x(5 - 2x - 3y) \\ \dot{y} = y(3 - x - 2y) \end{cases} \tag{5.30}$$

Problem 5.3. Describe the equilibria in the phase plane of $\ddot{x}+\alpha\dot{x}-x^2+x^4 = 0$

Problem 5.4. Consider the ODE

$$\ddot{x} + \alpha\dot{x} + V'(x) = 0 \tag{5.31}$$

Verify that maxima and minima of V correspond to equilibria in the phase plane. If V has a nondegenerate maximum $x = a^*$, what can be said about the corresponding equilibrium in the phase plane? If V has a minimum, what can be said about the equilibrium?

Solution. Write the equation as the system

$$\begin{cases} \dot{x} = y \\ \dot{y} = -V'(x) - \alpha y \end{cases} \tag{5.32}$$

The point $(a, 0)$ is an equilibrium since $V'(a) = 0$.

Linearized system Eq. (5.6) near the equilibrium in this case is

$$\begin{cases} \dot{u} = v \\ \dot{v} = -V''(a)u - \alpha v \end{cases} \tag{5.33}$$

Note that $V''(a)$ is a constant coefficient.

If $V''(a) < 0$ then the matrix of the linearized system has real eigenvalues (hence the maximum corresponds to saddle). If $V''(a) > 0$, then, depending on α, the eigenvalues are either real negative (the overdamped case) or complex with negative real parts (the underdamped case).

Eq. (5.32) describes the motion of a particle in a potential well with damping, and the statements of the problem are consistent with (and predictable by) physical intuition.

Problem 5.5. Consider a planar autonomous system of ODEs $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ with a periodic solution of period T: $\mathbf{p}(t) = \mathbf{p}(t + T)$ for all t. One of the Floquet multipliers of this periodic solution is 1, according to Theorem 5.2. Prove that the second multiplier is given by

$$\lambda = \exp \int_0^T \operatorname{div} \mathbf{f}(\mathbf{p}(t))\, dt.$$

Problem 5.6. Show that any ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in $\mathbb{R}^n$ can be extended as a Hamiltonian system in $\mathbb{R}^{2n}$.

*here "nondegenerate" means that $V'' \neq 0$.

Problem 5.7. Let D be a bounded domain in $\mathbb{R}^n$, with a smooth boundary, and let $\mathbf{f}$ be a smooth vector field in $\mathbb{R}^n$. Let $D_t = \varphi^t(D)$, where φ^t is the flow associated with $\mathbf{f}$, and let V_t be the (n–dimensional) volume of D_t. Show that

$$\frac{d}{dt}V_t = \int_{\partial D_t} \mathbf{f}\cdot\mathbf{n}\, dS.$$

Solution. We change variables in the first step below via $\mathbf{y} = \varphi^t\mathbf{x}$:

$$\frac{d}{dt}\int_{D_t} dV = \frac{d}{dt}\int_D \det\frac{\partial\varphi^t\mathbf{x}}{\partial\mathbf{x}}\, dV = \int_D \frac{d}{dt}\det\frac{\partial\varphi^t\mathbf{x}}{\partial\mathbf{x}}\, dV.$$

By Abel's theorem (5.19), the last integral can be rewritten as

$$\int_D (\operatorname{div}\mathbf{f})\det\frac{\partial\varphi^t\mathbf{x}}{\partial\mathbf{x}}\, dV = \int_{D_t}(\operatorname{div}\mathbf{f})\, dV;$$

in the last step we changed variables back via $\mathbf{y} = \varphi^n\mathbf{x}$. By the divergence theorem the last integral becomes

$$\int_{\partial D_t}\mathbf{f}\cdot\mathbf{n}\, dS,$$

Q.E.D.

Chapter 6

Index of planar vector fields

References: Strogatz; Coddington-Levinson, page 398.

6.1 Index and its properties

Let γ be a simple closed curve in the plane, and let $\mathbf{f} = (f_1, f_2)$ be a continuous vectorfield in the plane. Let us traverse the curve in the counterclockwise direction, keeping track of the direction of the vector $\mathbf{f}(\mathbf{x})$ in our moving frame. After $\mathbf{x}$ makes a full trip around γ, the vector $\mathbf{f}(\mathbf{x})$ also returns to its original direction, i.e. it makes an integer number of revolutions.[†] More precisely

Definition 6.1. *Index of a vectorfield* $\mathbf{f}$ *on the curve* γ *is the number of turns the vector* $\mathbf{f}(\mathbf{x})$ *makes as the point* $\mathbf{x}$ *traverses* γ *exactly once in a counterclockwise direction. More precisely, we parametrize* γ *by* $\mathbf{x}(t) = (x(t), y(t))$, $0 \leq t \leq 1$, *going counterclockwise. Let* $\theta(t) = arg\,\mathbf{f}$ *be the angle formed with the* x*–axis, and define*

$$i_\gamma(\mathbf{f}) = \frac{1}{2\pi}(\theta(1) - \theta(0)). \tag{6.1}$$

This can be rewritten as

$$i_\gamma(\mathbf{f}) = \frac{1}{2\pi} \oint \frac{f_1 df_2 - f_2 df_1}{f_1^2 + f_2^2} = \frac{1}{2\pi} \int_0^1 \frac{\mathbf{f} \wedge \dot{\mathbf{f}}}{\mathbf{f}^2}\, dt, \tag{6.2}$$

where $\mathbf{f} = \mathbf{f}(\mathbf{x}(t))$, *provided the curve and the vector field are differentiable.*

[†]it is important to note that these revolutions are counted around the origin of our moving frame.

Remarks. *1. The definition only makes sense if* $\mathbf{f} \neq \mathbf{0}$ *on* γ*, since the argument of a zero vector is undefined. 2. We insist on continuity of* $\theta(t)$*; without this continuity assumption* $\theta(t)$ *is multiple valued; with this continuity assumption, (6.1) is defined uniquely, independent of the particular choice of* $\theta(t)$*.*

Figure 6.1 gives examples of various indices.

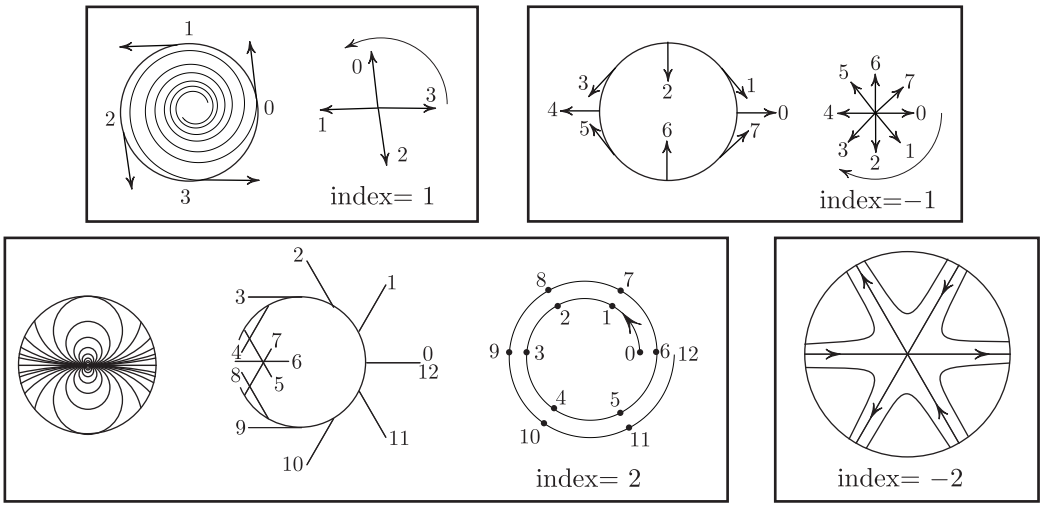

Figure 6.1: Index around a curve surrounding the spiral, the saddle, the dipole and the double saddle.

Theorem 6.1. *If there are no equilibria inside and on a closed non-selfintersecting curve* γ*, then* $i_\gamma(\mathbf{f}) = 0$*.*

The index therefore detects equilibria: a nonzero index implies the presence of equilibria inside γ. Is the converse of this theorem true?

Proof of Theorem 6.1. Referring to Figure 6.2, let us divide the domain D bounded by γ into small subdomains, denoting the boundaries of these by γ_n, $n = 1, 2, \ldots, N$. The domains are chosen to have diameters so small that the argument of $\mathbf{f}$ varies by less than 2π over each γ_k.* Thus

$$i_{\gamma_n}(\mathbf{f}) = 0 \quad \text{for all} \quad n = 1, \ldots, N \tag{6.3}$$

since the change of angle over γ_n is an integer multiple of 2π on the one hand, but is strictly less in absolute value than 2π on the other (by our construction of γ_n), and hence is zero.

*note that we use $\mathbf{f} \neq \mathbf{0}$ and the continuity of $\mathbf{f}$ here.

But

$$i_\gamma(\mathbf{f}) = \sum_{n=1}^{N} i_{\gamma_n}(\mathbf{f}),$$

as Figure 6.2 explains: when adding up the indices over each γ_n, the angle changes over shared edges cancel out, leaving only the angle changes over the unshared arcs of γ_n.

♢

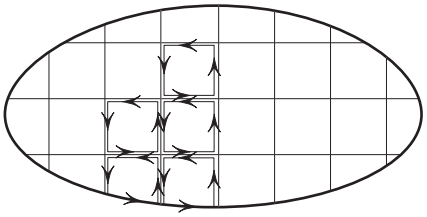

Figure 6.2: Proving that the index is zero if no equilibria inside γ are present.

Theorem 6.2. *Let* $\mathbf{f}$ *be a continuous vector field, and let* γ_A, γ_B *be two simple closed curves such that one can be continuously deformed into the other without passing through the equilibria of* $\mathbf{f}$. *Then*

$$i_{\gamma_A}(\mathbf{f}) = i_{\gamma_B}(\mathbf{f}) \tag{6.4}$$

Proof. Without the loss of generality, let us always parametrize all the curves by $t \in [0,1]$. Let $\gamma_\tau(t)$, $\tau \in [0,1]$ be a deformation mentioned in the statement of the theorem, i.e. a continuous function from $[0,1]\times[0,1] \to \mathbb{R}^2$ such that $\gamma_0(t) = \gamma_A(t)$ and $\gamma_1(t) = \gamma_B(t)$.

Since $\mathbf{f}(\gamma_\tau(t)) \neq \mathbf{0}$ for all $0 \leq t, \tau \leq 1$, the angle $\theta_\tau(t) = \theta(\mathbf{f}(\gamma_\tau(t)))$ is also a continuous function of t, τ; we conclude that the index

$$i_{\gamma_\tau} = \frac{1}{2\pi}(\theta_\tau(1) - \theta_\tau(0))$$

is also continuous in τ. But being also integer, the only way for it to be continuous is to be a constant: $i_{\gamma_0} = i_{\gamma_1}$, Q.E.D. ♢

The last theorem suggests that what really determines γ in $i_\gamma(\mathbf{f})$ is not the particular choice of γ, but rather the equilibria inside γ. This leads to the definition of the index of an equilibrium, as follows.

Definition 6.2. *The index of an isolated equilibrium point* P *of a vector field is the index of the vector field over a simple closed curve* γ *containing* P *and no other equilibrium points in its interior.*

For this definition to make sense we must show that the choice of γ in the definition does not matter, i.e. that for any two curves γ_0, γ_1 surrounding P and enclosing no other equilibria,

$$i_{\gamma_0}(\mathbf{f}) = i_{\gamma_1}(\mathbf{f}).$$

One way to show this is to argue that one curve can be deformed into the other without passing through equilibria. Rather than proving the existence of such a deformation, let us use a different way: surround the equilibrium inside γ_0 by a circle that lies entirely inside γ_0, Figure 6.3, and form a compound closed path as shown in the figure. The shaded region Γ (bounded by γ, the circle and the two segments) encloses no equilibria and thus

$$i_\Gamma(\mathbf{f}) = 0, \tag{6.5}$$

according to Threorem 6.1.* The angle changes over the back–and–forth trip along the cut cancel, and (6.5) becomes

$$i_\gamma(\mathbf{f}) - i_{\gamma_C}(\mathbf{f}) = 0 \tag{6.6}$$

Returning now to two curves $\gamma_0,\ \gamma_1$ surrounding the same equilibrium and no others, we pick the circle lying inside both curves; by (6.6) $i_{\gamma_k}(\mathbf{f}) = i_{\gamma_C}(\mathbf{f})$, $k = 0,\ 1.$ e

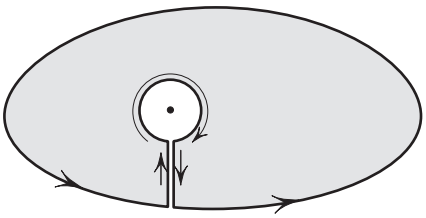

Figure 6.3: The index over γ equal the index over a circle.

Theorem 6.3. *The index of a vectorfield on a simple closed curve equals the sum of indices of the equilibrium points inside that curve.*

Proof is illustrated by Figure 6.4 and is left as an exercise.

*To make the path in Figure 6.3 non-selfintersecting, as required by Theorem 6.1, we can spread the two lines by a small distance ε, and then take the limit of (6.5) as $\varepsilon \downarrow 0$.

.

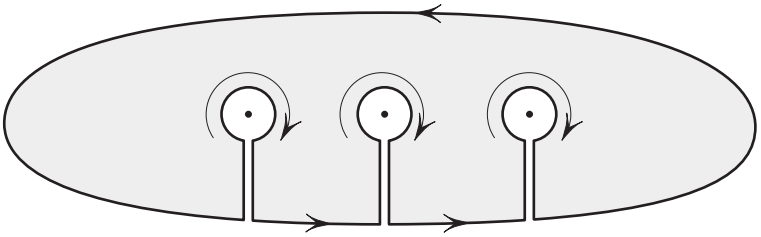

Figure 6.4: Proof of Theorem 6.3.

6.2 Index over a periodic orbit of a vector field

Theorem 6.4. *The index of the vectorfield tangent to a simple closed curve (and not vanishing on this curve) equals 1.*

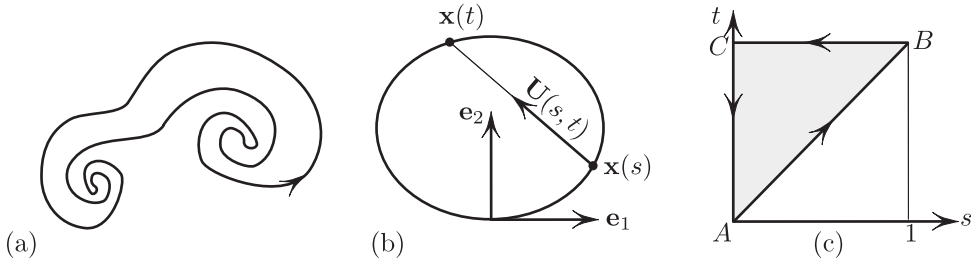

Figure 6.5: Proof of Theorem 6.4.

Proof. Although the statement is near–obvious for a convex curve, a glance at the twisted curve in Figure 6.5 makes it less clear how to prove the statement. The following proof is from the ODE book of Coddington and Levinson; I don't know who invented this beautiful proof.

Let $\mathbf{x} = \mathbf{x}(t), \quad 0 \leq t \leq 1$ be a parametrization of the curve, which we position so that it lies in the upper half-plane and is tangent to the x–axis at the origin, Figure 6.5(b). For an (s,t)–chord, consider its *unit direction vector*:

$$\mathbf{U}(s,t) = \frac{\mathbf{x}(t) - \mathbf{x}(s)}{|\mathbf{x}(t) - \mathbf{x}(s)|}, \quad 0 \leq s < t \leq 1. \tag{6.7}$$

The vector U is still undefined when $\mathbf{x}(s) = \mathbf{x}(t)$, i.e. when $s = t$ and when $(s,t) = (0,1)$. We extend the above definition to both of these cases by continuity, setting

$$\mathbf{U}(t,t) = \lim_{s \uparrow t} \mathbf{U}(s,t)$$

– this is precisely the unit tangent vector in the positive direction, the vector in whose rotation we are interested. And we let

$$\mathbf{U}(0,1) = \lim_{t\uparrow 1} \mathbf{U}(0,t) = -\mathbf{e}_1,$$

the unit vector along the negative x–axis. With this definition $\mathbf{U}$ is a vector field defined on the closed triangle $0 \le s \le t \le 1$ in the (s,t)–plane, Figure 6.5(c), and the key idea is to view $\mathbf{U}(s,t)$ as a vectorfield on this triangle. Note that we started with a vector field defined on an ugly curve, and ended up with a vector field defined on a triangle! The wiggliness of the curve is now hidden in the some wiggliness of the directions of $\mathbf{U}$, but that is easy to deal with, as we shall see next.

Since $|\mathbf{U}| = 1 \neq 0$, we have

$$i_{ABCA}(\mathbf{U}) = 0 \tag{6.8}$$

by Theorem 6.1. We conclude that

$$i_{AB}(\mathbf{U}) = i_{AC}(\mathbf{U}) + i_{CB}(\mathbf{U}); \tag{6.9}$$

we use the same notation for the index even though AB, AC and CB are not closed curves. But $i_{AC}(\mathbf{U}) = \pi/2\pi = 1/2$, since the vector $\mathbf{U}(0,t)$ starts with $\mathbf{e}_1$ at $t = 0$ and ends at $-\mathbf{e}_1$ at $t = 1$, while staying in the upper half plane for all $t \in [0,1]$. Similarly, $i_{CB} = 1/2$ since $\mathbf{U}(1,0) = -\mathbf{e}_1$, $\mathbf{U}(1,1) = \mathbf{e}_1$ and $\mathbf{U}(1,s)$ stays in the lower half plane for $s \in [0,1]$. Substituting these results into (6.9) yields $i_{AB}(\mathbf{U}) = 1$, which is the claim of the theorem, since $\mathbf{U}(t,t)$ is the tangent vector to the curve. ♢

6.3 The Bohl–Brower fixed point theorem

Theorem 6.5. *Any continuous map $\mathbf{x} \mapsto \phi(\mathbf{x})$ from a disk D into itself has a fixed point.*

Proof. Without the loss of generality, let D be the disk $x^2 + y^2 \le 1$. Consider the displacement vector $\mathbf{f}(x) = \phi(\mathbf{x}) - \mathbf{x}$. But $\mathbf{f}$ is a vector field on D, and our goal is to show that there exists an equilibrium point $\mathbf{x}_0 \in D$, i.e. the point for which $\mathbf{f}(\mathbf{x}_0) = 0$. Assume the contrary: $\mathbf{f}(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in D$. The index $i_C\mathbf{f}$ over the boundary circle C is well defined. On the one hand, $\mathbf{i}_C\mathbf{f} = 0$ since there are no equilibria inside C (Theorem 6.1). On the other hand, on the boundary C, the vector $\mathbf{f}$ points into the disk, i.e. always lies to the "inner" side of the tangent to the circle. Since the tangent makes one

full revolution, as the tangency point does, we conclude the same about $\mathbf{f}$, which shows that $\mathbf{i}_C\mathbf{f} = 1$. The contradiction proves that the assumption $\mathbf{f} \neq 0$ was wrong. ♢

6.4 The fundamental theorem of algebra

Theorem 6.6. *Any polynomial has at least one root in the complex plane. (The existence of n roots is then a simple consequence.)*

Proof (an outline). The polynomial $P(z) = z^n + a_1 z^{n-1} + \dots a_n$ with complex z (and with possibly complex coefficients) can be interpreted as a vector field in the plane, and the goal is to show that this vector field has an equilibrium. The main idea is to observe that the index of this over a very large circle is determined by the leading term z^n, since this term exceeds all the others combined and thus dictates the number of turns made. More formally, there exists R so large that

$$R^n > |a_1|R^{n-1} + \dots + |a_n|. \tag{6.10}$$

The existence of such R follows from the fact that the ratio of the right to left sides in (6.10) approaches 0 (and hence is less than 1 for some R). The idea now is to deform $P(z)$ into a simpler polynomial z^n, by a continuous deformation given by

$$P_s(z) = z^n + s(a_1 z^{n-1} + \dots + a_n), \tag{6.11}$$

with s going from 1 to 0. By (6.10) the vectorfield P_s never vanishes on $|z| = R$, and thus the index remains constant in s. But $i_{|z|=R}(z^n) = n$, and thus $i_{|z|=R}(P(z)) = n \neq 0$. This implies that there exists at least one root of P. In fact, we showed that the sum of indices of all equilibria is n. ♢

6.5 Trying to comb a sphere

Theorem 6.7. *Any continuous vector field on the sphere (i.e. a function which attaches to each point on the sphere a tangent vector at that point) has at least one critical point.*

Proof. Recall the definition of the stereographic projection, Figure 6.6: a point N on the sphere is singled out, and any other a is mapped to the point of intersection of the line Na with the tangent plane at S, the antipode of N. In other words, A is the shadow of a with the source of light at N. If

a vector field is defined on the sphere, i.e. if each point has a velocity, the shadows' velocities are thereby defined as well, and so a vector field on the sphere defines a vector field in the plane. And the equilibrium of one vector field corresponds to an equilibrium of the other. This last remark will reduce the problem of proving existence of equilibria on the sphere to the problem in the plane.

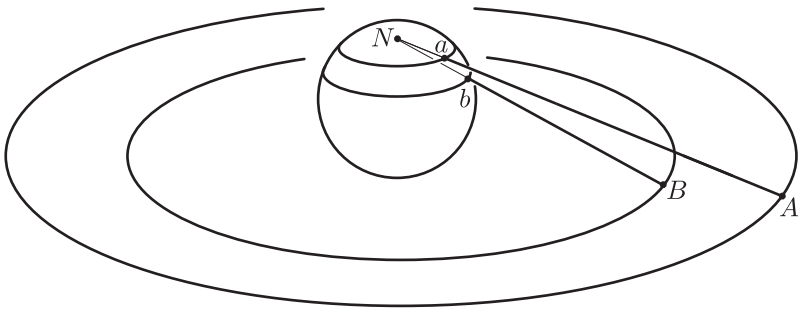

Figure 6.6: Stereorgaphic projection reflects the direction of the field around the tangent to the parallel: a is "inside" the b–circle, but A is "outside" the B–circle.

Let N be a nonsingular point on the sphere (if such doesn't exist, then we have an everywhere zero vector field and there is nothing left to prove). Treat N as the north pole, and surround it with a parallel.

Speaking intuitively, the stereographic projection flips the band between two parallels in Figure 6.6 inside out, i.e. the inner circle (as viewed by the observer on the North pole) maps to the outer circle after projection.* Putting it more formally, if a bug crawling on the sphere happens to cross a latitude circle southwards, then his shadow will cross the shadow of that circle *inwards*. This fact, along with the fact that the index around the parallel on the sphere is zero, allows one to conclude that the index of the projected "shadow" vector field is 2, as Figure 6.7) suggests. I omit the details of the proof (the idea is to first count full turns of the vector field in question *relative to the tangent vector to the circle*; the answer turns out to be 1. But the tangent itself makes 1 turn, so the true number of flips is $1 + 1 = 2$.) ♢

*This is only from the point of view of N: from the point of view of the South pole, no inversion happens: for the South Pole observer, both b and B lie on the inner rings.

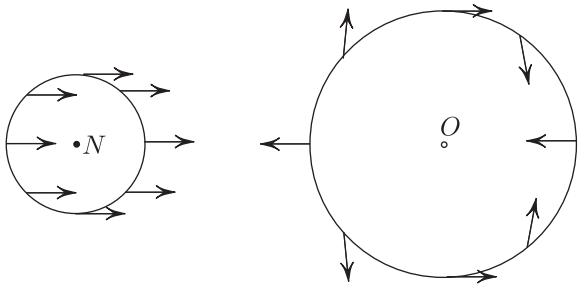

Figure 6.7: Velocities of the bugs on the sphere (left) and the projections (of these velocities) to the plane (right).

6.6 Problems

Problem 6.1. Consider two vector fields $\mathbf{f}$ and $\mathbf{g}$ in $\mathbb{R}^2$. Show that if the angle $\angle(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x})) < \pi$ for all $\mathbf{x}$ on a simple closed curve γ, then

$$i_\gamma \mathbf{f} = i_\gamma \mathbf{g}$$

Problem 6.2. Consider two vector fields $\mathbf{f}$, $\mathbf{g}$ satisfying $|\mathbf{g}(\mathbf{x})| < |\mathbf{f}(\mathbf{x})|$ for all $\mathbf{x}$ on a simple closed curve γ. Show that

$$i_\gamma \mathbf{f} = i_\gamma (\mathbf{f} + \mathbf{g})$$

Problem 6.3. Let γ be a closed orbit on an ODE $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in $\mathbb{R}^2$. Show that a closed orbit of an autonomous ODE in $\mathbb{R}^2$ cannot enclose a saddle equilibrium and no others.