DYNAMICAL SYSTEMS AND NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS

J. Banasiak
School of Mathematical Sciences
University of KwaZulu-Natal, Durban, South Africa

Contents

Chapter 1

Solvability of ordinary differential equations

1 What are differential equations?

In an algebraic equations, like linear or quadratic equations, the unknown quantity is a number (or a collection of numbers) and the equation expresses relations between these numbers and certain numerical coefficients arising in the problem. If the data appearing in the problem are variable, then the unknown will be a function and in the modelling usually we have to balance small increments of this function and the data of the problem. The result typically in an equation involving the derivatives of the unknown function. Such an equation is called a differential equation.

Differential equations are divided into several classes. The main two classes are ordinary differential equations (ODEs) and partial differential equations (PDEs). This course is devoted to PDEs but during the first few lectures we shall recall basic facts concerning ODEs.

As suggested by the name, ODEs are equations where the unknown function is a function of one variable and the derivatives involved in the equation are ordinary derivatives of this function. Symbolically, the general form of ODE is

$$F(y^{(n)}, y^{(n-1)}, \dots y', y, t) = 0,$$

where F is a given function of $n+2$ variables. To solve this ODE means to find an n-times continuously differentiable function $y(t)$ such that for any t (from some interval)

$$F(y^{(n)}(t), y^{(n-1)}(t), \dots y'(t), y(t), t) = 0.$$

Ordinary differential equations describe processes that depend only on one independent variable. For example, let us consider the evolution of the temperature of a body. The so-called Newton's Law of Cooling mathematically can be expressed by the linear first order ordinary differential equation

$$\frac{du}{dt} = k(u - u_s) \tag{1.1.1}$$

where $t \to u(t)$ is the temperature of an object at the time t, u_s is the temperature of the surrounding medium, and k is the proportionality constant. Thus, the temperature was assumed to be a function of only one variable: time t. A short reflection shows that this model in general cannot be correct. Typically, in any object there are regions of lower and higher temperature, so the model described by (1.1.1) could be correct under very special physical assumption: that the temperature is uniform throughout the body.

In the real life we can hardly expect such a uniformity – most functions depend on 4 variables: the time and three spatial coordinates. For example, the temperature in the room changes in time, but also changes

when we move from one point to another. It means that to describe properly the temperature we have to use function of four variables. As we shall see later, the time and position dependence of the temperature will introduce partial derivatives into the equation governing the temperature and make the equation a partial differential equation.

This is the key identifying property of partial differential equations: there are more than one independent variable $x_1, x_2, ...,$. We have also one dependent variable, that is, unknown function $u(x_1, x_2, \ldots)$. In general, it is also possible to have more than one unknown function, which leads to a system of PDEs. A PDE is an identity which relates the independent variables, dependent variable u and the partial derivatives of u. For example

$$F(x_1, x_2, u(x_1, x_2), u_{x_1}(x_1, x_2), u_{x_2}(x_1, x_2)) = 0,$$

where $F : \mathbb{R}^5 \to \mathbb{R}$ is any function, is the most general PDE in two independent variables of *first order* (the highest order of partial derivatives).

Similarly, we can write the most general form of second order PDE in three variables as

$$F(x_1, x_2, x_3, u, u_{x_1}, u_{x_2}, u_{x_3} u_{x_1 x_1}, u_{x_1 x_2}, u_{x_1 x_3}, u_{x_2 x_3}, u_{x_2 x_2}, u_{x_3 x_3}) = 0$$

and so on.

Some examples of DEs (in various variants of notation) are:

1. $\rho_t + \mathbf{a}\nabla\rho = 0$ (continuity equation)
2. $u_t - \Delta u = 0$ (heat or diffusion equation)
3. $u_{tt} - \Delta u = 0$ (waves of small amplitude)
4. $\Delta u = 0$ (electric potential)
5. $u_t = -i\Delta u$ (Schrödinger equation of quantum mechanics)
6. $u_t + u \cdot u_x + u_{xxx} = 0$ (dispersive wave equation)
7. $u_{tt} + u_{xxxx} = 0$ (vibrating beam equation)
8. $u' = au$ (equation of exponential growth)
9. $u' = u(N - u)$ (logistic equation)
10. $u_t = u_{xx} + u(1 - u)$ (Fisher equation).

We shall see how to derive some of these equations from physical principles; some of them require, however, a high degree of sophistication and we will leave them aside.

First of all let us note that, contrary to the general notation, we shall distinguished one variable t. This is dictated by the physical interpretation of the quantities appearing in the equations. Later it will become also clear that the properties of the equations and their solutions with respect to this variable are different from those with respect to the remaining variables reflecting the physical fact that the time is a little different variable from the spatial variables.

Let us look at these equations and try to find some common features. Equations 8. and 9. are first order ordinary differential equations whereas the remaining equations are partial differential equations. Equations 6. and 7. are equations in two variables, the remaining ones can have any number of independent variables as the general notation for the gradient

$$\nabla u = (u_1, \cdots, u_n)$$

and the Laplacian

$$\Delta u = \sum_{n=i}^{n} u_{x_i x_i}$$

does not specify the number of variables.

Equation 1. is a first order equations, Equations 2. through 5. are second order equations, Equation 6. is a third order and Equation 7. is a fourth order equation.

One of the most important property of a differential equation is *linearity*. A rough definition of a linear differential equation is that the function and its derivatives appear in the equation only at most in first (algebraic) power, that is, linearly.

A more sophisticated definition involves some linear algebra terminology. We recall that the set of (continuous, differentiable) functions is a linear space, that is, a sums and scalar multiples of such functions are again functions having the same properties. In such a space one can define operators, like super-functions, which take functions and transform them into other functions. Such an operator can be defined by e.g.

$$L[u] = u_{x_1} + u_{x_2},$$

that is, L takes any differentiable function of two variables, calculates both first partial derivatives and adds them together. This operator is a linear operator as

$$L[c_1u_1 + c_2u_2] = c_1L[u_1] + c_2L[u_2]$$

for any scalars c_1, c_2 and functions u_1 and u_2 (for which it is defined). On the other hand, the operator related to Equation 6. is not linear: if we take the function $u = x$, then

$$L[u] = x$$

but applied to the function $2u$ we have

$$L[2u] = 4x \neq 2L[u].$$

Using the concept of an operator we can write any differential equation in the form

$$L[u] = f,$$

by grouping all the terms containing the unknown function and/or its derivatives on the left-hand side. We then say that the differential equation is linear if L is a linear operator in u. We also say that the linear equation is homogeneous if $f \equiv 0$; otherwise the equation is said to be non-homogeneous.

The advantage of linearity for the equation $L[u] = 0$ is that if u and v are solution of this equation, then so is $u + v$. More generally, any linear combination of solutions is itself a solution; this is sometimes called the *superposition principle*. This is principle is **not** valid for nonlinear equations.

Using this definition we observe that equations 1.–5., 7. and 8. are linear whereas 6. 9. and 10. are nonlinear.

2 Cauchy problem for ordinary differential equations

In this section we shall be concerned with *first order* ordinary differential equations which are solved with respect to the derivative of the unknown function, that is, with equations which can be written as

$$y' = f(t, y), \tag{1.2.1}$$

where f is a given function of two variables.

Several comments are in place here. Firstly, even though in such a simplified form, equation (1.2.1) in general has no closed form solution, that is, it is impossible to write the solution in the form

$$y(t) = combination\ of\ elementary\ functions\ like \sin t, \cos t, \ln t, polynomials...$$

Example 2.1 A trivial example is the equation

$$y' = e^{-t^2}.$$

We know that the solution must be

$$y(t) = \int e^{-t^2} dt$$

but, on the other hand, it is known that this integral cannot be expressed as a combination of elementary functions.

If a solution to a given equation can be written in terms of integrals of elementary functions (as above), then we say that the equation is *solvable in quadratures*. Since we know that every continuous function has an antiderivative (though often we cannot find this antiderivative explicitly), it is almost as good as finding the explicit solution to the equation. However, there are many instances when we cannot solve an equation even in quadratures. How do we know then that the equation has a solution? The answer is that if the right hand side of the equation, that is the function f, is continuous, then there is at least one solution to (1.2.1). This result is called the Peano Theorem and involves some more advanced calculus. Thus, we can safely talk about solutions to ODEs of the form (1.2.1) even without knowing their explicit formulae.

Another important problem is related to the uniqueness of solutions, that is, whether there is only one solution to a given ODE. A quick reflection shows that clearly not: for the simplest equation

$$y' = 0,$$

the solutions are

$$y(t) = C,$$

where C is an arbitrary constant; thus there are infinitely many solutions. The uniqueness question, however, hasn't been properly posed. In fact, what we are looking for is usually a solution passing through a specified point.

Example 2.2 Assume that a point is moving along the horizontal line with speed given by $v(t) = t$. Find the position of the point at $t = 5$. To solve this problem let us recall that $v(t) = \frac{ds}{dt}$ where s is the distance travelled. Thus the problem results in an equation of the type discussed above:

$$v(t) = \frac{ds}{dt} = t$$

and

$$s(t) = 0.5t^2 + C$$

where C is an arbitrary constant. Hence $s(5) = 12.5 + C$ and there is no proper answer. In this physical setting the original question is clearly wrongly posed. What we need to give the proper answer is the information about the position of the point at some other time t, say, $t = 1$. If we know that (with respect to a fixed origin) $s(1) = 2$, then also $s(1) = 0.5 + C$ and $C = 1.5$. Therefore $s(5) = 12.5 + 1.5 = 14$.

From this example (and from physical or other considerations) it follows that if we are interested in getting a unique answer, we not only need the equation (which reflects usually some natural law) but also the state of the system (that is, the value of the solution) at some specified point. Thus, the complete *Cauchy* or *initial value* problem would be to solve

$$\begin{aligned} y' &= f(t,y), \quad \text{for all} \quad t \in [t_1, t_2] \\ y(t_0) &= y_0, \quad \text{for some } t_0 \in [t_1, t_2]. \end{aligned} \tag{1.2.2}$$

Once again we emphasize that to solve (1.2.2) is to find a continuously differentiable function $y(t)$ such that

$$\begin{aligned} y'(t) &= f(t, y(t)) \quad \text{for all} \quad t \in [t_1, t_2] \\ y(t_0) &= y_0, \quad \text{for some } t_0 \in [t_1, t_2]. \end{aligned}$$

Example 2.3 Check that the function $y(t) = \sin t$ is the solution to the problem

$$\begin{aligned} y' &= \sqrt{1 - y^2}, \quad t \in [0, \pi/2] \\ y(\pi/2) &= 1 \end{aligned}$$

Solution. LHS: $y'(t) = \cos t$, RHS: $\sqrt{1 - y^2} = \sqrt{1 - \sin^2 t} = |\cos t| = \cos t$ as $t \in [0, \pi/2]$. Thus the equation is satisfied. Also $\sin \pi/2 = 1$ so the "initial" condition is satisfied.

Note that the function $y(t) = \sin t$ is not a solution to this equation on a larger interval.

Returning to our uniqueness question we ask whether the problem (1.2.2) has always a unique solution. The answer is negative.

Example 2.4 The Cauchy problem

$$\begin{aligned} y' &= \sqrt{y}, \quad t \geq 0 \\ y(0) &= 0, \end{aligned}$$

has at least two solutions: $y \equiv 0$ and $y = \frac{1}{4}t^2$.

However, there is a large class of functions f for which (1.2.2) has exactly one solution. Before we formulate the main result of this section, we must introduce necessary definitions. We say that a function g defined on an interval $[a, b]$ is Lipschitz continuous if there is a constant L such that

$$|g(x_1) - g(x_2)| \leq L|x_1 - x_2|, \tag{1.2.3}$$

for any $x_1, x_2 \in [a, b]$. In particular, if g is continuously differentiable on a closed rectangle $[a, b]$, the derivative g' is bounded there by, say, a constant M and then, by the Mean Value Theorem, we have for $x_1, x_2 \in [a, b]$

$$|g(x_1) - g(x_2)| = |g'(\theta)(x_1 - x_2)| \leq M|x_1 - x_2|, \tag{1.2.4}$$

where θ is a point between x_1 and x_2, so that continuously differentiable functions are Lipschitz continuous. However, e.g., $g(x) = |x|$ is Lipschitz continuous but not continuously differentiable.

Theorem 2.1 *(Picard's theorem) Let f be continuous in the rectangle $R : |t - t_0| \leq a, |y - y_0| \leq b$ for some $a, b > 0$ and satisfy the Lipschitz condition with respect to y uniformly in t: for any $(t, y_1), (t, y_2) \in R$*

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|. \tag{1.2.5}$$

Denote

$$M = \max_{(t,y)\in R} |f(t, y)|$$

and define $\alpha = \min\{a, b/M\}$. Then the initial value problem (1.2.2) has exactly one solution at least on the interval $t_0 - \alpha \leq t \leq t_0 + \alpha$.

Remark 2.1 All continuous functions $f(t, y)$ having continuous partial derivative $\frac{\partial f}{\partial y}$ in some neighbourhood of (t_0, y_0) give rise to (1.2.2) with exactly one solution (at least close to t_0). In fact, if f is continuously differentiable with respect to y on some closed rectangle R so that the derivative is bounded there, say, by a constant M:

$$\left|\frac{\partial f}{\partial y}(t, y)\right| \leq M$$

for all $(t, y) \in R$ which, by (1.2.4), gives Lipschitz continuity w.r.t y

We split the proof of this result into several steps. First we shall prove a general result known as Gronwall's lemma.

Lemma 2.1 *If $f(t)$, $g(t)$ are continuous and nonnegative for $t \in [t_0, t_0 + \alpha]$, $\alpha > 0$, and $c > 0$, then*

$$f(t) \leq c + \int_{t_0}^{t} f(s)g(s)ds \tag{1.2.6}$$

on $[t_0, t_0 + \alpha]$ implies

$$f(t) \leq c \exp\left(\int_{t_0}^{t} g(s)ds\right) \tag{1.2.7}$$

for all $[t_0, t_0 + \alpha]$.

If f satisfies (1.2.6) with $c = 0$, then $f(t) = 0$ on $[t_0, t_0 + \alpha]$.

Proof. Define $F(t) = c + \int_{t_0}^{t} f(s)g(s)ds$ on $[t_0, t_0 + \alpha]$. Then $F(t) \geq f(t)$ and $F(t) > 0$ on this interval. Differentiating, we get $F'(t) = f(t)g(t)$ and therefore

$$\frac{F'(t)}{F(t)} = \frac{f(t)g(t)}{F(t)} \leq \frac{f(t)g(t)}{f(t)} = g(t).$$

However, the left-hand side is equal to $\frac{d}{dt} \ln F(t)$ so that

$$\frac{d}{dt} \ln F(t) \leq g(t),$$

or, integrating

$$\ln F(t) - \ln F(t_0) \leq \int_{t_0}^{t} g(s)ds.$$

Since $F(t_0) = c$, we obtain

$$F(t) \leq c \exp\left(\int_{t_0}^{t} g(s)ds \right)$$

but since, as we observed, $f(t) \leq F(t)$, we have

$$f(t) \leq c \exp\left(\int_{t_0}^{t} g(s)ds \right),$$

which proves the first part. If $c = 0$, then we cannot use the above argument directly, as it would involve taking logarithm of zero. However, if

$$f(t) \leq \int_{t_0}^{t} f(s)g(s)ds$$

then

$$f(t) \leq c + \int_{t_0}^{t} f(s)g(s)ds$$

for any $c > 0$ and so

$$0 \leq f(t) \leq c \exp\left(\int_{t_0}^{t} g(s)ds \right)$$

for any $c > 0$ but this yields $f(t) = 0$ for all t in $[t_0, t_0 + \alpha]$. ■

Gronwall's inequality can be used to show that, under the assumptions of Picard's theorem, there can be at most one solution to the Cauchy problem (1.2.2). Let $y_1(t)$ and $y_2(t)$ be two solutions of the Cauchy problem (1.2.2) on R with the same initial condition y_0, that is $y_1'(t) \equiv f(t, y_1(t))$, $y_1(t_0) = y_0$ and $y_2'(t) \equiv f(t, y_2(t))$, $y_2(t_0) = y_0$. Then $y_1(t_0) - y_2(t_0) = 0$ and

$$y_1'(t) - y_2'(t) = f(t, y_1(t)) - f_2(t, y_2(t)).$$

Integrating and using the condition at t_0 we see that

$$y_1(t) - y_2(t) = \int_{0}^{t} (f(t, y_1(s)) - f_2(t, y_2(s)))ds.$$

Using next (1.2.5) we have

$$\begin{aligned} |y_1(t) - y_2(t)| &= \left| \int_{t_0}^{t} (f(t, y_1(s)) - f_2(t, y_2(s)))ds \right| \leq \int_{t_0}^{t} |f(t, y_1(s)) - f_2(t, y_2(s))|ds \\ &\leq L \int_{t_0}^{t} |y_1(s)) - y_2(s)|ds, \end{aligned} \tag{1.2.8}$$

thus we can use the second part of Gronwall's lemma to claim that $|y_1(t) - y_2(t)| = 0$ or $y_1(t) = y_2(t)$ for all t satisfying $|t - t_0| < a$.

The proof of the existence is much more complicated. Firstly, we convert the Cauchy problem (1.2.2) to an integral equation by integrating both sides of the equation in (1.2.2) and using the initial condition to get

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))ds. \tag{1.2.9}$$

If $y(t)$ is a differentiable solution to (1.2.2), then of course (1.2.9) is satisfied. On the other hand, if $y(t)$ is a continuous solution to (1.2.9), then it is also differentiable, and we see that by differentiating (1.2.9) we obtain the solution of (1.2.2). Thus, we shall concentrate on finding continuous solutions to (1.2.9). The approach is to define the so-called Picard's iterates by

$$\begin{aligned} y_0(t) &= y_0, \\ y_n(t) &= \int_{t_0}^{t} f(s, y_{n-1}(s))ds, \end{aligned} \tag{1.2.10}$$

and proving that they converge to the solution.

As a first step, we shall show that the iterates remain in the rectangle R. Namely, if M, a, b, α are defined as in the formulation of Picard's theorem and y_n is defined as in (1.2.10), then for any n

$$|y_n(t) - y_0| \leq M|t - t_0| \tag{1.2.11}$$

for $|t - t_0| \leq \alpha$. Note that (1.2.11) means that y_n is sandwiched between lines $y_0 \pm M(t - t_0)$ and the wedge created by these lines is always inside the rectangle R if $|t - t_0| < \alpha$.

To prove (1.2.11), we note that for $n = 0$ the estimate (1.2.11) is obvious, so to proceed with induction, we shall assume that it is valid for some $n > 0$ and, taking $n + 1$, we have

$$|y_{n+1}(t) - y_0| = \left| \int_{t_0}^{t} f(s, y_n(s))ds \right| \leq \int_{t_0}^{t} |f(s, y_n(s))|ds.$$

However, by the remark above and the induction assumption, $y_n(s)$ is in R as long as $|t - t_0| \leq \alpha$ and thus $|f(s, y_n(s))| \leq M$. Thus easily

$$|y_{n+1}(t) - y_0| \leq M|t - t_0|.$$

In the next step we shall show that the sequence of Picard's iterates converges. To make things simpler, we shall convert the sequence into a series the convergence of which is easier to establish. To this end we write

$$y_n(t) = y_0 + (y_1(t) - y_0) + (y_2(t) - y_1(t)) + \ldots + (y_n(t) - y_{n-1}(t)) \tag{1.2.12}$$

and try to show that

$$\sum_{n=0}^{\infty} |y_{n+1}(t) - y_n(t)| < +\infty$$

for any $|t-t_0| \leq \alpha$, which would give the convergence of the series.

We use induction again. Assume that $t > t_0$, the analysis for $t < t_0$ being analogous. Firstly, proceeding as in (2.2.5), we observe that for $n > 1$

$$|y_n(t) - y_{n-1}(t)| \leq \int_{t_0}^{t} |f(s, y_{n-1}(s)) - f(s, y_{n-2}(s))| ds \leq L \int_{t_0}^{t} |y_{n-1}(s) - y_{n-2}(s)| ds. \tag{1.2.13}$$

Now, for $n = 1$ we obtain

$$|y_1(t) - y_0| \leq M(t - t_0)$$

and for $n = 2$

$$\begin{aligned} |y_2(t) - y_1(t)| &\leq \int_{t_0}^{t} |f(s, y_1(s)) - f(s, y_0)| ds \leq L \int_{t_0}^{t} |y_1(s) - y_0| ds \\ &\leq LM \int_{t_0}^{t} (s - s_0) ds = \frac{ML}{2}(t - t_0)^2. \end{aligned}$$

This justifies the induction assumption

$$|y_n(t) - y_{n-1}(t)| \leq \frac{ML^{n-1}}{n!}(t - t_0)^n$$

and by (1.2.13)

$$\begin{aligned} |y_{n+1}(t) - y_n(t)| &\leq \int_{t_0}^{t} |f(s, y_n(s)) - f(s, y_{n-1}(s))| ds \leq L \int_{t_0}^{t} |y_n(s) - y_{n-1}(s)| ds \\ &\leq \frac{ML^n}{n!} \int_{t_0}^{t} (s - t_0)^n = \frac{ML^n}{(n+1)!}(t - t_0)^{n+1}. \end{aligned}$$

Now, because $|t - t_0| < \alpha$, we see that

$$|y_n(t) - y_{n-1}(t)| \leq \frac{ML^{n-1}}{n!}\alpha^n$$

so that

$$\sum_{n=0}^{\infty} |y_{n+1}(t) - y_n(t)| \leq \sum_{n=1}^{\infty} \frac{ML^{n-1}}{n!}\alpha^n = \frac{M}{L}(e^{\alpha L} - 1)$$

which is finite. Thus, the sequence $y_n(t)$ converges for any t satisfying $|t - t_0| \leq \alpha$. Let us denote the limit by $y(t)$. By (1.2.12) we obtain

$$y(t) = y_0 + (y_1(t) - y_0) + (y_2(t) - y_1(t)) + \ldots + (y_n(t) - y_{n-1}(t)) + \ldots = y_0 + \sum_{n=0}^{\infty} (y_{n+1}(t) - y_n(t)) \tag{1.2.14}$$

and so

$$\begin{aligned} |y(t) - y_n(t)| &= \left| \sum_{j=n}^{\infty} (y_{j+1}(t) - y_j(t)) \right| \leq M \sum_{j=n}^{\infty} L^j \frac{(t - t_0)^{j+1}}{(j+1)!} \\ &\leq \frac{M}{L} \sum_{j=n}^{\infty} \frac{L^{j+1}\alpha^{j+1}}{(j+1)!} \end{aligned} \tag{1.2.15}$$

where the tail of the series is convergent to zero. It is clear that the left hand side does not depend on t as long as $|t-t_0| \le \alpha$. This fact can be used to show the continuity of the limit function $y(t)$. Firstly, we observe that, by induction, $y_n(t)$ is continuous if $y_{n-1}(t)$ is. In fact, we have

$$\begin{aligned} |y_n(t_1) - y_n(t_2)| &\le \left| \int_{t_0}^{t_1} f(s, y_n(s))ds - \int_{t_0}^{t_2} f(s, y_n(s))ds \right| \le \left| \int_{t_1}^{t_2} f(s, y_n(s))ds \right| \\ &\le M|t_2 - t_1|. \end{aligned}$$

Next, let t_1 and t_2 be arbitrary numbers satisfying $|t_i - t_0| \le \alpha$ for $i = 1, 2$. By (1.2.15) we can find n so large that

$$\frac{M}{L} \sum_{j=n}^{\infty} \frac{L^{j+1}\alpha^{j+1}}{(j+1)!} < \epsilon/3$$

so that

$$|y(t_i) - y_n(t_i)| < \epsilon/3, \qquad i = 1, 2.$$

Since $y_n(t)$ is continuous, we can find $\delta > 0$ such that if $|t_1 - t_2| < \delta$, then

$$|y_n(t_1) - y_n(t_2)| < \epsilon/3.$$

Combining, we see that whenever $|t_1 - t_2| < \delta$ and $|t_i - t_0| < \alpha$ for $i = 1, 2$, we have

$$|y(t_1) - y(t_2)| \le |y(t_1) - y_n(t_1)| + |y_n(t_1) - y_n(t_2)| + |y_n(t_2) - y(t_2)| \le \epsilon,$$

so that $y(t)$ is continuous.

The last step is to prove that the obtained function is indeed the solution of the Cauchy problem in the integral form (1.2.9)

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))ds.$$

Since by construction

$$y_{n+1}(t) = y_0 + \int_{t_0}^{t} f(s, y_n(s))ds$$

we obtain

$$y(t) = \lim_{n\to\infty} y_{n+1}(t) = y_0 + \lim_{n\to\infty} \int_{t_0}^{t} f(s, y_n(s))ds$$

so that we have to prove that

$$\lim_{n\to\infty} \int_{t_0}^{t} f(s, y_n(s))ds = \int_{t_0}^{t} f(s, y(s))ds. \tag{1.2.16}$$

Firstly, note that the right-hand side is well-defined as y is a continuous function, f is continuous so that the composition $f(s, y(s))$ is continuous and the integral is well-defined. Thus, we can write, by (1.2.15),

$$\begin{aligned} \left| \int_{t_0}^{t} f(s, y_n(s))ds - \int_{t_0}^{t} f(s, y(s))ds \right| &\le \int_{t_0}^{t} |f(s, y_n(s)) - f(s, y(s))|ds \le L \int_{t_0}^{t} |y_n(s) - y(s)|ds \\ &\le L\frac{M}{L} \sum_{j=n}^{\infty} \frac{L^{j+1}\alpha^{j+1}}{(j+1)!} \int_{t_0}^{t} ds \le M\alpha \sum_{j=n}^{\infty} \frac{L^{j+1}\alpha^{j+1}}{(j+1)!}. \end{aligned}$$

As before, the sum above approaches zero as $n \to \infty$ and therefore (1.2.16), and the whole theorem, is proved. ■

We illustrate the use of this theorem on several examples.

Example 2.5 We have seen in Example 2.4 that there are two solutions to the problem

$$\begin{aligned} y' &= \sqrt{y}, \quad t \geq 0 \\ y(0) &= 0. \end{aligned}$$

In this case $f(t,y) = \sqrt{y}$ and $f_y = 1/2\sqrt{y}$; obviously f_y is not continuous in any rectangle $|t| \leq a$, $|y| \leq b$ and we may expect troubles.

Another example of nonuniqueness is offered by

$$\begin{aligned} y' &= (\sin 2t) y^{1/3}, \quad t \geq 0 \\ y(0) &= 0, \end{aligned} \tag{1.2.17}$$

Direct substitution shows that we have 3 different solutions to this problem: $y_1 \equiv 0, y_2 = \sqrt{8/27}\sin^3 t$ and $y_3 = -\sqrt{8/27}\sin^3 t$.

Example 2.6 Show that the solution $y(t)$ of the initial value problem

$$\begin{aligned} y' &= t^2 + e^{-y^2}, \\ y(0) &= 0, \end{aligned}$$

exists for $0 \leq t \leq 0.5$, and in this interval, $|y(t)| \leq 1$.

Let R be the rectangle $0 \leq t \leq 0.5, |y| \leq 1$. The function $f(t,y) = t^2 + e^{-y^2}$ is continuous and has continuous derivative f_y. We find

$$M = \max_{(t,y)\in R} |f(t,y)| \leq (1/2)^2 + e^0 = 5/4,$$

thus the solution exists and is unique for

$$0 \leq t \leq \min\{1/2, 5/4\} = 1/2,$$

and of course in this interval $|y(t)| \leq 1$.

Example 2.7 The solution of the initial value problem

$$\begin{aligned} y' &= 1 + y^2, \\ y(0) &= 0, \end{aligned}$$

is given by $y(t) = \tan t$. This solution is defined only for $-\pi/2 < t < \pi/2$. Let us check this equation against the Picard Theorem. We have $f(t,y) = 1 + y^2$ and $f_y(t,y) = 2y$ and both functions are continuous on the whole plane. Let R be the rectangle $|t| \leq a$, $|y| \leq b$, then

$$M = \max_{(t,y)\in R} |f(t,y)| = 1 + b^2,$$

and the solution exists for

$$|t| \leq \alpha = \min\{a, \frac{b}{1+b^2}\}.$$

Since a can be arbitrary, the maximal interval of existence predicted by the Picard Theorem is the maximum of $b/(1+b^2)$ which is equal to $1/2$.

This shows that it may happen that the Picard theorem sometimes does not give the best possible answer - that is why it is sometimes called "the local existence theorem".

Example 2.8 Suppose that $|f(t,y)| \leq K$ in the whole plane $\mathbb{R}^2$. Show that the solution of the initial value problem

$$\begin{aligned} y' &= f(t,y), \\ y(t_0) &= y_0, \end{aligned}$$

where t_0 and y_0 are arbitrary, exists for all $t \in \mathbb{R}$.

Let R be the rectangle $|t - t_0| \leq a$, $|y - y_0| \leq b$ for some a, b. The quantity M is given by

$$M = \max_{(t,y)\in R} |f(t, y)| = K,$$

and the quantity

$$|t - t_0| \leq \alpha = \min\{a, \frac{b}{K}\},$$

can be made as large as we wish by choosing a and b sufficiently large. Thus the solution exists for all t.

To be able to extend the class of functions f for which the solution is defined on the whole real line we must introduce the concept of the continuation of the solution.

Remark 2.2 Picard's theorem gives local uniqueness that is for any point (t_0, y_0) around which the assumptions are satisfied, there is an interval over which there is only one solution of the given Cauchy problem. However, taking a more global view, it is possible that we have two solutions $y_1(t)$ and $y_2(t)$ which coincide over the interval of uniqueness mentioned above but branching for larger times. If we assume that any point of the plane is the uniqueness point, such a scenario is impossible. In fact, if $y_1(t) = y_2(t)$ over some interval I, then by continuity of solutions, there is the largest t, say t_1, having this property. Thus, $y_1(t_1) = y_2(t_1)$ with $y_1(t) \neq y_2(t)$ for some $t > t_1$. Thus, the point $(t_1, y_1(t_1))$ would be the point violating Picard's theorem, contrary to the assumption.

An important consequence of the above is that we can glue solutions together to obtain solution defined on a possibly larger interval. If $y(t)$ is a solution to (1.2.2) defined on an interval $[t_0 - \alpha, t_0 + \alpha]$ and $(t_0 + \alpha, y(t_0 + \alpha))$ is a point around which the assumption of Picard's theorem is satisfied, then there is a solution passing through this point defined on some interval $[t_0 + \alpha - \alpha', t_0 + \alpha + \alpha']$, $\alpha' > 0$. These two solutions coincide on $[t_0 + \alpha - \alpha', t_0 + \alpha]$ and therefore, by the first part, they must coincide over the whole interval of their common existence and therefore constitute a solution of the original Cauchy problem defined at least on $[t_0 - \alpha, t_0 + \alpha + \alpha']$.

Example 2.9 Using such a patchwork technique, global existence can be proved for a larger class of right-hand sides in (1.2.2). Assume that the function f is globally Lipschitz that is

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|$$

for all $t, y_1, y_2 \in \mathbb{R}$, where the constant L is independent of t and y_1, y_2. Let $y(t)$ be the solution passing through (t_0, y_0). It is defined on the interval $|t - t_0| \leq \alpha$ with $\alpha = \min\{a, b/M\}$. Here, a and b can be arbitrarily large as f is defined on the whole $\mathbb{R}^2$. Let as fix $a = 1/(L + 1)$. Lipschitz continuity yields

$$|f(t, y) - f(t, y_0)| \leq L|y - y_0| \leq Lb$$

for $(t, y) \in R$. Thus,

$$|f(t, y)| \leq Lb + |f(t, y_0)| \leq Lb + \max_{t_0-1/(L+1)\leq t\leq t_0+1/(L+1)} |f(t, y_0)| = Lb + m(t_0, y_0)$$

so that

$$M \leq Lb + m(t_0, y_0)$$

and

$$\frac{b}{M} \geq \frac{b}{Lb + m(t_0, y_0)} = \frac{1}{L + m(t_0, y_0)/b}.$$

For any fixed t_0, y_0 we can select b large enough so that $m(t_0, y_0)/b \leq 1$ and thus, for such a b

$$\alpha = \frac{1}{L + 1}.$$

The solution therefore is defined at least on the interval $[t_0-\alpha, t_0+\alpha]$, where $\alpha = \frac{1}{L+1}$, and the length of the interval of existence is **independent of** t_0 **and** y_0. Next we shall use the method of small steps. We take $t_{1,0}$, $t_{1,0} = t_0 + 0.9\alpha$ with corresponding $y_{1,0} = y(t_{1,0})$ as a new Cauchy data. By the above there is a solution of this Cauchy problem that is defined on $[t_{1,0}-\alpha, t_{1,0}+\alpha] = [t_0 - 0.1\alpha, t_0 + 0.9\alpha + \alpha]$ and by uniqueness the two solutions coincide on $[t_0 - 0.1\alpha, t_0 + \alpha]$ and therefore by gluing them together we obtain a the solution of the original Cauchy problem defined on $[t_0 - \alpha, t_0 + 0.9\alpha + \alpha]$. Continuing this way we eventually cover the whole real line with solutions as each time we make a step of constant length.

Note that the crucial rôle here was played by the fact that the numerator and denominator of the fraction b/M grew at the same rate. The procedure described above would be impossible if $f(y)$ grew faster than linearly as $y \to \infty$, like in Example 2.7. There, $b/M = b/(1+b^2)$ and if we enlarge b, then the possible time step will become smaller and there is a possibility that the time steps will sum up to a finite time determining the maximal interval of existence of the solution as in Example 2.7.

Picard's theorem ensures existence for only a bounded interval $|t - t_0| \le \alpha$, where in general α depends on the initial condition (through the rectangle R). In most applications it is important to determine whether the solution exists for all times, as discussed in the previous two examples. To be able to discuss this question we shall introduce the *maximal interval of existence* of a solution of the differential equation.

Next we present a powerful result allowing to assess whether a local solution to (1.2.2) can be extended to a global one; that is, defined for all t (and also providing an alternative proof of the result in the example above). First we have to introduce new terminology. We say that $[t_0, t_0+\alpha^*)$ is the maximal interval of existence for a solution $y(t)$ to (1.2.2) if there is no solution $y_1(t)$ on a longer time interval $[t_0, t_0+\alpha^+)$ where $\alpha^+ > \alpha^*$ satisfying $y(t) = y_1(t)$ for $t \in [t_0, t_0+\alpha^*)$. In other words, we cannot extend $y(t)$ beyond $t_0+\alpha^*$ so that it remains a solution of (1.2.2).

Theorem 2.2 *Assume that f in (1.2.2) satisfies the assumptions of Picard's theorem on $\mathbb{R}^2$. The solution $y(t)$ of (1.2.2) has a finite maximal interval of existence $[t_0, t_0+\alpha^*)$ if and only if*

$$\lim_{t \to t_0+\alpha^*} |y(t)| = \infty. \tag{1.2.18}$$

Proof. Clearly, if (1.2.18) is satisfied, then $y(t)$ cannot be extended beyond $t_0+\alpha^*$. On the other hand, assume that (1.2.18) does not hold. Let us reflect what it means. The meaning of (1.2.18) is that for any K there is t_K such that for any $t_0+\alpha^* > t \ge t_K$ we have $|y(t)| \ge K$. Thus, by saying that (1.2.18) does not hold, we mean that there is K such that for any $t < t_0+\alpha^*$ there is $t < t' < t_0+\alpha^*$ with $|y(t')| < K$. In particular, there is a sequence $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to t_0+\alpha^*$ we have $|y_n| := |y(t_n)| < K$. Consider Cauchy problems

$$y' = f(t,y), \quad y(t_n) = y_n. \tag{1.2.19}$$

Since $|y_n|$ are bounded by K and f satisfies the conditions of the Picard theorem are satisfied on $\mathbb{R}^2$, we can consider the above problem in rectangles $R_n = \{(t,y);\ |t-t_n| < a, |y-y_n| < b\}$ for some fixed a, b. Moreover, all R_ns are contained in the rectangle $R = \{(t,y);\ t_0 - a < t < t_0+\alpha^*, -K-b < y < K+b\}$ and the solutions of the problems (1.2.19) are defined on intervals $(t_n-\alpha, t_n+\alpha)$ where $\alpha = \min\{a, b/M\}$ and M can be taken as $\max_{(t,y)\in R}|f(t,y)|$ and is independent of n. If α^* was finite, then we could find t_n with $t_0+\alpha^*-t_n < \alpha$ so that the solution could be continued beyond $t_0+\alpha^*$ contradicting the assumption that $[t_0, t_0+\alpha^*)$ is the maximal interval of existence. □

Example 2.10 This result allows to give another proof of the fact that solutions of (1.2.2) with globally Lipschitz right-hand side are defined on the whole line. In fact, using Gronwall's lemma, we obtain

$$|y(t)| \le |y_0| + \int_{t_0}^{t} |f(s,y(s)|ds \le |y_0| + \int_{t_0}^{t} |f(s,y(s)) - f(s,y_0)|ds + \int_{t_0}^{t} |f(s,y_0)|ds$$

$$\leq |y_0| + \int_{t_0}^{t} |f(s,y_0)|ds + L\int_{t_0}^{t} |y(s)-y_0)|ds \leq |y_0| + \int_{t_0}^{t} |f(s,y_0)|ds + L\int_{t_0}^{t} |y_0|ds + L\int_{t_0}^{t} |y(s)|ds$$

$$\leq |y_0| + \int_{t_0}^{t} |f(s,y_0)|ds + L(t-t_0)|y_0| + L\int_{t_0}^{t} |y(s)|ds$$

If $y(t)$ is not defined for all t, then by the previous remark, $|y(t)|$ becomes unbounded as $t \to t_{max}$ for some t_{max}. Denoting

$$c = |y_0| + \int_{t_0}^{t_{max}} |f(s,y_0)|ds + L(t_{max}-t_0)|y_0|$$

which is finite as f is continuous for all t, we can write the above inequality as

$$|y(t)| \leq c + L\int_{t_0}^{t} |y(s)|ds$$

for any $t_0 \leq t \leq t_{max}$. Using now Gronwall's lemma, we obtain

$$|y(t)| \leq c\exp Lt \leq c\exp Lt_{max}$$

contradicting thus the definition of t_{max}.

3 A survey of ODEs that can be solved in a closed form

3.1 Separable equations

Separable equations are the ones that can be written in the form

$$\frac{dy}{dt} = \frac{g(t)}{h(y)}, \tag{1.3.20}$$

where g and h are known functions. Firstly, we note that any constant function $y = y_0$, such that $1/h(y_0) = 0$, is *stationary or equilibrium solutions.*

To find a general solution, we assume that $1/h(y) \neq 0$, that is $h(y) \neq \infty$. Multiplying then both sides of (1.3.20) by $h(y)$ to get

$$h(y)\frac{dy}{dt} = g(t) \tag{1.3.21}$$

and observe that, denoting by $H(y) = \int h(y)dy$ the antiderivative of h, we can write (1.3.20) in the form

$$\frac{d}{dt}(H(y(t))) = g(t),$$

that closely resembles (1.3.21). Thus, upon integration we obtain

$$H(y(t)) = \int g(t)dt + c, \tag{1.3.22}$$

where c is an arbitrary constant of integration. The next step depends on the properties of H: for instance, if $H : \mathbb{R} \to \mathbb{R}$ is monotonic, then we can find y explicitly for all t as

$$y(t) = H^{-1}\left(\int g(t)dt + c\right).$$

Otherwise, we have to do it locally, around the initial values. To explain this, we solve the initial value problem for separable equation.

$$\begin{aligned} \frac{dy}{dt} &= \frac{g(t)}{h(y)}, \\ y(t_0) &= y_0, \end{aligned} \tag{1.3.23}$$

Using the general solution (1.3.22) (with definite integral) we obtain

$$H(y(t)) = \int_{t_0}^{t} g(s)ds + c,$$

we obtain

$$H(y_0) = H(y(t_0)) = \int_{t_0}^{t_0} a(s)ds + c,$$

which, due $\int_{t_0}^{t_0} a(s)ds = 0$, gives

$$c = H(y_0),$$

so that

$$H(y(t)) = \int_{t_0}^{t} g(s)ds + H(y_0).$$

We are interested in the existence of the solution at least close to t_0, which means that H should be invertible close to y_0. From the Implicit Function Theorem we obtain that this is possible if H is differentiable in a neighbourhood of y_0 and $\partial H/\partial y(y_0) \neq 0$. But $\partial H/\partial y(y_0) = h(y_0)$, so we are back at Picard's theorem: if $h(y)$ is differentiable in the neighbourhood of y_0 with $h(y_0) \neq 0$ (if $h(y_0) = 0$, then the equation (1.3.20) does not make sense at y_0, and g is continuous, then $f(t, y) = g(t)/h(y)$ satisfies the assumptions of the theorem in some neighbourhood of (t_0, y_0).

3.2 Linear ordinary differential equations of first order

The general first order linear differential equation is

$$\frac{dy}{dt} + a(t)y = b(t). \tag{1.3.24}$$

Functions a and b are known continuous functions of t.

A solution of the equation

$$\frac{d\mu(t)}{dt} = \mu(t)a(t),$$

is called but an *integrating factor* of the equation (1.3.24). This is a separable equation, the general solution of which is given by (1.3.22). Since we need only one such function, we may take

$$\mu(t) = \exp\left(\int a(t)dt\right)$$

to be used as the integrating factor. With such function, (1.3.24) can be written as

$$\frac{d}{dt}\mu(t)y = \mu(t)b(t),$$

where c is an arbitrary constant of integration. Finally

$$\begin{aligned} y(t) &= \frac{1}{\mu(t)}\left(\int \mu(t)b(t)dt + c\right) \\ &= \exp\left(-\int a(t)dt\right)\left(\int b(t)\exp\left(\int a(t)dt\right)dt + c\right) \end{aligned} \tag{1.3.25}$$

It is worthwhile to note that the solution consists of two parts: the general solution to the reduced equation associated with (1.3.24)

$$c\exp\left(-\int a(t)dt\right)$$

and, what can be checked by direct differentiation, a particular solution to the full equation.

If we want to find a particular solution satisfying $y(t_0) = y_0$, then we write (1.3.25) using definite integrals

$$y(t) = \exp\left(-\int_{t_0}^{t} a(s)ds\right)\left(\int_{t_0}^{t} b(s)\exp\left(\int_{t_0}^{s} a(r)dr\right)ds + c\right)$$

and use the fact that $\int_{t_0}^{t_0} f(s)ds = 0$ for any function f. This shows that the part of the solution satisfying the nonhomogeneous equation:

$$y_b(t) = \exp\left(-\int_{t_0}^{t} a(s)ds\right)\int_{t_0}^{t} b(s)\exp\left(\int_{t_0}^{s} a(r)dr\right)ds$$

takes on the zero value at $t = t_0$. Thus

$$y_0 = y(t_0) = c$$

and the solution to the initial value problem is given by

$$y(t) = y_0\exp\left(-\int_{t_0}^{t} a(s)ds\right) + \exp\left(-\int_{t_0}^{t} a(s)ds\right)\int_{t_0}^{t} b(s)\exp\left(\int_{t_0}^{s} a(r)dr\right)ds. \tag{1.3.26}$$

3.3 Equations of homogeneous type

In differential equations, as in integration, a smart substitution can often convert a complicated equation into a manageable one. For some classes of differential equations there are standard substitutions that transform them into separable equations. We shall discuss one such a class in detail.

A differential equation that can be written in the form

$$\frac{dy}{dt} = f\left(\frac{y}{t}\right), \tag{1.3.27}$$

where f is a function of the single variable $z = y/t$ is said to be of *homogeneous type*. Note that in some textbooks such equations are called *homogeneous equations* but this often creates confusion as the name homogeneous equation is generally used in another context.

To solve (1.3.27) we make substitution

$$y = tz \tag{1.3.28}$$

where z is the new unknown function. Then, by the product rule for derivatives

$$\frac{dy}{dt} = z + t\frac{dz}{dt}$$

and (1.3.27) becomes

$$z + t\frac{dz}{dt} = f(z),$$

or

$$t\frac{dz}{dt} = f(z) - z. \tag{1.3.29}$$

In (1.3.29) the variables are separable so it can be solved as in Subsection 3.1.

3.4 Equations that can be reduced to first order equations

Some higher order equations can be reduced to equations of the first order. We shall discuss two such cases for second order equations.

Equations that do not contain the unknown function

If we have the equation of the form

$$F(y'', y', t) = 0, \tag{1.3.30}$$

then the substitution $z = y'$ reduces this equation to an equation of the first order

$$F(z', z, t) = 0. \tag{1.3.31}$$

If we can solve this equation

$$z = \phi(t, C),$$

where C is an arbitrary constant, then, returning to the original unknown function y, we obtain another first order equation

$$y' = \phi(t, C),$$

which is immediately solvable as

$$y(t) = \int \phi(t, C)dt + C_1.$$

Equations that do not contain the independent variable

Let us consider the equation

$$F(y'', y', y) = 0, \tag{1.3.32}$$

that does not involve the independent variable t. Such an equation can be also reduced to a first order equation, the idea, however, is a little more complicated. Firstly, we note that the derivative y' is uniquely defined by the function y. This means that we can write $y' = g(y)$ for some function g. Using the chain rule we obtain

$$y'' = \frac{d}{dt}y' = \frac{dg}{dy}(y)\frac{dy}{dt} = y'\frac{dg}{dy}(y) = g(y)\frac{dg}{dy}(y). \tag{1.3.33}$$

Substituting (1.3.33) into (1.3.32) gives the first order equation with y as an independent variable

$$F\left(g\frac{dg}{dy}, g, y\right) = 0. \tag{1.3.34}$$

If we solve this equation in the form $g(y) = \phi(y, C)$, then to find y we have to solve one more first order equation with t as the independent variable

$$\frac{dy}{dt} = \phi(y, C).$$

Chapter 2

Systems of differential equations

1 Why systems?

Two possible generalizations of first order scalar equation

$$y' = f(t, y) :$$

one is a differential equation of higher order

$$y^{(n)} = F(t, y', y'', \ldots, y^{(n-1)}) = 0, \tag{2.1.1}$$

(where, for simplicity, we consider only equations solved with respect to the highest derivative), and the other is a system of first order equations, that is,

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}) \tag{2.1.2}$$

where, in general,

$$\mathbf{y}(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix},$$

and

$$\mathbf{f}(t, \mathbf{y}) = \begin{pmatrix} f_1(t, y_1, \ldots, y_n) \\ \vdots \\ f_n(t, y_1, \ldots, y_n) \end{pmatrix}.$$

is a nonlinear function of t and $\mathbf{y}$. It turns out that, at least from the theoretical point of view, there is no need to consider these two separately as equation of higher order can be always written as as system (the converse, in general, is not true). To see how this can be accomplish, we introduce new unknown variables $z_1(t) = y(t), z_2(t) = y'(t), z_n = y^{(n-1)}(t)$ so that $z_1'(t) = y'(t) = z_2(t), z_2'(t) = y''(t) = z_3(t), \ldots$ and (2.1.1) converts into

$$\begin{aligned} z_1' &= z_2, \\ z_2' &= z_3, \\ &\vdots \\ z_n' &= F(t, z_1, \ldots, z_n) \end{aligned}$$

Clearly, solving this system, we obtain simultaneously the solution of (2.1.1) by taking $y(t) = z_1(t)$.

2 Local existence and basic properties

In this chapter we shall consider the systems of differential equations (2.1.2).

There is an version of Picard's systems giving local solvability of the Cauchy problem for (2.1.2):

$$\begin{aligned} \mathbf{y}' &= \mathbf{f}(t,\mathbf{y}), \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \end{aligned} \tag{2.2.3}$$

but before we formulate and discuss it in detail, we shall recall a few notions from vector analysis. Let $\mathbf{y}(t)$ is a vector-valued function with values in some $\mathbb{R}^n$. $\mathbf{y}(t)$ is continuous if each of its components is continuous. As $\mathbf{y}'(t) = (y_1'(t), \dots, y_n'(t))$, we define similarly

$$\int_a^b \mathbf{y}(t)dt = \left(\int_a^b y_1(t)dt, \dots, \int_a^b y_n(t)dt \right).$$

An important concept for any vector $\mathbf{v}$ is its norm. In general, a norm is any real valued function $\|\cdot\|$ of a vector satisfying $\|\mathbf{v}\| \geq 0$ with $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v}\| = 0$, $\|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\|$ for any scalar λ and $\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for any two vectors $\mathbf{v}, \mathbf{w}$. A classical example of a norm in the euclidean norm

$$\|\mathbf{v}\|_2 = \sqrt{v_1^2 + \dots v_n^2}$$

but other often used norms are

$$\|\mathbf{v}\|_\infty = \max\{|v_1|, \dots, |v_n|\}$$

and

$$\|\mathbf{v}\|_1 = |v_1| + \dots + |v_n|.$$

An important property of the norm is

$$\| \int_a^b \mathbf{y}(t)dt\| \leq \int_a^b \|\mathbf{y}(t)\|dt.$$

It can be easily verified for the two latter norms but for the euclidean norm the proof requires some work.

Theorem 2.1 *If each of the functions* $\mathbf{f}$ *is continuous in a region* $\mathbf{R} : |t-t_0| \leq a, \|\mathbf{y}-\mathbf{y}^0\| \leq b$ *and*

$$\|\mathbf{f}(t,\mathbf{y}_1) - \mathbf{f}(t,\mathbf{y}_2)\| \leq L\|\mathbf{y}_1 - \mathbf{y}_2\|$$

for some constant L *and all* $(t,\mathbf{y}_1), (t,\mathbf{y}_2) \in \mathbf{R}$*, then the initial value problem (2.2.3) has one and only one solution* $\mathbf{y}(t) = (y_1(t), \dots, y_n(t))$ *defined at least on the interval* $|t-t_0| \leq \alpha$ *where* $\alpha = \min\{a, b/M\}$ *and* $M = \max_{(t,y)\in\mathbf{R}} \|f(t,y)\|$.

The proof of this theorem is a repetition of the scalar case so that we shall skip the details. To give, however, some flavour of operations with vector functions, we shall prove uniqueness. Starting as in the scalar case, we let $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ be two solutions of the Cauchy problem (2.2.3) on $\mathbf{R}$ with the same initial condition $\mathbf{y}^0$, that is $\mathbf{y}_1'(t) \equiv \mathbf{f}(t,\mathbf{y}_1(t))$, $\mathbf{y}_1(t_0) = \mathbf{y}_0$ and $\mathbf{y}_2'(t) \equiv \mathbf{f}(t,\mathbf{y}_2(t))$, $\mathbf{y}_2(t_0) = \mathbf{y}^0$. Then, integrating and using the equality at t_0, we see that

$$\mathbf{y}_1(t) - \mathbf{y}_2(t) = \int_0^t (\mathbf{f}(t,\mathbf{y}_1(s)) - \mathbf{f}_2(t,\mathbf{y}_2(s)))ds.$$

Thus

$$\begin{aligned}\|\mathbf{y}_1(t)-\mathbf{y}_2(t)\| &= \left\|\int_{t_0}^{t}(\mathbf{f}(t,\mathbf{y}_1(s))-\mathbf{f}_2(t,\mathbf{y}_2(s)))ds\right\| \leq \int_{t_0}^{t}\|\mathbf{f}(t,\mathbf{y}_1(s))-\mathbf{f}_2(t,\mathbf{y}_2(s))\|ds \\ &\leq L\int_{t_0}^{t}\|\mathbf{y}_1(s))-\mathbf{y}_2(s)\|ds, \end{aligned} \tag{2.2.4}$$

thus we can use the second part of Gronwall's lemma to claim that $\|\mathbf{y}_1(t)-\mathbf{y}_2(t)\|=0$ and from the properties of the norm, we obtain $\mathbf{y}_1(t)=\mathbf{y}_2(t)$ for all t satisfying $|t-t_0|<\alpha$.

Many properties of the solution can be proved in a way that is analogous to the scalar case. In particular, if f is such that the assumptions of Picard's theorem are satisfied everywhere and $I=(t_0-\alpha,t_0+\alpha)$ is the maximum interval of existence of the solution $\mathbf{y}(t,t_0,\mathbf{y}^0)$ of the Cauchy problem (2.2.3), then

$$\lim_{t\to t_0\pm\alpha}\|\mathbf{y}(t)\|=\infty.$$

Thus, in particular, if the right-hand side is bounded on $\mathbb{R}^{n+1}$: $\sup_{(t,\mathbf{y})\in\mathbb{R}^{n+1}}\|f(t,\mathbf{y})\|\leq M$ for some constant M, or globally Lipschitz:

$$\|\mathbf{f}(t,\mathbf{y}_1)-\mathbf{f}(t,\mathbf{y}_2)\|\leq L\|\mathbf{y}_1-\mathbf{y}_2\|$$

for some constant L and all $(t,\mathbf{y}_1),(t,\mathbf{y}_2)\in\mathbb{R}^{n+1}$, then the solution to (2.2.3) is defined for all times.

Lemma 2.1 *Let $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ be two solutions of the equation*

$$\mathbf{y}'=\mathbf{f}(t,\mathbf{y}),$$

satisfying the initial conditions $\mathbf{y}_1(t_0)=\mathbf{y}_{1,0}$ and $\mathbf{y}_2(t_0)=\mathbf{y}_{2,0}$, where $\mathbf{f}$ satisfies the assumptions of Picard's theorem on some rectangle $\mathbf{R}$ containing both $(t_0,\mathbf{y}_{1,0})$ and $(t_0,\mathbf{y}_{2,0})$. Then

$$\|\mathbf{y}_1(t)-\mathbf{y}_2(t)\|\leq\|\mathbf{y}_1-\mathbf{y}_2\|e^{L|t-t_0|}$$

for all $|t-t_0|\leq\alpha'$, where L is the Lipschitz constant for $\mathbf{f}$ on $\mathbf{R}$ and $|t-t_0|<\alpha'$ is the common interval of existence of both solutions.

Proof. Starting as for the uniqueness, we let $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ be the solutions of the Cauchy problem for (2.4.26) on $\mathbf{R}$; that is, $\mathbf{y}_1'(t)\equiv\mathbf{f}(t,\mathbf{y}_1(t))$, $\mathbf{y}_1(t_0)=\mathbf{y}_{1,0}$ and $\mathbf{y}_2'(t)\equiv\mathbf{f}(t,\mathbf{y}_2(t))$, $\mathbf{y}_1(t)=\mathbf{y}_{2,0}$. Then, integrating and using the initial conditions, we see that

$$\mathbf{y}_1(t)-\mathbf{y}_2(t)=\mathbf{y}_{1,0}-\mathbf{y}_{2,0}+\int_0^t(\mathbf{f}(t,\mathbf{y}_1(s))-\mathbf{f}_2(t,\mathbf{y}_2(t)))ds.$$

Thus

$$\begin{aligned}\|\mathbf{y}_1(t)-\mathbf{y}_2(t)\| &= \|\mathbf{y}_{1,0}-\mathbf{y}_{2,0}\|+\left\|\int_{t_0}^{t}(\mathbf{f}(t,\mathbf{y}_1(s))-\mathbf{f}(t,\mathbf{y}_2(s)))ds\right\| \\ &\leq \|\mathbf{y}_{1,0}-\mathbf{y}_{2,0}\|+\int_{t_0}^{t}\|\mathbf{f}(t,\mathbf{y}_1(s))-\mathbf{f}(t,\mathbf{y}_2(s))\|ds \\ &\leq \|\mathbf{y}_{1,0}-\mathbf{y}_{2,0}\|+L\int_{t_0}^{t}\|\mathbf{y}_1(s)-\mathbf{y}_2(s)\|ds, \end{aligned} \tag{2.2.5}$$

thus we can use the first part of Gronwall's lemma to obtain the thesis. ■

3 Solvability of linear systems

We shall consider only linear systems of first order differential equations.

$$\begin{array}{rcl} y_1' &=& a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n + g_1(t), \\ \vdots & \vdots & \vdots, \\ y_n' &=& a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n + g_n(t), \end{array} \tag{2.3.6}$$

where $y_1, \ldots, y_n$ are unknown functions, $a_{11}, \ldots a_{nn}$ are constant coefficients and $g_1(t) \ldots, g_n(t)$ are known continuous functions. If $g_1 = \ldots = g_n = 0$, then the corresponding system (2.8.42) is called the associated homogeneous system. The structure of (2.8.42) suggest that a more economical way of writing is is to use the vector-matrix notation. Denoting $\mathbf{y} = (y_1, \ldots, y_n)$, $\mathbf{g} = (g_1, \ldots, g_n)$ and $\mathcal{A} = \{a_{ij}\}_{1 \le i,j \le n}$, that is

$$\mathcal{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix},$$

we can write (2.8.42) is a more concise notation as

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}. \tag{2.3.7}$$

Here we have n unknown functions and the system involves first derivative of each of them so that it is natural to consider (2.8.43) in conjunction with the following initial conditions

$$\mathbf{y}(t_0) = \mathbf{y^0}, \tag{2.3.8}$$

or, in the expanded form,

$$y_1(t_0) = y_1^0, \ldots, y_n(t_0) = y_n^0, \tag{2.3.9}$$

where t_0 is a given argument and $\mathbf{y^0} = (y_1^0, \ldots, y_n^0)$ is a given vector.

Let us denote by $\mathbf{X}$ the set of all solutions to the homogeneous system (2.8.42). Due to linearity of differentiation and multiplication by $\mathcal{A}$, it is easy to see that $\mathbf{X}$ is a vector space. We have two fundamental results.

Theorem 3.1 *The dimension of* $\mathbf{X}$ *is equal to* n.

Theorem 3.2 *Let* $\mathbf{y_1}, \ldots, \mathbf{y_k}$ *be* k *linearly independent solutions of* $\mathbf{y}' = \mathcal{A}\mathbf{y}$ *and let* $t_0 \in \mathbb{R}$ *be an arbitrary number. Then,* $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ *for a linearly independent set of functions if and only if* $\{\mathbf{y_1}(t_0), \ldots, \mathbf{y_k}(t_0)\}$ *is a linearly independent set of vectors in* $\mathbb{R}$.

These two results show that if we construct solutions emanating from n linearly independent initial vectors, then these solutions are linearly independent and therefore they span the space of all solutions to the homogeneous system (2.8.42).

Let $\mathcal{A}$ be an $n \times n$ matrix. We say that a number λ (real or complex) is an *eigenvalue* of $\mathcal{A}$ is there exist a non-zero solution of the equation

$$\mathcal{A}\mathbf{v} = \lambda\mathbf{v}. \tag{2.3.10}$$

Such a solution is called an *eigenvector* of $\mathcal{A}$. The set of eigenvectors corresponding to a given eigenvalue is a vector subspace. Eq. (2.8.46) is equivalent to the homogeneous system $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$, where $\mathcal{I}$ is the identity matrix, therefore λ is an eigenvalue of $\mathcal{A}$ if and only if the determinant of $\mathcal{A}$ satisfies

$$det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} a_{11} - \lambda & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{2.3.11}$$

Evaluating the determinant we obtain a polynomial in λ of degree n. This polynomial is also called the characteristic polynomial of the system (2.8.42) (if (2.8.42) arises from a second order equation, then this is the same polynomial as the characteristic polynomial of the equation). We shall denote this polynomial by $p(\lambda)$. From algebra we know that there are exactly n, possibly complex, root of $p(\lambda)$. Some of them may be multiple, so that in general $p(\lambda)$ factorizes into

$$p(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdot \ldots \cdot (\lambda_k - \lambda)^{n_k}, \tag{2.3.12}$$

with $n_1 + \ldots + n_k = n$. It is also worthwhile to note that since the coefficients of the polynomial are real, then complex roots appear always in conjugate pairs, that is, if $\lambda_j = \xi_j + i\omega_j$ is a characteristic root, then so is $\bar{\lambda}_j = \xi_j - i\omega_j$. Thus, eigenvalues are roots of the characteristic polynomial of $\mathcal{A}$. The exponent n_i appearing in the factorization (2.8.48) is called the *algebraic multiplicity* of λ_i. For each eigenvalue λ_i there corresponds an eigenvector $\mathbf{v_i}$ and eigenvectors corresponding to distinct eigenvalues are linearly independent. The set of all eigenvectors corresponding to λ_i spans a subspace, called the *eigenspace* corresponding to λ_i which we will denote by E_{λ_i}. The dimension of E_{λ_i} is called the *geometric multiplicity* of λ_i. In general, algebraic and geometric multiplicities are different with geometric multiplicity being at most equal to the algebraic one. Thus, in particular, if λ_i is a single root of the characteristic polynomial, then the eigenspace corresponding to λ_1 is one-dimensional.

If the geometric multiplicities of eigenvalues add up to n, that is, if we have n linearly independent eigenvectors, then these eigenvectors form a basis for $\mathbb{R}^n$. In particular, this happens if all eigenvalues are single roots of the characteristic polynomial. If this is not the case, then we do not have sufficiently many eigenvectors to span $\mathbb{R}^n$ and if we need a basis for $\mathbb{R}^n$, then we have to find additional linearly independent vectors. A procedure that can be employed here and that will be very useful in our treatment of systems of differential equations is to find solutions to equations of the form $(\mathcal{A} - \lambda_i \mathcal{I})^k \mathbf{v} = 0$ for $1 < k \leq n_i$, where n_i is the algebraic multiplicity of λ_i. Precisely speaking, if λ_i has algebraic multiplicity n_i and if

$$(\mathcal{A} - \lambda_i \mathcal{I})\mathbf{v} = 0$$

has only $\nu_i < n_i$ linearly independent solutions, then we consider the equation

$$(\mathcal{A} - \lambda_i \mathcal{I})^2 \mathbf{v} = 0.$$

It follows that all the solutions of the preceding equation solve this equation but there is at least one more independent solution so that we have at least $\nu_i + 1$ independent vectors (note that these new vectors are no longer eigenvectors). If the number of independent solutions is still less than n_1, we consider

$$(\mathcal{A} - \lambda_i \mathcal{I})^3 \mathbf{v} = 0,$$

and so on, till we get a sufficient number of them. Note, that to make sure that in the step j we select solutions that are independent of the solutions obtained in step $j-1$ it is enough to find solutions to $(\mathcal{A} - \lambda_i \mathcal{I})^j \mathbf{v} = 0$ that satisfy $(\mathcal{A} - \lambda_i \mathcal{I})^{j-1} \mathbf{v} \neq 0$.

Matrix exponentials

The above theory can be used to provide a unified framework for solving systems of differential equations.

Recall that for a single equation $y' = ay$, where a is a constant, the general solution is given by $y(t) = e^{at}C$, where C is a constant. In a similar way, we would like to say that the general solution to

$$\mathbf{y}' = \mathcal{A}\mathbf{y},$$

where $\mathcal{A}$ is an $n \times n$ matrix, is $\mathbf{y} = e^{\mathcal{A}t}\mathbf{v}$, where $\mathbf{v}$ is any constant vector in $\mathbb{R}^n$. The problem is that we do not know what it means to evaluate an exponential of a matrix. However, if we reflect for a moment that the exponential of a number can be evaluated as the power (Maclaurin) series

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots + \frac{x^k}{k!} + \ldots,$$

where the only involved operations on the argument x are additions, scalar multiplications and taking integer powers, we come to the conclusion that the above expression can be written also for a matrix, that is, we can define

$$e^{\mathcal{A}} = \mathcal{I} + \mathcal{A} + \frac{1}{2}\mathcal{A}^2 + \frac{1}{3!}\mathcal{A}^3 + \ldots + \frac{1}{k!}\mathcal{A}^k + \ldots. \tag{2.3.13}$$

It can be shown that if $\mathcal{A}$ is a matrix, then the above series always converges and the sum is a matrix. For example, if we take

$$\mathcal{A} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \lambda\mathcal{I},$$

then

$$\mathcal{A}^k = \lambda^k\mathcal{I}^k = \lambda^k\mathcal{I},$$

and

$$\begin{aligned} e^{\mathcal{A}} &= \mathcal{I} + \lambda\mathcal{I} + \frac{\lambda^2}{2}\mathcal{I} + \frac{\lambda^3}{3!}\mathcal{I} + \ldots + \frac{\lambda^k}{k!} + \ldots \\ &= \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \ldots + \frac{\lambda^k}{k!} + \ldots\right)\mathcal{I} \\ &= e^{\lambda}\mathcal{I}. \end{aligned} \tag{2.3.14}$$

Unfortunately, in most cases finding the explicit form for $e^{\mathcal{A}}$ directly is impossible.

Matrix exponentials have the following algebraic properties

$$\left(e^{\mathcal{A}}\right)^{-1} = e^{-\mathcal{A}}$$

and

$$e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{A}}e^{\mathcal{B}} \tag{2.3.15}$$

provided the matrices $\mathcal{A}$ and $\mathcal{B}$ commute: $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$.

Let us define a function of t by

$$e^{t\mathcal{A}} = \mathcal{I} + t\mathcal{A} + \frac{t^2}{2}\mathcal{A}^2 + \frac{t^3}{3!}\mathcal{A}^3 + \ldots + \frac{t^k}{k!}\mathcal{A}^k + \ldots\ . \tag{2.3.16}$$

It follows that this function can be differentiated with respect to t by termwise differentiation of the series, as in the scalar case, that is,

$$\begin{aligned} \frac{d}{dt}e^{\mathcal{A}t} &= \mathcal{A} + t\mathcal{A}^2 + \frac{t^2}{2!}\mathcal{A}^3 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^k + \ldots \\ &= \mathcal{A}\left(\mathcal{I} + t\mathcal{A} + \frac{t^2}{2!}\mathcal{A}^2 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^{k-1} + \ldots\right) \\ &= \mathcal{A}e^{t\mathcal{A}} = e^{t\mathcal{A}}\mathcal{A}, \end{aligned}$$

proving thus that $\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{v}$ is a solution to our system of equations for any constant vector $\mathbf{v}$. Since $\mathbf{y}(0) = e^{0\mathcal{A}}\mathbf{v} = \mathbf{v}$, from Picard's theorem $\mathbf{y}(t)$ is a unique solution to the Cauchy problem

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{v}.$$

As we mentioned earlier, in general it is difficult to find directly the explicit form of $e^{t\mathcal{A}}$. However, we can always find n linearly independent vectors $\mathbf{v}$ for which the series $e^{t\mathcal{A}}\mathbf{v}$ can be summed exactly. This is based on the following two observations. Firstly, since $\lambda\mathcal{I}$ and $\mathcal{A} - \lambda\mathcal{I}$ commute, we have by (2.8.50) and (2.8.51)

$$e^{t\mathcal{A}}\mathbf{v} = e^{t(\mathcal{A}-\lambda\mathcal{I})}e^{t\lambda\mathcal{I}}\mathbf{v} = e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v}.$$

Secondly, if $(\mathcal{A}-\lambda\mathcal{I})^m\mathbf{v}=\mathbf{0}$ for some m, then

$$(\mathcal{A}-\lambda\mathcal{I})^r\mathbf{v}=\mathbf{0}, \tag{2.3.17}$$

for all $r\geq m$. This follows from

$$(\mathcal{A}-\lambda\mathcal{I})^r\mathbf{v}=(\mathcal{A}-\lambda\mathcal{I})^{r-m}[(\mathcal{A}-\lambda\mathcal{I})^m\mathbf{v}]=\mathbf{0}.$$

Consequently, for such a $\mathbf{v}$

$$e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v}=\mathbf{v}+t(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}+\ldots+\frac{t^{m-1}}{(m-1)!}(\mathcal{A}-\lambda\mathcal{I})^{m-1}\mathbf{v}.$$

and

$$e^{t\mathcal{A}}\mathbf{v}=e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})}=e^{\lambda t}\left(\mathbf{v}+t(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}+\ldots+\frac{t^{m-1}}{(m-1)!}(\mathcal{A}-\lambda\mathcal{I})^{m-1}\mathbf{v}\right). \tag{2.3.18}$$

Thus, to find all solutions to $\mathbf{y}'=\mathcal{A}\mathbf{y}$ it is sufficient to find n independent vectors $\mathbf{v}$ satisfying (2.8.53) for some scalars λ. But these are precisely the eigenvectors or associated eigenvectors and we know that it is possible to find exactly n of them.

Thus, for example, if $\lambda=\lambda_1$ is a simple eigenvalue of $\mathcal{A}$ with a corresponding eigenvector $\mathbf{v^1}$, then $(\mathcal{A}-\lambda_1\mathcal{I})\mathbf{v^1}=1$, thus m of (2.8.53) is equal to 1. Consequently, the sum in (2.8.54) terminates after the first term and we obtain

$$\mathbf{y^1}(t)=e^{\lambda_1}\mathbf{v^1}.$$

From our discussion of eigenvalues and eigenvectors it follows that if λ_i is a multiple eigenvalue of $\mathcal{A}$ of algebraic multiplicity n_i and the geometric multiplicity is less then n_i, that is, there is less than n_i linearly independent eigenvectors corresponding to λ_i, then the missing independent vectors can be found by solving successively equations $(\mathcal{A}-\lambda_i\mathcal{I})^k\mathbf{v}=\mathbf{0}$ with k running at most up to n_1. Thus, we have the following algorithm for finding n linearly independent solutions to $\mathbf{y}'=\mathcal{A}\mathbf{y}$:

1. Find all eigenvalues of $\mathcal{A}$;
2. If λ is a single real eigenvalue, then there is an eigenvector $\mathbf{v}$ so that the solution is given by
$$\mathbf{y}(t)=e^{\lambda t}\mathbf{v} \tag{2.3.19}$$
3. If λ is a single complex eigenvalue $\lambda=\xi+i\omega$, then there is a complex eigenvector $\mathbf{v}=\Re\mathbf{v}+i\Im\mathbf{v}$ such that two solutions corresponding to λ (and $\bar{\lambda}$) are given by
$$\begin{aligned}\mathbf{y^1}(t) &= e^{\xi t}(\cos\omega t\,\Re\mathbf{v}-\sin\omega t\,\Im\mathbf{v})\\ \mathbf{y^2}(t) &= e^{\xi t}(\cos\omega t\,\Im\mathbf{v}+\sin\omega t\Re\mathbf{v})\end{aligned} \tag{2.3.20}$$
4. If λ is a multiple eigenvalue with algebraic multiplicity k (that is, λ is a multiple root of the characteristic equation, of multiplicity k), then we first find eigenvectors by solving $(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}=\mathbf{0}$. For these eigenvectors the solution is again given by (2.8.55) (or (2.8.56), if λ is complex). If we found k independent eigenvectors, then our work with this eigenvalue is finished. If not, then we look for vectors that satisfy $(\mathcal{A}-\lambda\mathcal{I})^2\mathbf{v}=\mathbf{0}$ but $(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}\neq\mathbf{0}$. For these vectors we have the solutions
$$e^{t\mathcal{A}}\mathbf{v}=e^{\lambda t}\left(\mathbf{v}+t(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}\right).$$
If we still do not have k independent solutions, then we find vectors for which $(\mathcal{A}-\lambda\mathcal{I})^3\mathbf{v}=\mathbf{0}$ and $(\mathcal{A}-\lambda\mathcal{I})^2\mathbf{v}\neq\mathbf{0}$, and for such vectors we construct solutions
$$e^{t\mathcal{A}}\mathbf{v}=e^{\lambda t}\left(\mathbf{v}+t(\mathcal{A}-\lambda\mathcal{I})\mathbf{v}+\frac{t^2}{2}(\mathcal{A}-\lambda\mathcal{I})^2\mathbf{v}\right).$$
This procedure is continued till we have k solutions (by the properties of eigenvalues we have to repeat this procedure at most k times).
If λ is a complex eigenvalue of multiplicity k, then also $\bar{\lambda}$ is an eigenvalue of multiplicity k and we obtain pairs of real solutions by taking real and imaginary parts of the formulae presented above.

Fundamental solutions and nonhomogeneous problems

Let us suppose that we have n linearly independent solutions $\mathbf{y^1}(t), \ldots, \mathbf{y^n}(t)$ of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$, where $\mathcal{A}$ is an $n \times n$ matrix, like the ones constructed in the previous paragraphs. Let us denote by $\mathcal{Y}(t)$ the matrix

$$\mathcal{Y}(t) = \begin{pmatrix} y_1^1(t) & \ldots & y_1^n(t) \\ \vdots & & \vdots \\ y_n^1(t) & \ldots & y_n^n(t) \end{pmatrix},$$

that is, the columns of $\mathcal{Y}(t)$ are the vectors $\mathbf{y^i}$, $i = 1, \ldots, n$. Any such matrix is called a *fundamental matrix* of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$.

We know that for a given initial vector $\mathbf{y^0}$ the solution is given by

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{y^0}$$

on one hand, and, by Theorem 8.1, by

$$\mathbf{y}(t) = C_1\mathbf{y^1}(t) + \ldots + C_n\mathbf{y^n}(t) = \mathcal{Y}(t)\mathbf{C},$$

on the other, where $\mathbf{C} = (C_1, \ldots, C_n)$ is a vector of constants to be determined. By putting $t = 0$ above we obtain the equation for $\mathbf{C}$

$$\mathbf{y^0} = \mathcal{Y}(0)\mathbf{C}$$

Since $\mathcal{Y}$ has independent vectors as its columns, it is invertible, so that

$$\mathbf{C} = \mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Thus, the solution of the initial value problem

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{y^0}$$

is given by

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Since $e^{t\mathbf{A}}\mathbf{y^0}$ is also a solution, by the uniqueness theorem we obtain explicit representation of the exponential function of a matrix

$$e^{t\mathcal{A}} = \mathcal{Y}(t)\mathcal{Y}^{-1}(0). \tag{2.3.21}$$

Let us turn our attention to the non-homogeneous system of equations

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}(t). \tag{2.3.22}$$

The general solution to the homogeneous equation ($\mathbf{g}(t) \equiv 0$) is given by

$$\mathbf{y_h}(t) = \mathcal{Y}(t)\mathbf{C},$$

where $\mathcal{Y}(t)$ is a fundamental matrix and $\mathbf{C}$ is an arbitrary vector. Using the technique of variation of parameters, we will be looking for the solution in the form

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{u}(t) = u_1(t)\mathbf{y^1}(t) + \ldots + u_n(t)\mathbf{y^n}(t) \tag{2.3.23}$$

where $\mathbf{u}(t) = (u_1(t), \ldots, u_n(t))$ is a vector-function to be determined so that (2.8.59) satisfies (2.8.58). Thus, substituting (2.8.59) into (2.8.58), we obtain

$$\mathcal{Y}'(t)\mathbf{u}(t) + \mathcal{Y}(t)\mathbf{u}'(t) = \mathcal{A}\mathcal{Y}(t)\mathbf{u}(t) + \mathbf{g}(t).$$

Since $\mathcal{Y}(t)$ is a fundamental matrix, $\mathcal{Y}'(t) = \mathcal{A}\mathcal{Y}(t)$ and we find

$$\mathcal{Y}(t)\mathbf{u}'(t) = \mathbf{g}(t).$$

As we observed earlier, $\mathcal{Y}(t)$ is invertible, hence

$$\mathbf{u}'(t) = \mathcal{Y}^{-1}(t)\mathbf{g}(t)$$

and

$$\mathbf{u}(t) = \int^t \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds + \mathbf{C}.$$

Finally, we obtain

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{C} + \mathcal{Y}(t)\int^t \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds \tag{2.3.24}$$

This equation becomes much simpler if we take $e^{t\mathcal{A}}$ as a fundamental matrix because in such a case $\mathcal{Y}^{-1}(t) = \left(e^{t\mathcal{A}}\right)^{-1} = e^{-t\mathcal{A}}$, that is, to calculate the inverse of $e^{t\mathcal{A}}$ it is enough to replace t by $-t$. The solution (2.8.60) takes then the form

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{C} + \int e^{(t-s)\mathcal{A}}\mathbf{g}(s)ds. \tag{2.3.25}$$

4 Flow of an autonomous equation – basic properties

From now on our interest lies with the Cauchy problem for the autonomous system of equations in $\mathbb{R}^n$

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}), \tag{2.4.26}$$
$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.4.27}$$

To simplify considerations, we assume that $\mathbf{f}$ satisfies the assumptions of the Picard theorem on $\mathbb{R}^n$ so that the solutions exist for all $-\infty < t < \infty$. The *flow* of (2.4.26) is the map

$$\mathbb{R} \times \mathbb{R}^n \ni (t, \mathbf{x}_0) \to \phi(t, \mathbf{x_0}) \in \mathbb{R}^n$$

where $\mathbf{x}(t) = \phi(t, \mathbf{x}_0)$ is the solution to (2.4.26) satisfying $\mathbf{x}(0) = \mathbf{x}_0$. We note the following important properties of the flow:

$$\phi(0, \mathbf{x}) = \mathbf{x}, \tag{2.4.28}$$
$$\phi(t_1, \phi(t_2, \mathbf{x})) = \phi(t_1 + t_2, \mathbf{x}) \tag{2.4.29}$$

for any $t_1, t_2 \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}$. Property (2.4.29) follows from the simple lemma which w note for further reference

Lemma 4.1 *If $\mathbf{x}(t)$ is a solution to*

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}),$$

then for any c the function $\hat{\mathbf{x}}(t) = \mathbf{x}(t + c)$ also satisfies this equation.

Proof. Define $\tau = t + c$ and use the chain rule for $\hat{x}$. We get

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \frac{d\mathbf{x}(t+c)}{dt} = \frac{d\mathbf{x}(\tau)}{d\tau}\frac{d\tau}{dt} = \frac{d\mathbf{x}(\tau)}{d\tau} = \mathbf{f}(\mathbf{x}(\tau)) = \mathbf{f}(\mathbf{x}(t+c)) = \mathbf{f}(\hat{\mathbf{x}}(t)).$$

In terms of the flow we can rephrase the lemma by noting that $\mathbf{u}(t) := \phi(t + t_2, \mathbf{x})$ is the solution of (2.4.26), with $\mathbf{u}(0) = \phi(t_2, \mathbf{x})$, and thus, by Picard's theorem and the definition of the flow, must coincide with $\phi(t, \phi(t_2, \mathbf{x}))$.

Remark 4.1 *Occasionally we could need solutions satisfying the initial condition at $t = t_0 \neq 0$. Then we should use the notation $\phi(t, t_0, \mathbf{x})(= \phi(t + t_0, \mathbf{x}))$ so that the first property above is $\phi(t_0, t_0, \mathbf{x}) = \mathbf{x}$.*

The next important notion is that of *equilibrium point* or *stationary solutions*. We note that if (2.4.26) has a solution that is constant in time, $\mathbf{x}(t) \equiv \mathbf{x}_0$, called a stationary solution, then such a solution satisfies $\mathbf{x}'(t) = 0$ and consequently

$$\mathbf{f}(\mathbf{x}_0) = 0. \tag{2.4.30}$$

Conversely, if the equation $\mathbf{f}(\mathbf{x}) = 0$ has a solution, which we call an equilibrium point, then, since $\mathbf{f}$ is independent of time, such a solution is a vector, say $\mathbf{y}_0$. If we consider now a function defined by $\mathbf{x}(t) = \mathbf{x}_0$, then we see that $\mathbf{x}'(t) \equiv 0$ and consequently

$$0 = \mathbf{x}'(t) = \mathbf{x}_0' = \mathbf{f}(\mathbf{y}_0)$$

and such a solution is a stationary solution. Summarizing, equilibrium points are solutions to algebraic equation (2.4.30) and, treated as constant functions, are (the only) stationary solutions to (3.2.6). Another way of looking at stationary solutions is that if we start the system from the initial state being an equilibrium point, then nothing will change – the system will stay in this state so that the solution will not depend on time. In terms of the flow, the equilibrium point can be defined as the point $\mathbf{x}_0$ such that

$$\phi(t, \mathbf{x}_0) = \mathbf{x}_0.$$

Example 4.1 Find all equilibrium values of the system of differential equations

$$\begin{aligned} y_1' &= 1 - y_2, \\ y_2' &= y_1^3 + y_2. \end{aligned}$$

We have to solve the system of algebraic equations

$$\begin{aligned} 0 &= 1 - y_2, \\ 0 &= y_1^3 + y_2. \end{aligned}$$

From the first equation we find $y_2 = 1$ and therefore $y_1^3 = -1$ which gives $y_1 = -1$ and the only equilibrium solution is

$$\mathbf{y^0} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Lemma 2.1, expressed in terms of the flow, states that the flow is a continuous function with respect to the initial condition.

Example 4.2 *Consider the scalar initial value problem*

$$y' = ay, \qquad y(t_0) = y_0.$$

In this case the solution can be found explicitly and the flow is given by

$$\phi(t, y_0) = y_0 \exp(a(t - t_0)),$$

and clearly

$$\phi(t, y_0) - \phi(t, y_1) = (y_0 - y_1) \exp at.$$

5 The phase space and orbits

In this section we shall give rudiments of the 'geometric' theory of differential equations. The aim of this theory is to obtain as complete a description as possible of all solutions of the autonomous system of differential equations (2.4.26) without solving it explicitly but by analysing geometric properties of its orbits. To explain the latter, we note that every solution $\mathbf{y} = (y_1(t), \ldots, y_n(t))$ defines a curve in the $n+1$-dimensional space $(t, y_1, \ldots, y_n)$.

Example 5.1 The solution $y_1(t) = \cos t$ and $y_2(t) = \sin t$ of the system

$$\begin{aligned} y_1' &= -y_2, \\ y_2' &= y_1 \end{aligned}$$

describes a helix in the (t, y_1, y_2) space.

The foundation of the geometric theory of differential equations is the observation that every solution $\mathbf{y}(t)$, $t_0 \le t \le t_1$, of (2.4.26), (2.4.27) also describes a curve in the n-dimensional space, that is, as t runs from t_0 to t_1, the points $(y_1(t), \dots, y_n(t))$ trace out a curve in this space. We call this curve the *integral curve* of (2.4.26). Typically we are interested more in the integral curve as the geometric object consisting of points visited by the solution. In such a case we talk about the *orbit*, or the *trajectory*, of the solution $\mathbf{y}(t)$. The formal definition is given below.

Definition 5.1 *The set*

$$\Gamma_{\mathbf{x}_0} = \{\mathbf{x} \in \mathbb{R}^n;\ \mathbf{x} = \phi(t, \mathbf{x}_0), t \in \mathbb{R}\}$$

is called the trajectory, or orbit, of the flow through $\mathbf{x}_0$. *If* $\mathbf{x}_0$ *plays no role in the considerations, we shall drop it from the notation. By positive (negative) half-orbit we understand the curve*

$$\Gamma^{\pm}_{\mathbf{x}_0} = \{\mathbf{x} \in \mathbb{R}^n;\ \mathbf{x} = \phi(t, \mathbf{x}_0), t \gtreqless 0\}.$$

The n*-dimensional* $\mathbf{y}$*-space, in which the orbits are situated, is called the* phase space *of the solutions of (2.4.26).*

Remark 5.1 *Note that the orbit of an equilibrium solution reduces to a point.*

Example 5.2 The solution of the previous example, $y_1(t) = \cos t, y_2(t) = \sin t$ traces out the unit circle $y_1^2 + y_2^2 = 1$ when t runs from 0 to 2π, hence the unit circle is the orbit of this solution. If t runs from 0 to ∞, then the pair $(\cos t, \sin t)$ trace out this circle infinitely often.

Example 5.3 Functions $y_1(t) = e^{-t}\cos t$ and $y_2(t) = e^{-t}\sin t$, $-\infty < t < \infty$, is a solution of the system

$$\begin{aligned} y_1' &= -y_1 - y_2, \\ y_2' &= y_1 - y_2. \end{aligned}$$

Since $r^2(t) = y_1^2(t) + y_2^2(t) = e^{-2t}$, we see that the orbit of this solution is a spiral traced towards the origin as t runs towards ∞.

One of the advantages of considering the orbit of the solution rather than the solution itself is that it is often possible to find the orbit explicitly without prior knowledge of the solution. We shall describe this for a two dimensional system. Let $(y_1(t), y_2(t))$ be a solution of (2.4.26) defined in a neighbourhood of a point $\bar{t}$. If e.g. $y_1'(\bar{t}) \neq 0$, then we can solve $y_1 = y_1(t)$ getting $t = t(y_1)$ in some neighbourhood of $\bar{y} = y_1(\bar{t})$. Thus, for t near $\bar{t}$, the orbit of the solution $(y_1(t), y_2(t))$ is given as the graph of $y_2 = y_2(t(y_1))$. Next, using the chain rule and the inverse function theorem

$$\frac{dy_2}{dy_1} = \frac{dy_2}{dt}\frac{dt}{dy_1} = \frac{y_2'}{y_1'} = \frac{f_2(y_1, y_2)}{f_1(y_1, y_2)}.$$

Thus, the orbits of the solution $y_1 = y_1(t), y_2(t) = y_2(t)$ of (2.4.26) are the solution curves of the first order scalar equation

$$\frac{dy_2}{dy_1} = \frac{f_2(y_1, y_2)}{f_1(y_1, y_2)} \tag{2.5.31}$$

and therefore to find the orbit of a solution there is no need to solve (2.4.26), (2.4.27); we have to solve the single first-order scalar equation (2.5.31).

The same procedure reduces a n-dimensional system to an $n-1$-dimensional in a neighbourhood of a point where, say, $f_i(y_1,\dots,y_n)\neq 0$:

$$\frac{dy_j}{dy_i}=\frac{f_j(\mathbf{y})}{f_i(\mathbf{y})},\qquad i\neq j=1,\dots,n.$$

Example 5.4 The orbits of the system of differential equations

$$\begin{aligned} y_1' &= y_2^2,\\ y_2' &= y_1^2. \end{aligned} \tag{2.5.32}$$

are the solution curves of the scalar equation $dy_2/dy_1=y_1^2/y_2^2$. This is a separable equation and it is easy to see that every solution is of the form $y_2=(y_1^3+c)^{1/3}$, c constant. Thus, the orbits are the curves $y_2=(y_1^3+c)^{1/3}$.

Example 5.5 A solution curve of (2.5.31) is an orbit of (2.4.26) if and only if $y_1'\neq 0$ and $y_2'\neq 0$ simultaneously along the solution. If a solution curve of (2.5.31) passes through an equilibrium point of (2.4.26), where $y_1'(\bar t)=0$ and $y_2'(\bar t)=0$ for some $\bar t$, then the entire solution curve is not a solution but rather it is a union of several distinct orbits. For example, consider the system of differential equations

$$\begin{aligned} y_1' &= y_2(1-y_1^2-y_2^2),\\ y_2' &= -y_1(1-y_1^2-y_2^2). \end{aligned} \tag{2.5.33}$$

The solution curves of the scalar equation

$$\frac{dy_2}{dy_1}=-\frac{y_1}{y_2}$$

are the family of concentric circles $y_1^2+y_2^2=c^2$. Observe however that to get the latter equation we should have assumed $y_1^2+y_2^2=1$ and that each point of this circle is an equilibrium point of (7.1.5). Thus, the orbits of (7.1.5) are the circles $y_1^2+y_2^2=c^2$ for $c\neq 1$ and each point of the unit circle is the orbit of a stationary solution.

Similarly, the full answer for the system (7.1.3) of the previous example is that $y_2=(y_1^3+c)^{1/3}$ are orbits for $c\neq 0$ as then neither solution curve passes through the only equilibrium point $(0,0)$. For $c=0$ the solution curve $y_2=y_1$ consists of the equilibrium point $(0,0)$ and two orbits $y_2=y_1$ for $y_1>0$ and $y_1<0$.

Note that in general it is impossible to solve (2.5.31) explicitly. Hence, usually we cannot find the equation of orbits in a closed form. Nevertheless, it is still possible to obtain an accurate description of all orbits of (2.4.26). In fact, the system (2.4.26) provides us with explicit information about how fast and in which direction solution is moving at each point of the trajectory. In fact, as the orbit of the solution $(y_1(t),y_2(t))$ is in fact a curve of which $(y_1(t),y_2(t))$ is a parametric description, $(y_1'(t),y_2'(t))=(f_1(y_1,y_2),f_2(y_1,y_2))$ is the tangent vector to the orbit at the point (y_1,y_2) showing, moreover, the direction at which the orbit is traversed. In particular, the orbit is vertical at each point (y_1,y_2) where $f_1(y_1,y_2)=0$ and $f_2(y_1,y_2)\neq 0$ and it is horizontal at each point (y_1,y_2) where $f_1(y_1,y_2)\neq 0$ and $f_2(y_1,y_2)=0$. As we noted earlier, each point (y_1,y_2) where $f_1(y_1,y_2)=0$ and $f_2(y_1,y_2)=0$ gives an equilibrium solution and the orbit reduces to this point.

6 Qualitative properties of orbits

We shall now prove two properties of orbits that are crucial to analyzing system (2.4.26).

Theorem 6.1 *Assume that the assumptions of Theorem 2.1 are satisfied. Then*

(i) there exists one and only one orbit through every point $\mathbf{x^0}\in\mathbb{R}^2$. In particular, if the orbits of two solutions $\mathbf{x}(t)$ and $\mathbf{y}(t)$ have one point in common, then they must be identical.

(ii) Let $\mathbf{x}(t)$ *be a solution to (2.4.26). If for some* $T > 0$ *and some* t_0 *we have* $\mathbf{x}(t_0 + T) = \mathbf{x}(t_0)$*, then* $\mathbf{x}(t + T) = \mathbf{x}(t)$ *for all* t*. In other words, if a solution* $\mathbf{x}(t)$ *returns to its starting value after a time* $T > 0$*, then in must be periodic (that is, it must repeat itself over every time interval of length* T*).*

Proof. ad (i) Let $\mathbf{x^0}$ be any point in $\mathbb{R}^2$. Then, from Theorem 2.1, we know that there is a solution of the problem $\mathbf{x}' = \mathbf{f}(\mathbf{x}), \mathbf{x}(0) = \mathbf{x^0}$ and the orbit of this solution passes through $\mathbf{x^0}$ from the definition of the orbit. Assume now that there is another orbit passing through $\mathbf{x^0}$, that is, there is a solution $\mathbf{y}(t)$ satisfying $\mathbf{y}(t_0) = \mathbf{x^0}$ for some t_0. From Lemma 4.1 we know that $\hat{\mathbf{y}}(t) = \mathbf{y}(t + t_0)$ is also a solution. However, this solution satisfies $\hat{\mathbf{y}}(0) = \mathbf{y}(t_0) = \mathbf{x^0}$, that is, the same initial condition as $\mathbf{x}(t)$. By the uniqueness part of Theorem 2.1 we must then have $\mathbf{x}(t) = \hat{\mathbf{y}}(t) = \mathbf{y}(t + t_0)$ for all t for which the solutions are defined. This implies that the orbits are identical. In fact, if ξ is an element of the orbit of $\mathbf{x}$, then for some t' we have $\mathbf{x}(t') = \xi$. However, we have also $\xi = \mathbf{y}(t' + t_0)$ so that ξ belongs to the orbit of $\mathbf{y}(t)$. Conversely, if ξ belongs to the orbit of $\mathbf{y}$ so that $\xi = \mathbf{y}(t'')$ for some t'', then by $\xi = \mathbf{y}(t'') = \mathbf{x}(t'' - t_0)$, we see that ξ belongs to the orbit of $\mathbf{x}$.

ad (ii) Assume that for some numbers t_0 and $T > 0$ we have $\mathbf{x}(t_0) = \mathbf{x}(t_0 + T)$. The function $\mathbf{y}(t) = \mathbf{x}(t + T)$ is again a solution satisfying $\mathbf{y}(t_0) = \mathbf{x}(t_0 + T) = \mathbf{x}(t_0)$, thus from Theorem 2.1, $\mathbf{x}(t) = \mathbf{y}(t)$ for all t for which they are defined and therefore $\mathbf{x}(t) = \mathbf{x}(t + T)$ for all such t. ■

Example 6.1 A curve in the shape of a figure 8 cannot be an orbit. In fact, suppose that the solution passes through the intersection point at some time t_0, then completing the first loop returns after time T, that is, we have $\mathbf{x}(t_0) = \mathbf{x}(t_0 + T)$. From (ii) it follows then that this solution is periodic, that is, it must follow the same loop again and cannot switch to the other loop.

Corollary 6.1 *A solution* $\mathbf{x}(t)$ *of (2.4.26) is periodic if and only if its orbit is a closed curve in* $\mathbb{R}$*.*

Proof. Assume that $\mathbf{x}(t)$ is a periodic solution of (2.4.26) of period T, that is $\mathbf{x}(t) = t + T$. If we fix t_0, then, as t runs from t_0 to $t_0 + T$, the point $\mathbf{x}(t) = (x_1(t), x_2(t))$ traces a curve, say C, from $\xi = \mathbf{x}(t)$ back to the same point ξ without intersections and, if t runs from $-\infty$ to ∞, the curve C is traced infinitely many times.

Conversely, suppose that the orbit C is a closed curve (containing no equilibrium points). The orbit is parametrically described by $\mathbf{x}(t)$, $-\infty < t < \infty$ in a one-to-one way (as otherwise we would have $\mathbf{x}(t') = \mathbf{x}(t'')$ for some $t' \neq t''$ and, by the previous theorem, the solution would be periodic). Consider a sequence $(t_n)_{n\in\mathbb{N}}$ with $t_n \to \infty$. Since the sequence $\mathbf{x}(t)$ is bounded, we find a subsequence $t'_n \to \infty$ such that $\mathbf{x}(t'_n) \to \mathbf{x} \in C$. Then, however, $\mathbf{x} = \mathbf{x}(t_0)$ for some finite t_0 since C does not contain equilibria. Consider a neighbourhood of $\mathbf{x}(t_0)$ which is the image of some interval $(t_0 - \epsilon, t_0 + \epsilon)$. Since $\mathbf{x}(t'_n) \to \mathbf{x}(t_0)$, $t'_n \in (t_0 - \epsilon, t_0 + \epsilon)$ for sufficiently large n which contradicts $t'_n \to \infty$. ■

We conclude the theoretical part by presenting another result showing how an orbit can/cannot look like.

Proposition 6.1 *Suppose that a solution* $\mathbf{y}(t)$ *of (2.4.26) approaches a vector* $\mathbf{v}$ *as* $t \to \infty$*. Then* $\mathbf{v}$ *is an equilibrium point of (2.4.26).*

Proof. $\lim_{t\to\infty} \mathbf{y}(t) = \mathbf{v}$ is equivalent to $\lim_{t\to\infty} y_i(t) = v_i$, $i = 1, \ldots, n$. This implies $\lim_{t\to\infty} y_i(t + h) = v_i$ for any fixed h. Using the mean value theorem we have

$$y_i(t + h) - y_i(t) = hy'_i(\tau) = hf_i(y_1(\tau), \ldots, y_n(\tau)),$$

where $\tau \in [t, t + h]$. If $t \to \infty$, then also $\tau \to \infty$ and passing to the limit in the above equality, we obtain

$$0 = v_i - v_i = hf_i(v_1, \ldots, v_n), \qquad i = 1, \ldots, n,$$

so that $\mathbf{v}$ is an equilibrium point. ■

Example 6.2 Show that every solution $z(t)$ of the second order differential equation

$$z'' + z + z^3 = 0$$

is periodic. We convert this equation into a system: let $z = x_1$ and

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -x_1 - x_1^3. \end{aligned}$$

The orbits are the solution curves of the equation

$$\frac{dx_2}{dx_1} = -\frac{x_1 + x_1^3}{x_2},$$

so that

$$\frac{x_2^2}{2} + \frac{x_1^2}{2} + \frac{x_1^4}{4} = c^2$$

is the equation of orbits. If $c \neq 0$, then neither of them contains the unique equilibrium point $(0, 0)$. By writing the above equation in the form

$$\frac{x_2^2}{2} + \left(\frac{x_1^2}{2} + \frac{1}{2}\right)^2 = c^2 + \frac{1}{4}$$

we see that for each $c \neq 0$ it describes a closed curve consisting of two branches $x_2 = \pm\frac{1}{\sqrt{2}}\sqrt{4c^2 + 1 - (x^2+1)^2}$ that stretch between $x_1 = \pm\sqrt{1 + \sqrt{4c^2+1}}$. Consequently, every solution is a periodic function.

7 Applications of the phase-plane analysis

In this section we shall present typical techniques of phase plane analysis to determine long time behaviour of solutions.

Example 7.1 Consider the system of differential equations

$$\begin{aligned} x' &= ax - by - ex^2, \\ y' &= -cy + dxy - fy^2, \end{aligned} \tag{2.7.34}$$

where a, b, e, c, d, f are positive constants. This system can describe the population growth of two species x and y is an environment of limited capacity, where the species y depends on the species x for its survival. Assume that $c/d > a/e$. We prove that every solution $(x(t), y(t))$ of (2.7.34), with $x(0), y(0) > 0$ approaches the equilibrium solution $x = a/e, y = 0$, as t approaches infinity. As a first step, we show that the solutions with positive initial data must stay positive, that is, the orbit of any solution originating in the first quadrant must stay in this quadrant. Otherwise the model would not correspond to reality. First, let us observe that putting $y(t) \equiv 0$ we obtain the logistic equation for x that can be solved giving

$$x(t) = \frac{ax_0}{ex_0 + (a - ex_0)\exp(-at)}$$

where $x_0 \geq 0$. The orbits of these solutions is the equilibrium point $(0, 0)$, the segment $0 \leq x < a/e, y = 0$ for $x_0 < a/e$, the equilibrium point $(a/e, 0)$ for $x_0 = a/e$ and the segments $a/e < x < \infty, y = 0$ for $x > a/e$. Thus, the positive x-semiaxis $x \geq 0$ is the union of these four orbits. Similarly, putting $x(t) \equiv 0$ we obtain the equation

$$y' = -cy - fy^2.$$

To use the theory of the first chapter, we observe that the equilibrium points of this equation are $y = 0$ and $y = -c/f$ so that there are no equilibria on the positive y-semiaxis and $-cy - fy^2 < 0$ for $y > 0$. Therefore

Fig.5 Regions described in Example 7.1

any solution with initial value $y_0 > 0$ will decrease converging to 0 and the semiaxis $y > 0$ is a single orbit of (2.7.34). Thus, if a solution of (2.7.34) left the first quadrant, its orbit would cross one of the orbits the positive semiaxes consist of which is precluded by uniqueness of orbits.

In the next step we divide the first quadrant into regions where the derivatives x' and y' are of a fixed sign. This is done by drawing lines l_1 and l_2, as in Fig. 5, across which one of the other derivative vanishes. The line l_1 is determined by $-ex/b + a/b$ so that $x' > 0$ in region I and $x' <$ in regions II and III. The line l_2 is given by $y = dx/f - c/f$ and $y' < 0$ in regions I and II, and $y' > 0$ in region III.

We describe the behaviour of solutions in each regions in the sequence of observations.

Observation 1. *Any solution to (2.7.34) which starts in the region I at $t = t_0$ will remain in this region for all $t > t_0$ and ultimately approach the equilibrium $x = a/e, y = 0$.*

Proof. If the solution $x(t), y(t)$ leaves region I at some time $t = t^*$, then $x'(t^*) = 0$, since the only way to leave this region is to cross the line l_1. Differentiation the first equation in (2.7.34) gives

$$x'' = ax' - bx'y - bxy' - 2exx'$$

so that at $t = t^*$ we obtain

$$x''(t^*) = -bx(t^*)y'(t^*).$$

Since $y'(t^*) < 0$, $x''(t^*) > 0$ which means that $x(t^*)$ is a local minimum. However, $x(t)$ reaches this point from region I where it is increasing, which is a contradiction. Thus, the solution $x(t), y(t)$ stays in region I for all times $t \geq t_0$. However, any solution staying in I must be bounded and $x' > 0$ and $y' < 0$ so that $x(t)$ is increasing and $y(t)$ is decreasing and therefore they must tend to a finite limit. By Proposition 6.1, this limit must be an equilibrium point. The only equilibria are $(0, 0)$ and $(a/e, 0)$ and the solution cannot tend to the former as $x(t)$ is positive and increasing. Thus, any solution starting in region I for some time $t = t_0$ tends to the equilibrium $(a/e, 0)$ as $t \to \infty$.

Observation 2. *Any solution of (2.7.34) that starts in region III at time t_0 must leave this region at some later time.*

Proof. Suppose that a solution $x(t), y(t)$ stays in region III for all time $t \geq 0$. Since the sign of both derivatives x' and y' is fixed, $x(t)$ decreases and $y(t)$ increases, thus $x(t)$ must tend to a finite limit. $y(t)$

cannot escape to infinity as the only way it could be achieved would be if also $x(t)$ tended to infinity, which is impossible. Thus, $x(t), y(t)$ tend to a finite limit that, by Proposition 6.1, has to be an equilibrium. However, there are no equilibria to be reached from region III and thus the solution must leave this region at some time.

Observation 3. *Any solution of (2.7.34) that starts in region II at time $t = t_0$ and remains in this region for all $t \geq 0$ must approach the equilibrium solution $x = a/e, y = 0$.*

Proof. Suppose that a solution $x(t), y(t)$ stays in region II for all $t \geq t_0$. Then both $x(t)$ and $y(t)$ are decreasing and, since the region is bounded from below, we see that this solution must converge to an equilibrium point, in this case necessarily $(a/e, 0)$.

Observation 4. *A solution cannot enter region III from region II.*

Proof. This case is similar to Observation 1. Indeed, if the solution crosses l_2 from II to III at $t = t^*$, then $y'(t^*) = 0$ but then, from the second equation of (2.7.34)

$$y''(t^*) = dy(t^*)x'(t^*) < 0$$

so that $y(t^*)$ is a local maximum. This is, however, impossible, as $y(t)$ is decreasing in region II.

Summarizing, if the initial values are in regions I or II, then the solution tends to the equilibrium $(a/e, 0)$ as $t \to \infty$, by Observations 1,3 and 4. If the solution starts from region III, then at some point it must enter region II and we can apply the previous argument to claim again that the solution will eventually approach the equilibrium $(a/e, 0)$. Finally, if a solution starts on l_1, it must immediately enter region I as $y' < 0$ and $x' < 0$ in region II (if the solution ventured into II from l_1, then either x' or y' would have to be positive somewhere in II). Similarly, any solution starting from l_2 must immediately enter II. Thus, all the solution starting in the first quadrant (with strictly positive initial data) will converge to $(a/e, 0)$ as $t \to \infty$.

Example 7.2 Lotka-Volterra model. In this example we shall discuss the predator-prey model introduced by Lotka and Volterra. It reads

$$\begin{aligned} \frac{dx_1}{dt} &= (r - f)x_1 - \alpha x_1 x_2, \\ \frac{dx_2}{dt} &= -(s + f)x_2 + \beta x_1 x_2 \end{aligned} \tag{2.7.35}$$

where α, β, r, s, f are positive constants. In the predator-prey model x_1 is the density of the prey, x_2 is the density of the predators, r is the growth rate of the prey in the absence of predators, $-s$ is the growth rate of predators in the absence of prey (the population of predators dies out without the supply of the sole food source – prey). The quadratic terms account for predator–prey interaction and f represents indiscriminate killing of both prey and predators. The model was introduced in 1920s by Italian mathematician Vito Volterra to explain why, in the period of reduced (indiscriminate) fishing, the relative number predators (sharks) significantly increased.

Let us consider first the model without fishing

$$\begin{aligned} \frac{dx_1}{dt} &= rx_1 - \alpha x_1 x_2, \\ \frac{dx_2}{dt} &= -sx_2 + \beta x_1 x_2 \end{aligned} \tag{2.7.36}$$

Observe that there are two equilibrium solution $x_1(t) = 0, x_2(t) = 0$ and $x_1(t) = s/\beta, x_2(t) = r/\alpha$. The first solution is not interesting as it corresponds to the total extinction. We observe also that we have two other solutions $x_1(t) = c_1 e^{rt}, x_2(t) = 0$ and $x_1(t) = 0, x_2(t) = c_2 e^{-st}$ that correspond to the situation when one of the species is extinct. Thus, both positive x_1 and x_2 semi-axes are orbits and, by Theorem 6.1 (i), any orbit starting in the first quadrant will stay there or, in other words, any solution with positive initial data will remain strictly positive for all times.

The orbits of (2.7.36) are the solution curves of the first order separable equation

$$\frac{dx_2}{dx_1} = \frac{x_2(-s+\beta x_1)}{x_1(r-\alpha x_2)} \tag{2.7.37}$$

Separating variables and integrating we obtain

$$r\ln x_2 - \alpha x_2 + s\ln x_1 - \beta x_1 = k$$

which can be written as

$$\frac{x_2^r}{e^{\alpha x_2}}\frac{x_1^s}{e^{\beta x_1}} = K. \tag{2.7.38}$$

Next prove that the curves defined by (2.7.38) are closed. It is not an easy task. To prove this we shall show that for each x_1 from a certain open interval $(x_{1,m}, x_{1,M})$ we have exactly two solutions $x_{2,m}(x_1)$ and $x_{2,M}(x_1)$ and that these two solutions tend to common limits as x_1 approaches $x_{1,m}$ and $x_{1,M}$.

First, let as define $f(x_2) = x_2^r e^{-\alpha x_2}$ and $g(x_1) = x_1^s e^{-\beta x_1}$. We shall analyze only f as g is of the same for. Due to positivity of all the coefficients, we see that $f(0) = 0$ also $\lim_{x_2\to\infty} f(x_2) = 0$ and $f(x_2) > 0$ for $x_2 > 0$. Further

$$f'(x_2) = x_2^{r-1}e^{-\alpha x_2}(r-\alpha x_2),$$

so that f is increasing from 0 to $x_2 = r/\alpha$ where it attains global maximum, say M_2, and then starts to decrease monotonically to 0. Similarly, $g(0) = \lim_{x_1\to\infty} g(x_1) = 0$ and $g(x_1) > 0$ for $x_1 > 0$ and it attains global maximum M_1 at $x_1 = s/b$. We have to analyze solvability of

$$f(x_2)g(x_1) = K.$$

Firstly, there are no solutions if $K > M_1M_2$, and for $K = M_1M_2$ we have the equilibrium solution $x_1 = s/\beta, x_2 = r/\alpha$. Thus, we have to consider $K = \lambda M_2$ with $\lambda < 1$. Let us write this equation as

$$f(x_2) = \frac{\lambda}{g(x_1)}M_2. \tag{2.7.39}$$

From the shape of the graph g we find that the equation $g(x_1) = \lambda$ has no solution if $\lambda > M_1$ but then $\lambda/g(x_1) \geq \lambda/M_1 > 1$ so that (2.7.39) is not solvable. If $\lambda = M_1$, then we have again the equilibrium solution. Finally, for $\lambda < M_1$ there are two solutions $x_{1,m}$ and $x_{1,M}$ satisfying $x_{1,m} < s/\beta < x_{1,M}$. Now, for x_1 satisfying $x_{1,m} < x_1 < x_{1,M}$ we have $\lambda/g(x_1) < 1$ and therefore for such x_1 equation (2.7.39) has two solutions $x_{2,m}(x_1)$ and $x_{2,M}(x_1)$ satisfying $x_{2,m} < r/\alpha < x_{2,M}$, again on the basis of the shape of the graph of f. Moreover, if x_1 moves towards either $x_{1,m}$ or $x_{1,M}$, then both solutions $x_{2,m}$ and $x_{2,M}$ move towards r/α, that is the set of points satisfying (2.7.39) is a closed curve.

Summarizing, the orbits are closed curves encircling the equilibrium solution $(s/\beta, r/\alpha)$ and are traversed in the anticlockwise direction. Thus, the solutions are periodic in time. The evolution can be described as follows. Suppose that we start with initial values $x_1 > s/\beta, x_2 < r/\alpha$, that is, in the lower right quarter of the orbit. Then the solution will move right and up till prey population reaches maximum $x_{1,M}$. Because there is a lot of prey, the number of predator will be still growing but then the number of prey will start decreasing slowing down the growth of the predator's population. The decrease in the prey population will eventually bring the growth of predator's population to stop at the maximum $x_{2,M}$. From now on the number of predators will decrease but the depletion of the prey population from the previous period will continue to prevail till the population reaches the minimum $x_{1,m}$ when it will start to take advantage of the decreasing number of predators and will start to grow; this growth will, however, slow down when the population of predators will reach its minimum. However, the the number of prey will be increasing beyond the point when the number of predators is the least till the growing number of predators eventually cause the prey population to decrease having reached its peak at $x_{1,M}$ and the cycle will repeat itself.

Now we are ready to provide the explanation of the observational data. Including fishing into the model, according to (**??**), amounts to changing parameters r and s to $r-f$ and $s+f$ but the structure of the system does not change, so that the equilibrium solution becomes

$$\left(\frac{s+f}{\beta}, \frac{r-f}{\alpha}\right).$$

Thus, with a moderate amount of fishing ($f < r$), the in the equilibrium solution there is more fish and less sharks in comparison with no-fishing situation. Thus, if we reduce fishing, the equilibrium moves towards larger amount of sharks and lower amount of fish. Of course, this is true for equilibrium situation, which not necessarily corresponds to reality, but as the orbits are closed curves around the equilibrium solution, we can expect that the amounts of fish and sharks in a non-equilibrium situation will change in a similar pattern. We can confirm this hypothesis by comparing average numbers of sharks and fish over the full cycle. For any function f its average over an interval (a, b) is defined as

$$\bar{f} = \frac{1}{b-a}\int_a^b f(t)dt,$$

so that the average numbers if fish and sharks over one cycle is given by

$$\overline{x_1} = \frac{1}{T}\int_0^T x_1(t)dt, \qquad \overline{x_2} = \frac{1}{T}\int_0^T x_2(t)dt.$$

It turns out that these averages can be calculated explicitly. Dividing the first equation of (2.7.36) by x_1 gives $x_1'/x_1 = r - \alpha x_2$. Integrating both sides, we get

$$\frac{1}{T}\int_0^T \frac{x_1'(t)}{x_1(t)}dt = \frac{1}{T}\int_0^T (r - \alpha x_2(t))dt.$$

The left-hand side can be evaluated as

$$\int_0^T \frac{x_1'(t)}{x_1(t)}dt = \ln x_1(T) - \ln x_1(0) = 0$$

on account of the periodicity of x_1. Hence,

$$\frac{1}{T}\int_0^T (r - \alpha x_2(t))dt = 0,$$

and

$$\overline{x_2} = \frac{r}{\alpha}. \tag{2.7.40}$$

In the same way,

$$\overline{x_1} = \frac{s}{\beta}, \tag{2.7.41}$$

so that the average values of x_1 and x_2 are exactly the equilibrium solutions. Thus, we can state that introducing fishing is more beneficial to prey than predators as the average numbers of prey increases while the average number of predators will decrease in accordance with (**??**), while reducing fishing will have the opposite effect of increasing the number of predators and decreasing the number of prey.

8 Solvability of linear systems

We shall consider only linear systems of first order differential equations.

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n + g_1(t), \\ &\vdots \quad \vdots \quad \vdots, \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n + g_n(t), \end{aligned} \tag{2.8.42}$$

where $y_1, \ldots, y_n$ are unknown functions, $a_{11}, \ldots a_{nn}$ are constant coefficients and $g_1(t) \ldots, g_n(t)$ are known continuous functions. If $g_1 = \ldots = g_n = 0$, then the corresponding system (2.8.42) is called the associated homogeneous system. The structure of (2.8.42) suggest that a more economical way of writing is is to use the vector-matrix notation. Denoting $\mathbf{y} = (y_1, \ldots, y_n)$, $\mathbf{g} = (g_1, \ldots, g_n)$ and $\mathcal{A} = \{a_{ij}\}_{1 \leq i,j \leq n}$, that is

$$\mathcal{A} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix},$$

we can write (2.8.42) is a more concise notation as

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}. \tag{2.8.43}$$

Here we have n unknown functions and the system involves first derivative of each of them so that it is natural to consider (2.8.43) in conjunction with the following initial conditions

$$\mathbf{y}(t_0) = \mathbf{y^0}, \tag{2.8.44}$$

or, in the expanded form,

$$y_1(t_0) = y_1^0, \ldots, y_n(t_0) = y_n^0, \tag{2.8.45}$$

where t_0 is a given argument and $\mathbf{y^0} = (y_1^0, \ldots, y_n^0)$ is a given vector.

Let us denote by $\mathbf{X}$ the set of all solutions to the homogeneous system (2.8.42). Due to linearity of differentiation and multiplication by $\mathcal{A}$, it is easy to see that $\mathbf{X}$ is a vector space. We have two fundamental results.

Theorem 8.1 *The dimension of* $\mathbf{X}$ *is equal to* n.

Theorem 8.2 *Let* $\mathbf{y_1}, \ldots, \mathbf{y_k}$ *be* k *linearly independent solutions of* $\mathbf{y}' = \mathcal{A}\mathbf{y}$ *and let* $t_0 \in \mathbb{R}$ *be an arbitrary number. Then,* $\{\mathbf{y_1}(t), \ldots, \mathbf{y_k}(t)\}$ *for a linearly independent set of functions if and only if* $\{\mathbf{y_1}(t_0), \ldots, \mathbf{y_k}(t_0)\}$ *is a linearly independent set of vectors in* $\mathbb{R}$.

These two results show that if we construct solutions emanating from n linearly independent initial vectors, then these solutions are linearly independent and therefore they span the space of all solutions to the homogeneous system (2.8.42).

Let $\mathcal{A}$ be an $n \times n$ matrix. We say that a number λ (real or complex) is an *eigenvalue* of $\mathcal{A}$ is there exist a non-zero solution of the equation

$$\mathcal{A}\mathbf{v} = \lambda\mathbf{v}. \tag{2.8.46}$$

Such a solution is called an *eigenvector* of $\mathcal{A}$. The set of eigenvectors corresponding to a given eigenvalue is a vector subspace. Eq. (2.8.46) is equivalent to the homogeneous system $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$, where $\mathcal{I}$ is the identity matrix, therefore λ is an eigenvalue of $\mathcal{A}$ if and only if the determinant of $\mathcal{A}$ satisfies

$$det(\mathcal{A} - \lambda\mathcal{I}) = \begin{vmatrix} a_{11} - \lambda & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{2.8.47}$$

Evaluating the determinant we obtain a polynomial in λ of degree n. This polynomial is also called the characteristic polynomial of the system (2.8.42) (if (2.8.42) arises from a second order equation, then this is the same polynomial as the characteristic polynomial of the equation). We shall denote this polynomial by $p(\lambda)$. From algebra we know that there are exactly n, possibly complex, root of $p(\lambda)$. Some of them may be multiple, so that in general $p(\lambda)$ factorizes into

$$p(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdot \ldots \cdot (\lambda_k - \lambda)^{n_k}, \tag{2.8.48}$$

with $n_1+\ldots+n_k=n$. It is also worthwhile to note that since the coefficients of the polynomial are real, then complex roots appear always in conjugate pairs, that is, if $\lambda_j=\xi_j+i\omega_j$ is a characteristic root, then so is $\bar{\lambda}_j=\xi_j-i\omega_j$. Thus, eigenvalues are roots of the characteristic polynomial of $\mathcal{A}$. The exponent n_i appearing in the factorization (2.8.48) is called the *algebraic multiplicity* of λ_i. For each eigenvalue λ_i there corresponds an eigenvector $\mathbf{v_i}$ and eigenvectors corresponding to distinct eigenvalues are linearly independent. The set of all eigenvectors corresponding to λ_i spans a subspace, called the *eigenspace* corresponding to λ_i which we will denote by E_{λ_i}. The dimension of E_{λ_i} is called the *geometric multiplicity* of λ_i. In general, algebraic and geometric multiplicities are different with geometric multiplicity being at most equal to the algebraic one. Thus, in particular, if λ_i is a single root of the characteristic polynomial, then the eigenspace corresponding to λ_1 is one-dimensional.

If the geometric multiplicities of eigenvalues add up to n, that is, if we have n linearly independent eigenvectors, then these eigenvectors form a basis for $\mathbb{R}^n$. In particular, this happens if all eigenvalues are single roots of the characteristic polynomial. If this is not the case, then we do not have sufficiently many eigenvectors to span $\mathbb{R}^n$ and if we need a basis for $\mathbb{R}^n$, then we have to find additional linearly independent vectors. A procedure that can be employed here and that will be very useful in our treatment of systems of differential equations is to find solutions to equations of the form $(\mathcal{A}-\lambda_i\mathcal{I})^k\mathbf{v}=0$ for $1<k\leq n_i$, where n_i is the algebraic multiplicity of λ_i. Precisely speaking, if λ_i has algebraic multiplicity n_i and if

$$(\mathcal{A}-\lambda_i\mathcal{I})\mathbf{v}=0$$

has only $\nu_i<n_i$ linearly independent solutions, then we consider the equation

$$(\mathcal{A}-\lambda_i\mathcal{I})^2\mathbf{v}=0.$$

It follows that all the solutions of the preceding equation solve this equation but there is at least one more independent solution so that we have at least ν_i+1 independent vectors (note that these new vectors are no longer eigenvectors). If the number of independent solutions is still less than n_1, we consider

$$(\mathcal{A}-\lambda_i\mathcal{I})^3\mathbf{v}=0,$$

and so on, till we get a sufficient number of them. Note, that to make sure that in the step j we select solutions that are independent of the solutions obtained in step $j-1$ it is enough to find solutions to $(\mathcal{A}-\lambda_i\mathcal{I})^j\mathbf{v}=0$ that satisfy $(\mathcal{A}-\lambda_i\mathcal{I})^{j-1}\mathbf{v}\neq 0$.

Matrix exponentials

The above theory can be used to provide a unified framework for solving systems of differential equations.

Recall that for a single equation $y'=ay$, where a is a constant, the general solution is given by $y(t)=e^{at}C$, where C is a constant. In a similar way, we would like to say that the general solution to

$$\mathbf{y}'=\mathcal{A}\mathbf{y},$$

where $\mathcal{A}$ is an $n\times n$ matrix, is $\mathbf{y}=e^{\mathcal{A}t}\mathbf{v}$, where $\mathbf{v}$ is any constant vector in $\mathbb{R}^n$. The problem is that we do not know what it means to evaluate an exponential of a matrix. However, if we reflect for a moment that the exponential of a number can be evaluated as the power (Maclaurin) series

$$e^x=1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\ldots+\frac{x^k}{k!}+\ldots,$$

where the only involved operations on the argument x are additions, scalar multiplications and taking integer powers, we come to the conclusion that the above expression can be written also for a matrix, that is, we can define

$$e^{\mathcal{A}}=\mathcal{I}+\mathcal{A}+\frac{1}{2}\mathcal{A}^2+\frac{1}{3!}\mathcal{A}^3+\ldots+\frac{1}{k!}\mathcal{A}^k+\ldots. \tag{2.8.49}$$

It can be shown that if $\mathcal{A}$ is a matrix, then the above series always converges and the sum is a matrix. For example, if we take

$$\mathcal{A}=\begin{pmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{pmatrix}=\lambda\mathcal{I},$$

then

$$\mathcal{A}^k = \lambda^k \mathcal{I}^k = \lambda^k \mathcal{I},$$

and

$$\begin{aligned} e^{\mathcal{A}} &= \mathcal{I} + \lambda\mathcal{I} + \frac{\lambda^2}{2}\mathcal{I} + \frac{\lambda^3}{3!}\mathcal{I} + \ldots + \frac{\lambda^k}{k!} + \ldots \\ &= \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \ldots + \frac{\lambda^k}{k!} + \ldots\right)\mathcal{I} \\ &= e^{\lambda}\mathcal{I}. \end{aligned} \tag{2.8.50}$$

Unfortunately, in most cases finding the explicit form for $e^{\mathcal{A}}$ directly is impossible.

Matrix exponentials have the following algebraic properties

$$\left(e^{\mathcal{A}}\right)^{-1} = e^{-\mathcal{A}}$$

and

$$e^{\mathcal{A}+\mathcal{B}} = e^{\mathcal{A}} e^{\mathcal{B}} \tag{2.8.51}$$

provided the matrices $\mathcal{A}$ and $\mathcal{B}$ commute: $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$.

Let us define a function of t by

$$e^{t\mathcal{A}} = \mathcal{I} + t\mathcal{A} + \frac{t^2}{2}\mathcal{A}^2 + \frac{t^3}{3!}\mathcal{A}^3 + \ldots + \frac{t^k}{k!}\mathcal{A}^k + \ldots\ . \tag{2.8.52}$$

It follows that this function can be differentiated with respect to t by termwise differentiation of the series, as in the scalar case, that is,

$$\begin{aligned} \frac{d}{dt}e^{\mathcal{A}t} &= \mathcal{A} + t\mathcal{A}^2 + \frac{t^2}{2!}\mathcal{A}^3 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^k + \ldots \\ &= \mathcal{A}\left(\mathcal{I} + t\mathcal{A} + \frac{t^2}{2!}\mathcal{A}^2 + \ldots + \frac{t^{k-1}}{(k-1)!}\mathcal{A}^{k-1} + \ldots\right) \\ &= \mathcal{A}e^{t\mathcal{A}} = e^{t\mathcal{A}}\mathcal{A}, \end{aligned}$$

proving thus that $\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{v}$ is a solution to our system of equations for any constant vector $\mathbf{v}$. Since $\mathbf{y}(0) = e^{0\mathcal{A}}\mathbf{v} = \mathbf{v}$, from Picard's theorem $\mathbf{y}(t)$ is a unique solution to the Cauchy problem

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{v}.$$

As we mentioned earlier, in general it is difficult to find directly the explicit form of $e^{t\mathcal{A}}$. However, we can always find n linearly independent vectors $\mathbf{v}$ for which the series $e^{t\mathcal{A}}\mathbf{v}$ can be summed exactly. This is based on the following two observations. Firstly, since $\lambda\mathcal{I}$ and $\mathcal{A} - \lambda\mathcal{I}$ commute, we have by (2.8.50) and (2.8.51)

$$e^{t\mathcal{A}}\mathbf{v} = e^{t(\mathcal{A}-\lambda\mathcal{I})}e^{t\lambda\mathcal{I}}\mathbf{v} = e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v}.$$

Secondly, if $(\mathcal{A} - \lambda\mathcal{I})^m\mathbf{v} = \mathbf{0}$ for some m, then

$$(\mathcal{A} - \lambda\mathcal{I})^r\mathbf{v} = \mathbf{0}, \tag{2.8.53}$$

for all $r \geq m$. This follows from

$$(\mathcal{A} - \lambda\mathcal{I})^r\mathbf{v} = (\mathcal{A} - \lambda\mathcal{I})^{r-m}[(\mathcal{A} - \lambda\mathcal{I})^m\mathbf{v}] = \mathbf{0}.$$

Consequently, for such a $\mathbf{v}$

$$e^{t(\mathcal{A}-\lambda\mathcal{I})}\mathbf{v} = \mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} + \ldots + \frac{t^{m-1}}{(m-1)!}(\mathcal{A} - \lambda\mathcal{I})^{m-1}\mathbf{v}.$$

and

$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}e^{t(\mathcal{A}-\lambda\mathcal{I})} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A}-\lambda\mathcal{I})\mathbf{v} + \ldots + \frac{t^{m-1}}{(m-1)!}(\mathcal{A}-\lambda\mathcal{I})^{m-1}\mathbf{v}\right). \tag{2.8.54}$$

Thus, to find all solutions to $\mathbf{y}' = \mathcal{A}\mathbf{y}$ it is sufficient to find n independent vectors $\mathbf{v}$ satisfying (2.8.53) for some scalars λ. But these are precisely the eigenvectors or associated eigenvectors and we know that it is possible to find exactly n of them.

Thus, for example, if $\lambda = \lambda_1$ is a simple eigenvalue of $\mathcal{A}$ with a corresponding eigenvector $\mathbf{v^1}$, then $(\mathcal{A} - \lambda_1\mathcal{I})\mathbf{v^1} = 1$, thus m of (2.8.53) is equal to 1. Consequently, the sum in (2.8.54) terminates after the first term and we obtain

$$\mathbf{y^1}(t) = e^{\lambda_1}\mathbf{v^1}.$$

From our discussion of eigenvalues and eigenvectors it follows that if λ_i is a multiple eigenvalue of $\mathcal{A}$ of algebraic multiplicity n_i and the geometric multiplicity is less then n_i, that is, there is less than n_i linearly independent eigenvectors corresponding to λ_i, then the missing independent vectors can be found by solving successively equations $(\mathcal{A} - \lambda_i\mathcal{I})^k\mathbf{v} = \mathbf{0}$ with k running at most up to n_1. Thus, we have the following algorithm for finding n linearly independent solutions to $\mathbf{y}' = \mathcal{A}\mathbf{y}$:

1. Find all eigenvalues of $\mathcal{A}$;
2. If λ is a single real eigenvalue, then there is an eigenvector $\mathbf{v}$ so that the solution is given by
$$\mathbf{y}(t) = e^{\lambda t}\mathbf{v} \tag{2.8.55}$$
3. If λ is a single complex eigenvalue $\lambda = \xi + i\omega$, then there is a complex eigenvector $\mathbf{v} = \Re\mathbf{v} + i\Im\mathbf{v}$ such that two solutions corresponding to λ (and $\bar{\lambda}$) are given by
$$\begin{aligned}\mathbf{y^1}(t) &= e^{\xi t}(\cos\omega t\,\Re\mathbf{v} - \sin\omega t\,\Im\mathbf{v})\\ \mathbf{y^2}(t) &= e^{\xi t}(\cos\omega t\,\Im\mathbf{v} + \sin\omega t\Re\mathbf{v})\end{aligned} \tag{2.8.56}$$
4. If λ is a multiple eigenvalue with algebraic multiplicity k (that is, λ is a multiple root of the characteristic equation, of multiplicity k), then we first find eigenvectors by solving $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} = \mathbf{0}$. For these eigenvectors the solution is again given by (2.8.55) (or (2.8.56), if λ is complex). If we found k independent eigenvectors, then our work with this eigenvalue is finished. If not, then we look for vectors that satisfy $(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v} = \mathbf{0}$ but $(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} \neq \mathbf{0}$. For these vectors we have the solutions
$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v}\right).$$
If we still do not have k independent solutions, then we find vectors for which $(\mathcal{A} - \lambda\mathcal{I})^3\mathbf{v} = \mathbf{0}$ and $(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v} \neq \mathbf{0}$, and for such vectors we construct solutions
$$e^{t\mathcal{A}}\mathbf{v} = e^{\lambda t}\left(\mathbf{v} + t(\mathcal{A} - \lambda\mathcal{I})\mathbf{v} + \frac{t^2}{2}(\mathcal{A} - \lambda\mathcal{I})^2\mathbf{v}\right).$$
This procedure is continued till we have k solutions (by the properties of eigenvalues we have to repeat this procedure at most k times).
If λ is a complex eigenvalue of multiplicity k, then also $\bar{\lambda}$ is an eigenvalue of multiplicity k and we obtain pairs of real solutions by taking real and imaginary parts of the formulae presented above.

Fundamental solutions and nonhomogeneous problems

Let us suppose that we have n linearly independent solutions $\mathbf{y^1}(t), \ldots, \mathbf{y^n}(t)$ of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$, where $\mathcal{A}$ is an $n \times n$ matrix, like the ones constructed in the previous paragraphs. Let us denote by $\mathcal{Y}(t)$ the matrix

$$\mathcal{Y}(t) = \begin{pmatrix} y_1^1(t) & \ldots & y_1^n(t) \\ \vdots & & \vdots \\ y_n^1(t) & \ldots & y_n^n(t) \end{pmatrix},$$

that is, the columns of $\mathcal{Y}(t)$ are the vectors $\mathbf{y^i}$, $i = 1, \ldots, n$. Any such matrix is called a *fundamental matrix* of the system $\mathbf{y}' = \mathcal{A}\mathbf{y}$.

We know that for a given initial vector $\mathbf{y^0}$ the solution is given by

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{y^0}$$

on one hand, and, by Theorem 8.1, by

$$\mathbf{y}(t) = C_1\mathbf{y^1}(t) + \ldots + C_n\mathbf{y^n}(t) = \mathcal{Y}(t)\mathbf{C},$$

on the other, where $\mathbf{C} = (C_1, \ldots, C_n)$ is a vector of constants to be determined. By putting $t = 0$ above we obtain the equation for $\mathbf{C}$

$$\mathbf{y^0} = \mathcal{Y}(0)\mathbf{C}$$

Since $\mathcal{Y}$ has independent vectors as its columns, it is invertible, so that

$$\mathbf{C} = \mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Thus, the solution of the initial value problem

$$\mathbf{y}' = \mathcal{A}\mathbf{y}, \qquad \mathbf{y}(0) = \mathbf{y^0}$$

is given by

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathcal{Y}^{-1}(0)\mathbf{y^0}.$$

Since $e^{t\mathbf{A}}\mathbf{y^0}$ is also a solution, by the uniqueness theorem we obtain explicit representation of the exponential function of a matrix

$$e^{t\mathcal{A}} = \mathcal{Y}(t)\mathcal{Y}^{-1}(0). \tag{2.8.57}$$

Let us turn our attention to the non-homogeneous system of equations

$$\mathbf{y}' = \mathcal{A}\mathbf{y} + \mathbf{g}(t). \tag{2.8.58}$$

The general solution to the homogeneous equation ($\mathbf{g}(t) \equiv 0$) is given by

$$\mathbf{y_h}(t) = \mathcal{Y}(t)\mathbf{C},$$

where $\mathcal{Y}(t)$ is a fundamental matrix and $\mathbf{C}$ is an arbitrary vector. Using the technique of variation of parameters, we will be looking for the solution in the form

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{u}(t) = u_1(t)\mathbf{y^1}(t) + \ldots + u_n(t)\mathbf{y^n}(t) \tag{2.8.59}$$

where $\mathbf{u}(t) = (u_1(t), \ldots, u_n(t))$ is a vector-function to be determined so that (2.8.59) satisfies (2.8.58). Thus, substituting (2.8.59) into (2.8.58), we obtain

$$\mathcal{Y}'(t)\mathbf{u}(t) + \mathcal{Y}(t)\mathbf{u}'(t) = \mathcal{A}\mathcal{Y}(t)\mathbf{u}(t) + \mathbf{g}(t).$$

Since $\mathcal{Y}(t)$ is a fundamental matrix, $\mathcal{Y}'(t) = \mathcal{A}\mathcal{Y}(t)$ and we find

$$\mathcal{Y}(t)\mathbf{u}'(t) = \mathbf{g}(t).$$

As we observed earlier, $\mathcal{Y}(t)$ is invertible, hence

$$\mathbf{u}'(t) = \mathcal{Y}^{-1}(t)\mathbf{g}(t)$$

and

$$\mathbf{u}(t) = \int^t \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds + \mathbf{C}.$$

Finally, we obtain

$$\mathbf{y}(t) = \mathcal{Y}(t)\mathbf{C} + \mathcal{Y}(t)\int^{t} \mathcal{Y}^{-1}(s)\mathbf{g}(s)ds \tag{2.8.60}$$

This equation becomes much simpler if we take $e^{t\mathcal{A}}$ as a fundamental matrix because in such a case $\mathcal{Y}^{-1}(t) = \left(e^{t\mathcal{A}}\right)^{-1} = e^{-t\mathcal{A}}$, that is, to calculate the inverse of $e^{t\mathcal{A}}$ it is enough to replace t by $-t$. The solution (2.8.60) takes then the form

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{C} + \int e^{(t-s)\mathcal{A}}\mathbf{g}(s)ds. \tag{2.8.61}$$

Chapter 3

Stability of systems of autonomous ordinary differential equations

1 Introduction

Unfortunately, in general there are no known methods of solving (2.4.26), (2.4.27). in general form. Though it is, of course, disappointing, it turns out that knowing exact solution is not really necessary. For example, let $y_1(t)$ and $y_2(t)$ denote the populations, at time t, of two species competing amongst themselves for the limited food and living space in some region. Further, suppose that the rates of growth of $y_1(t)$ and $y_2(t)$ are governed by (2.4.26). In such a case, for most purposes it is irrelevant to know the population sizes at each time t but rather it is important to know some qualitative properties of them. Specifically, the most important questions biologists ask are:

1. Do there exist values ξ_1 and ξ_2 at which the two species coexist in a steady state? That is to say, are there numbers ξ_1 and ξ_2 such that $\mathbf{y}_1(t) \equiv \xi_1$ and $\mathbf{y}_2(t) \equiv \xi_2$ is a solution to (2.4.26)? Such values, if they exist, are called *equilibrium points* of (2.4.26).
2. Suppose that the two species coexist in an equilibrium and suddenly a few members of one or both species are introduced to the environment. Will $\mathbf{y}_1(t)$ and $\mathbf{y}_2(t)$ remain close to their equilibrium values for all future times? Or may be these extra few members will give one of the species a large advantage so that it will proceed to annihilate the other species?
3. Suppose that $\mathbf{y}_1$ and $\mathbf{y}_2$ have arbitrary values at $t = 0$. What happens for large times? Will one species ultimately emerge victorious, or will the struggle for existence end in a draw?

Mathematically speaking, we are interested in determining the following properties of system (2.4.26).

Existence of equilibrium points and stationary solutions. Do there exist constant vectors $\mathbf{y^0} = (y_1^0, y_2^0)$ for which $\mathbf{y}(t) \equiv \mathbf{y^0}$ is a solution of (2.4.26)?

Stability. Let $\mathbf{x}(t)$ and $\mathbf{y}(t)$ be two solutions of (2.4.26) with initial values $\mathbf{x}(0)$ and $\mathbf{y}(0)$ very close to each other. Will $\mathbf{x}(t)$ and $\mathbf{y}(t)$ remain close for all future times or will $\mathbf{y}(t)$ eventually diverge from $\mathbf{x}(t)$?

Long time behaviour. What happens to solutions $\mathbf{y}(t)$ as t approaches infinity. Do all solutions approach equilibrium values? If they do not approach equilibrium, do they at least exhibit some regular behaviour, like e.g. periodicity, for large times.

First question is relatively easy to answer, and has been discussed in Section 4. The other two require more sophisticated tools which we will introduce now. We start with notions of stability.

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{3.1.1}$$

Definition 1.1 *The solution* $\mathbf{y}(t)$ *of (3.1.1) is (Lyapunov) stable if every other solution* $\mathbf{x}(t)$ *that starts sufficiently close to* $\mathbf{y}(t)$ *will remain close to it for all times. Precisely,* $\mathbf{y}(t)$ *is stable if for any* ϵ *there is* δ *such that for any solution* $\mathbf{x}$ *of (3.1.1) from*

$$\|\mathbf{x}(t_0) - \mathbf{y}(t_0)\| \leq \delta,$$

it follows

$$\|\mathbf{x}(t) - \mathbf{y}(t)\| \leq \epsilon.$$

We say that $\mathbf{y}(t)$ *is quasi-asymptotically stable ('tends to eventually') if there is* δ *such that if* $\|\mathbf{x}(t_0) - \mathbf{y}(t_0)\| \leq \delta$*, then*

$$\lim_{t\to\infty} \|\mathbf{y}(t) - \mathbf{x}(t)\| = 0.$$

Moreover, we say that $\mathbf{y}(t)$ *is asymptotically stable, if it is Lyapunov stable and quasi-asymptotically stable.*

The main interest in applications is to determine the stability of stationary solutions.

Example 1.1 Linear systems For linear systems the question of stability of solutions can be fully resolved. Firstly, we observe that any solution $\mathbf{y}(t)$ of the linear system

$$\mathbf{y}' = \mathcal{A}\mathbf{y} \tag{3.1.2}$$

is stable if and only if the stationary solution $\mathbf{x}(t) = (0, 0)$ is stable. To show this, let $\mathbf{z}(t)$ be any other solution; then $\mathbf{v}(t) = \mathbf{y}(t) - \mathbf{z}(t)$ is again a solution of (3.1.2). Therefore, if the null solution is stable, then $\mathbf{v}$ remains close to zero for all t if $\mathbf{v}(t_0)$ is small. This, however, implies that $\mathbf{y}(t)$ will remain close to $\mathbf{z}(t)$ if $\mathbf{y}(t_0)$ is sufficiently close to $\mathbf{z}(t_0)$. A similar argument applies to the asymptotic stability. Conversely, let the null solution be unstable; then there is a solution $\mathbf{h}(t)$ such that $\mathbf{h}(t_0)$ is small, but $h(t)$ becomes very large as t approaches to infinity. For any solution $\mathbf{y}(t)$, the function $\mathbf{z}(t) = \mathbf{y}(t) + \mathbf{h}(t)$ is again a solution to (3.1.2) which sways away from $\mathbf{y}(t)$ for large t. Thus, any solution is unstable.

An asymptotically stable stationary point is called a *sink* or *attracting*. On the contrary, if there is a neighbourhood of a stationary point such that any trajectory staring from this neighbourhood diverge from it, then such a stationary point is called *repelling* or a *source*.

Definition 1.2 *Suppose that* $\mathbf{x}$ *is an asymptotically stable stationary point of the equation (2.4.26). The set*

$$D_{\mathbf{x}} := \{\mathbf{y};\ \lim_{t\to\infty} \phi(t, \mathbf{y}) = \mathbf{x}\} \tag{3.1.3}$$

is called the domain of asymptotic stability (or the basin of attraction) of $\mathbf{x}$*. If* $D_{\mathbf{x}} = \mathbb{R}^n$*, then* $\mathbf{x}$ *is called globally asymptotically stable.*

Another important concept is that of an invariant set of the flow.

Definition 1.3 *A set* $M \subset \mathbb{R}^n$ *is called invariant with respect to the flow* $(t, \mathbf{x}) \to \phi(t, \mathbf{x})$ *generated by the system (2.4.26) if and only if* $\phi(t, \mathbf{x}) \in M$ *for any* $\mathbf{x} \in M$*. In other words,* M *is invariant if* $\Gamma_{\mathbf{x}} \subset M$ *for any* $\mathbf{x} \in M$*.*

A set M *is forward (res. backward) invariant if for any* $\mathbf{x} \in M$ *we have* $\Gamma^+_{\mathbf{x}} \subset M$ *(resp.* $\Gamma^-_{\mathbf{x}} \subset M$*).*

Example 1.2 Consider the system

$$\begin{aligned} x_1' &= x_1 - x_2 - x_1(x_1^2 + x_2^2) + \frac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}}, \\ x_2' &= x_1 + x_2 - x_2(x_1^2 + x_2^2) - \frac{x_1^2}{\sqrt{x_1^2 + x_2^2}}. \end{aligned} \tag{3.1.4}$$

Converting this equation into polar form, we obtain

$$\begin{aligned} r' &= r(1-r^2), \\ \theta' &= 2\sin^2\left(\frac{1}{2}\theta\right). \end{aligned} \tag{3.1.5}$$

We observe that $r' = 0$ when $r = 0$ or $r = 1$ and $\theta' = 0$ when $\theta = 0$. Also, $\theta' > 0$ for all $\theta \neq 0$. Thus, the half-line $\theta = 0$ is an invariant line and all trajectories move around to approach this half-line from below. On the other hand, in the r direction the trajectories tend to $r = 1$ (unless initially $r = 0$.) Hence, the system has two stationary points: the origin $r = 0$ and the point $(1, 0)$. It has also two invariant curves: half-line $\theta = 0$ and the circle $r = 1$. All other trajectories tend to $(1, 0)$. Consider any small neighbourhood of $(1, 0)$. Points with $\theta \leq 0$ will immediately tend to $(1, 0)$ without leaving this neighbourhood. On the other hand, points with $\theta > 0$ will make a circuit along the invariant curve $r = 1$ before tending to $(1, 0)$ from below. Hence, $(1, 0)$ is quasi-asymptotically stable but not Lyapunov stable.

2 One dimensional dynamics

In this section we shall focus on scalar autonomous differential equations:

$$y' = f(y) \tag{3.2.6}$$

As before, we assume that f is an everywhere defined function satisfying assumptions of Picard's theorem on the whole real line. Recall that $\phi(t, y_0)$ is the flow generated by this equation.

We note that it follows from Theorem 6.1 (i) that if y_0 is not an equilibrium point of (3.2.6), then $\phi(t, y_0)$ is never equal to an equilibrium point. In other words,

$$f(y_0) \neq 0 \Rightarrow f(\phi(t, y_0)). \tag{3.2.7}$$

From the above observation it follows that if f has several equilibrium points, then the stationary solutions corresponding to these points divide the (t, y) plane into strips such that any solution remains always confined to one of them. If we look at this from the point of phase space and orbits, we note that We note that the phase space in the 1 dimensional case is the real line $\mathbb{R}$, divided by equilibrium points and thus and orbits are simply segments (possibly infinite) between equilibrium points.

We shall formulate and prove a theorem that strengthens this observation.

Theorem 2.1 *All non-stationary solutions of the scalar autonomous equation (3.2.6) are either strictly decreasing or strictly increasing functions of t. For any $y_0 \in \mathbb{R}$, the solution $\phi(t, y_0)$ either diverges to $+\infty$ or $-\infty$, or converges to an equilibrium point, as $t \to \infty$.*

Proof. Assume that for some t_* the solution $y(t)$ has a local maximum or minimum $y_* = y(t_*)$. Since $y(t)$ is differentiable, we must have $\frac{dy}{dt}(t_*) = 0$ but then $f(y_*) = 0$ which makes y_* the equilibrium point of f so that $y_1(t) = y_*$ is a constant stationary solution. Then, however, $y(t_*) = y_1(t_*) = y_*$ so that by, (3.2.7), $y(t) = y_*$ for all t. Thus, if $y(t)$ is not a stationary solution, then it cannot attain local maxima or minima and thus must be either strictly increasing or strictly decreasing.

Since the solution is monotonic it either diverges to $\pm\infty$ (depending on whether it decreases or increases) or converges to a finite limit. Assume that the latter is the case and that y is an increasing function, thus

$$\lim_{t\to\infty} \phi(t, y_0) = \bar{y}.$$

Then it follows from Proposition 6.1 that $\bar{y}$ is an equilibrium point. We will provide, however, another proof of this fact, which is instructive in its own way.

Assume that $\bar{y}$ is not an equilibrium point. From continuity of (Darboux property), values of $y(t)$ must fill the interval $[y_0, \bar{y}]$ and this interval cannot contain any equilibrium point as the existence of such would

violate Picard's theorem. Thus, for any $y' \leq \bar{y}$, $f(y')$ is strictly positive and integrating the equation we obtain

$$t(y') - t(y_0) = \int_{y_0}^{y'} \frac{dy}{f(y)}.$$

Passing with t to infinity, we see that the left-hand side becomes infinite and so

$$\int_{y_0}^{\bar{y}} \frac{dy}{f(y)} = \infty.$$

By assumption, the interval of integration is finite so that the only way the integral could be infinite is when $1/f(\bar{y}) = \infty$ or $f(\bar{y}) = 0$. Thus $\bar{y}$ is an equilibrium point. ■

Remark 2.1 In the formulation and the proof of Theorem 2.1 we tacitly assumed that the solution exists for all $t \in \mathbb{R}$. Clearly, this is not true in general. However, we observe that any solution starting between equilibria is bounded and thus exists for all times. If the initial point is to the left of an equilibrium and the solution is increasing, or the initial point is to the right of an equilibrium and the solution is decreasing, then the solution is defined as $t \to +\infty$, and correspondingly, if the initial point is to the left of an equilibrium and the solution is decreasing, or the initial point is to the right of an equilibrium and the solution is increasing, then the solution is defined as $t \to -\infty$. Consider now the case when we have a decreasing solution $y(t)$ starting from an initial point which is left from the least equilibrium. Then either the solution exists for all $t \to +\infty$ or the maximal interval of existence $[t_0, t_{\max})$ is finite in which case $y(t) \to -\infty$ as $t \to t_{\max}$. Hence, Theorem 2.1 remains valid if we replace $t \to \infty$ by $t \to t_{\max}$ (with understanding that $t_{\max}$ may be ∞.) The remaining three cases are dealt with in the same way.

Let us summarize the possible scenarios for an autonomous equation (3.2.6). Assume that y_* is a single equilibrium point of f with $f(y) < 0$ for $y < y_*$ and $f(y) > 0$ for $y > y_*$. If the initial condition satisfies $y_0 < y_*$, then $\phi(t, y_0)$ decreases so it diverges either to $-\infty$ or to an equilibrium point. Since there is no equilibrium point smaller than y_0, the solution must diverge to $-\infty$. Similarly, for $y_0 > y_*$ we see that $y(t, y_0)$ must diverge to infinity. Hence, y_* is the source (or repellent).

Conversely, assuming that y_* is a single stationary point of f with $f(y) > 0$ for $y < y_*$ and $f(y) < 0$ for $y > y_*$, we see that if $y_0 < y_*$, then $\phi(t, y_0)$ increases so, being bounded, it must converge to y_*. Similarly, for $y_0 > y_*$, we see that $\phi(t, y_0)$ must decrease converging again to y_*. Hence y_* is attracting. If there are more equilibrium points, then the behaviour of the solution is a combination of the above scenarios. Assume, for example, that f has two equilibrium points $y_1 < y_2$ and it is positive for $y < y_1$, negative for $y_1 < y < y_2$ and again positive for $y > y_2$. Thus, for $y_0 < y_1$, $\phi(t, y_0)$ increases converging to y_1, for $y_1 < y_0 < y_2$ we have $\phi(t, y_0)$ decreasing and converging to y_1 and, finally, for $y_0 > y_2$, $\phi(t, y_0)$ increases to infinity.

Example 2.1 Let us consider the Cauchy problem for the logistic equation

$$y' = y(1 - y), \qquad y(0) = y_0. \tag{3.2.8}$$

Let us get as many information as possible about the solutions to this problem without actually solving it. Firstly, we observe that the right-hand side is given by $f(y) = y(1-y)$ which is a polynomial and therefore at each point of $\mathbb{R}^2$ the assumptions of Picard's theorem are satisfied, that is, through each point $(0, y_0)$ there passes only one solution of (3.2.8). However, f is not a globally Lipschitz function so that this solutions may be defined only locally, on small time intervals.

The second step is to determine equilibrium points and stationary solutions. From

$$y(1 - y) = 0.$$

we see that $y \equiv 0$ and $y \equiv 1$ are the only equilibrium solutions. Moreover, $f(y) < 0$ for $y < 0$ and $y > 0$ and $f(y) > 0$ for $0 < y < 1$. From Picard's theorem (uniqueness) it follows then that solutions staring from

$y_0 < 0$ will stay strictly negative, starting from $0 < y_0 < 1$ will stay in this interval and, finally those with $y_0 > 1$ will be larger than 1, for all times of their respective existence, as they cannot cross equilibrium solutions. Then, from Theorem 2.1, we see that the solutions with negative initial condition are decreasing and therefore tend to $-\infty$ for increasing times (in fact, they blow-up (become infinite) for finite times) as integrating the equation, we obtain

$$t(y) = \int_{y_0}^{y} \frac{d\eta}{\eta(1-\eta)}$$

and we see that passing with y to $-\infty$ on the right-hand side we obtain a finite number (the improper integral exists) giving the time of blow-up.

Next, solutions with $0 < y_0 < 1$ are bounded and thus defined for all times by Remark **??**. They are increasing and thus must converge to the larger equilibrium point, that is

$$\lim_{t\to\infty} \phi(t, y_0) = 1.$$

Finally, if we start with $y_0 > 1$, then the solution $\phi(t, y_0)$ will be decreasing and thus bounded, satisfying again

$$\lim_{t\to\infty} \phi(t, y_0) = 1.$$

We can learn even more about the shape of the solution curves. Differentiating the equation with respect to time and using the product rule, we obtain

$$y'' = y'(1-y) - yy' = y'(1-2y).$$

Since for each solution (apart from the stationary ones), y' has fixed sign, we see that the inflection points can exist only on solutions staring at $y_0 \in (0, 1)$ and occur precisely at $y = 1/2$ - for this value of y the solution changes from being convex downward to being convex upward. In the two other cases, the second derivative is of constant sign, giving the solution convex upward for negative solutions and convex downward for solutions larger than 1.

We can give a real life interpretation of these result if we realize that Cauchy problem 3.2.8 represents a model of a population size that starts with y_0 individuals and whose rate of growth is determined by the capacity of habitat (in this case normalized to 1). In such a model, negative initial values and solutions do not make any sense so that we shall not discuss them.

Also, the case with $y_0 > 1$, that is, with initial population exceeding the maximum capacity, is not particularly interesting - the population in this case decreases steadily to the maximum capacity value.

In the most natural case, when we start with a non-zero population below the maximum capacity, we have two cases to distinguish. If $y_0 < 1/2$ (that is we start below half of the maximum capacity), then the population grows very fast with increasing growth rate (second derivative of y is positive) till it reaches the size of $1/2$ whereupon the growth starts to slow down; the population is still growing but the rate of growth starts to decrease. The population is still growing but slower and slower, approaching the maximum capacity. Of course, if we start with $y_0 > 1/2$, then the population will not experience the fast, almost exponential, growth.

These considerations have an interesting application in determining sustainable harvesting of some population, e.g. sustainable fishing. Sustainable harvesting is a harvesting that does not destroy the harvested population while at the same time brings maximum profit to the community. To keep the harvested population at constant size, we can only catch the amount by which the population increases. Over a short time interval Δt the population size is changes approximately from $u(t)$ to $y(t) + y'(t)\Delta t$ hence, to keep the population in a constant size, we can catch only $y'(t)\Delta t$ individuals in time Δt. Thus, the yield will be the best when the population is kept at the size at which its growth is the highest. If the population grows according to the logistic law, then from the above considerations we see that the highest yield will be if the size of the population is kept at half of the capacity of the habitat, because then $y'' = (y')' = 0$. We can only harvest at the rate of increase which, in this case, will be

$$y'(1/2) = \frac{1}{4}.$$

3 Stability by Linearization

3.1 Planar linear systems

In this section we shall present a complete description of all orbits of the linear differential system

$$\mathbf{y}' = \mathcal{A}\mathbf{y} \tag{3.3.9}$$

where $\mathbf{y}(t) = (y_1(t), y_2(t))$ and

$$\mathcal{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We shall assume that $\mathcal{A}$ is invertible, that is, $ad - bc \neq 0$. In such a case $\mathbf{y} = (0, 0)$ is the only equilibrium point of (3.3.9).

The phase portrait is fully determined by the eigenvalues of the matrix $\mathcal{A}$. Let us briefly describe all possible cases, as determined by the theory of the preceding section. The general solution can be obtained as a linear combination of two linearly independent solutions. To find them, we have to find first the eigenvalues of $\mathcal{A}$, that is, solutions to

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = (a - \lambda)(d - \lambda) - bc = \lambda^2 - \lambda(d + a) + ad - bc.$$

Note that by the assumption on invertibility, $\lambda = 0$ is not an eigenvalue of $\mathcal{A}$. We have the following possibilities:

a) $\lambda_1 \neq \lambda_2$. In this case each eigenvalue must be simple and therefore we have two linearly independent eigenvectors $\mathbf{v}^1$, $\mathbf{v}^2$. The expansion $e^{t\mathcal{A}}\mathbf{v}^i$ for $i = 1, 2$ terminates after the first term. We distinguish two cases.

$\diamond$ If λ_1, λ_2 are real numbers, then the general solution is given simply by

$$\mathbf{y}(t) = c_1 e^{\lambda_1 t}\mathbf{v}^1 + c_2 e^{\lambda_2 t}\mathbf{v}^2. \tag{3.3.10}$$

$\diamond$ If λ_1, λ_2 are complex numbers, then the general solution is still given by the above formula but the functions above are complex and we would rather prefer solution to be real. To achieve this, we note that λ_1, λ_2 must be necessarily complex conjugate $\lambda_1 = \xi + i\omega, \lambda_2 = \xi - i\omega$, where ξ and ω are real. It can be also proved that the associated eigenvectors $\mathbf{v}^1$ and $\mathbf{v}^2$ are also complex conjugate. Let $\mathbf{v}^1 = \mathbf{u} + i\mathbf{v}$; then the real-valued general solution is given by

$$\mathbf{y}(t) = c_1 e^{\xi t}(\mathbf{u}\cos\omega t - \mathbf{v}\sin\omega t) + c_2 e^{\xi t}(\mathbf{u}\sin\omega t + \mathbf{v}\cos\omega t). \tag{3.3.11}$$

This solution can be written in a more compact form

$$\mathbf{y}(t) = e^{\xi t}\left(A_1\cos(\omega t - \phi_1), A_2\cos(\omega t - \phi_2)\right), \tag{3.3.12}$$

for some choice of constants $A_1, A_2 > 0$ and ϕ_1, ϕ_2.

b) $\lambda_1 = \lambda_2 = \lambda$. There are two cases to distinguish.

$\diamond$ There are two linearly independent eigenvectors $\mathbf{v}^1$ and $\mathbf{v}^2$ corresponding to λ. In this case the general solution is given by

$$\mathbf{y}(t) = e^{\lambda t}(c_1\mathbf{v}^1 + c_2\mathbf{v}^2). \tag{3.3.13}$$

$\diamond$ If there is only one eigenvector, then following the discussion above, we must find a vector $\mathbf{v}^2$ satisfying $(\lambda I - \mathcal{A})\mathbf{v}^2 \neq 0$ and $(\lambda I - \mathcal{A})^2\mathbf{v}^2 = 0$. However, since we are in the two-dimensional space, the latter is satisfied by any vector $\mathbf{v}^2$ and, since the eigenspace is one dimensional, from

$$(\lambda I - \mathcal{A})^2\mathbf{v}^2 = (\lambda I - \mathcal{A})(\lambda I - \mathcal{A})\mathbf{v}^2 = 0$$

it follows that $(\lambda I - \mathcal{A})\mathbf{v}^2 = k\mathbf{v}^1$. Thus, the formula for $e^{\mathcal{A}t}\mathbf{v}^2$ simplifies as

$$e^{t\mathcal{A}}\mathbf{v}^2 = e^{\lambda t}\left(\mathbf{v}^2 + t(\lambda I - \mathcal{A})\mathbf{v}^2\right) = e^{\lambda t}\left(\mathbf{v}^2 + kt\mathbf{v}^1\right).$$

Thus, the general solution in this case can be written as

$$\mathbf{y}(t) = e^{\lambda t}\left((c_1 + c_2 kt)\mathbf{v}^1 + c_2\mathbf{v}^2\right). \tag{3.3.14}$$

Remark 3.1 Before we embark on describing phase portraits, let us observe that if we change the direction of time in (3.3.9): $\tau = -t$ and $\mathbf{z}(\tau) = \mathbf{y}(-\tau) = \mathbf{y}(t)$, then we obtain

$$\mathbf{z}'_\tau = -\mathcal{A}\mathbf{z}$$

and the eigenvalues of $-\mathcal{A}$ are precisely the negatives of the eigenvalues of $\mathcal{A}$. Thus, the orbits of solutions corresponding to systems governed by $\mathcal{A}$ and $-\mathcal{A}$ or, equivalently, with eigenvalues that differ only by sign, are the same with only difference being the direction in which they are traversed.

We are now in a position to describe all possible phase portraits of (3.3.9). Again we have to go through several cases.

i) $\lambda_2 < \lambda_1 < 0$. Let $\mathbf{v}^1$ and $\mathbf{v}^2$ be eigenvectors of $\mathcal{A}$ with eigenvalues λ_1 and λ_2, respectively. In the $y_1 - y_2$ plane we draw four half-lines l_1, l_1', l_2, l_2' parallel to $\mathbf{v}^1$, $-\mathbf{v}^1$, $\mathbf{v}^2$ and $-\mathbf{v}^2$, respectively, and emanating from the origin, as shown in Fig 2.1. Observe first that $\mathbf{y}(t) = ce^{\lambda_i t}\mathbf{v}^i$, $i = 1, 2$, are the solutions to (3.3.9) for any choice of a non-zero constant c and, as they are parallel to $\mathbf{v}^i$, the orbits are the half-lines l_1, l_1', l_2, l_2 (depending on the sign of the constant c) and all these orbits are traced towards the origin as $t \to \infty$. Since every solution $\mathbf{y}(t)$ of (3.3.9) can be written as

$$\mathbf{y}(t) = c_1 e^{\lambda_1 t}\mathbf{v}^1 + c_2 e^{\lambda_2 t}\mathbf{v}^2$$

for some choice of constants c_1 and c_2. Since $\lambda_1, \lambda_2 < 0$, every solution tends to $(0, 0)$ as $t \to \infty$, and so every orbit approaches the origin for $t \to \infty$. We can prove an even stronger fact – as $\lambda_2 < \lambda_1$, the second term becomes negligible for large t and therefore the tangent of the orbit of $\mathbf{y}(t)$ approaches the direction of l_1 if $c_1 > 0$ and of l_1' if $c_1 < 0$. Thus, every orbit except that with $c_1 = 0$ approaches the origin along the same fixed line. Such a type of an equilibrium point is called a *stable node*. If we have $0 < \lambda_1 < \lambda_2$, then by Remark 3.1, the orbits of (3.3.9) will have the same shape as in case i) but the arrows will be reversed so that the origin will repel all the orbits and the orbits will be unbounded as $t \to \infty$. Such an equilibrium point is called an *unstable node*.

ii) $\lambda_1 = \lambda_2 = \lambda < 0$. In this case the phase portrait of (3.3.9) depends on whether $\mathcal{A}$ has one or two linearly independent eigenvectors. In the latter case, the general solution in given (see b) above) by

$$\mathbf{y}(t) = e^{\lambda t}(c_1\mathbf{v}^1 + c_2\mathbf{v}^2),$$

so that orbits are half-lines parallel to $c_1\mathbf{v}^1 + c_2\mathbf{v}^2$. These half-lines cover every direction of the $y_1 - y_2$ plane and, since $\lambda < 0$, each solution will converge to $(0, 0)$ along the respective line. Thus, the phase portrait looks like in Fig. 2.2a. If there is only one independent eigenvector corresponding to λ, then by (3.3.14)

$$\mathbf{y}(t) = e^{\lambda t}\left((c_1 + c_2 kt)\mathbf{v}^1 + c_2\mathbf{v}^2\right)$$

for some choice of constants c_1, c_2, k. Obviously, every solution approaches $(0, 0)$ as $t \to \infty$. Putting $c_2 = 0$, we obtain two half-line orbits $c_1 e^{\lambda t}\mathbf{v}^1$ but, contrary to the case i), there are no other half-line orbits. In addition, the term $c_1\mathbf{v}^1 + c_2\mathbf{v}^2$ becomes small in comparison with $c_2 kt\mathbf{v}^1$ as $t \to \infty$ so that the orbits approach the origin in the direction of $\pm\mathbf{v}^1$. The phase portrait is presented in Fig. 2.2b. The equilibrium in both cases is called the *stable degenerate node*. If $\lambda_1 = \lambda_2 > 0$, then again by Remark 3.1, the picture in this case will be the same as in Fig. 2.a-b but with the direction of arrows reversed. Such equilibrium point is called an *unstable degenerate node*.

Fig. 2.1 Stable node

2.2 Stable degenerate node

A saddle point

iii) $\lambda_1 < 0 < \lambda_2$. As in case i), in the $y_1 - y_2$ plane we draw four half-lines l_1, l_1', l_2, l_2' that emanate from the origin and are parallel to $\mathbf{v}^1$, $-\mathbf{v}^1$, $\mathbf{v}^2$ and $-\mathbf{v}^2$, respectively, as shown in Fig 2.3. Any solution is given by

$$\mathbf{y}(t) = c_1 e^{\lambda_1 t}\mathbf{v}^1 + c_2 e^{\lambda_2 t}\mathbf{v}^2$$

for some choice of c_1 and c_2. Again, the half-lines are the orbits of the solutions: l_1, l_1' for $c_1 e^{\lambda_1 t}\mathbf{v}^1$ with $c_1 > 0$ and $c_1 < 0$, and l_2, l_2' for $c_2 e^{\lambda_2 t}\mathbf{v}^2$ with $c_2 > 0$ and $c_2 < 0$, respectively. However, the direction of arrows is different on each pair of half-lines: while the solution $c_1 e^{\lambda_1 t}\mathbf{v}^1$ converges towards $(0,0)$ along l_1 or l_1' as $t \to \infty$, the solution $c_2 e^{\lambda_2 t}\mathbf{v}^2$ becomes unbounded moving along l_2 or l_2', as $t \to \infty$. Next, we observe that if $c_1 \neq 0$, then for large t the second term $c_2 e^{\lambda_2 t}\mathbf{v}^2$ becomes negligible and so the solution becomes unbounded as $t \to \infty$ with asymptotes given by the half-lines l_2, l_2', respectively. Similarly, for $t \to -\infty$ the term $c_1 e^{\lambda_1 t}\mathbf{v}^1$ becomes negligible and the solution again escapes to infinity, but this time with asymptotes l_1, l_1', respectively. Thus, the phase portrait, given in Fig. 2.3, resembles a saddle near $y_1 = y_2 = 0$ and, not surprisingly, such an equilibrium point is called a *saddle*. The case $\lambda_2 < 0 < \lambda_1$ is of course symmetric.

iv) $\lambda_1 = \xi + i\omega, \lambda_2 = \xi - i\omega$. In (3.3.12) we derived the solution in the form

$$\mathbf{y}(t) = e^{\xi t}\left(A_1\cos(\omega t - \phi_1), A_2\cos(\omega t - \phi_2)\right).$$

We have to distinguish three cases:

α) If $\xi = 0$, then

$$y_1(t) = A_1\cos(\omega t - \phi_1), \qquad y_2(t) = A_2\cos(\omega t - \phi_2),$$

both are periodic functions with period $2\pi/\omega$ and y_1 varies between $-A_1$ and A_1 while y_2 varies between $-A_2$ and A_2. Consequently, the orbit of any solution $\mathbf{y}(t)$ is a closed curve containing the origin inside and the phase portrait has the form presented in Fig. 3.4a. For this reason we say that the equilibrium point of (3.3.9) is a *center* when the eigenvalues of $\mathcal{A}$ are purely imaginary. The direction of arrows must be determined from the equation. The simplest way of doing this is to check the sign of y_2' when $y_2 = 0$. If at $y_2 = 0$ and $y_1 > 0$ we have $y_2' > 0$, then all the orbits are traversed in the anticlockwise direction, and conversely.

Fig.4 Center, stable and unstable foci

β) If $\xi < 0$, then the factor $e^{\xi t}$ forces the solution to come closer to zero at every turn so that the solution spirals into the origin giving the picture presented in Fig. 2.4b. The orientation of the spiral must be again determined directly from the equation. Such an equilibrium point is called a *stable focus*.

γ) If $\xi > 0$, then the factor $e^{\xi t}$ forces the solution to spiral outwards creating the picture shown in Fig. 4c. Such an equilibrium point is called an *unstable focus*.

4 Stability of equilibrium solutions

4.1 Linear systems

The discussion of phase-portraits for two-dimensional linear, given in the previous section allows to determine easily under which conditions $(0, 0)$ is stable. Clearly, the only stable cases are when real parts of both eigenvalues are nonnegative with asymptotic stability offered by eigenvalues with strictly negative ones (the case of the centre is an example of a stable but not asymptotically stable equilibrium point).

Analogous results can be formulated for linear systems in higher dimensions. By considering formulae for solutions we ascertain that the equilibrium point is (asymptotically stable) if all the eigenvalues have negative real parts and is unstable if at least one eigenvalue has positive real part. The case of eigenvalues with zero real part is more complicated as in higher dimension we can have multiple complex eigenvalues. Here, again from the formula for solutions, we can see that if for each eigenvalue with zero real part of algebraic multiplicity k there is k linearly independent eigenvectors, the solution is stable. However, if geometric and algebraic multiplicities of at least such eigenvalue are different, then in the solution corresponding to this eigenvalue there will appear a polynomial in t which will cause the solution to be unstable.

4.2 Nonlinear systems-stability by linearization

The above considerations can be used to determine stability of equilibrium points of arbitrary differential equations

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{3.4.15}$$

Let us first note the following result.

Lemma 4.1 *If* $\mathbf{f}$ *has continuous partial derivatives of the first order in some neighbourhood of* $\mathbf{y}^0$*, then*

$$\mathbf{f}(\mathbf{x}+\mathbf{y}^0) = \mathbf{f}(\mathbf{y}^0) + \mathcal{A}\mathbf{x} + \mathbf{g}(\mathbf{x}) \tag{3.4.16}$$

where

$$\mathcal{A} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{y^0}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{y^0}) \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n}(\mathbf{y^0}) & \dots & \frac{\partial f_n}{\partial x_n}(\mathbf{y^0}) \end{pmatrix},$$

and $\mathbf{g}(\mathbf{x})/\|\mathbf{x}\|$ *is continuous in some neighbourhood of* $\mathbf{y}^0$ *and vanishes at* $\mathbf{x} = \mathbf{y}^0$*.*

Proof. The matrix $\mathcal{A}$ has constant entries so that $\mathbf{g}$ defined by

$$\mathbf{g}(\mathbf{x}) = \mathbf{f}(\mathbf{x}+\mathbf{y}^0) - \mathbf{f}(\mathbf{y}^0) - \mathcal{A}\mathbf{x}$$

is a continuous function of $\mathbf{x}$. Hence, $\mathbf{g}(\mathbf{x})/\|\mathbf{x}\|$ is also continuous for $\mathbf{x} \neq \mathbf{0}$. Using now Taylor's formula for each component of $\mathbf{f}$ we obtain

$$f_i(\mathbf{x}+\mathbf{y}^0) = f_i(\mathbf{y}^0) + \frac{\partial f_i}{\partial x_1}(\mathbf{y}^0)x_1 + \ldots + \frac{\partial f_i}{\partial x_n}x_n(\mathbf{y}^0) + R_i(\mathbf{x}), \quad i = 1, \ldots, n,$$

where, for each i, the remainder R_i satisfies

$$|R_i(x)| \leq M(\|\mathbf{x}\|)\|\mathbf{x}\|$$

and M tends to zero is $\|\mathbf{x}\| \to 0$. Thus,

$$\mathbf{g}(\mathbf{x}) = (R_1(\mathbf{x}), \ldots, R_n(\mathbf{x}))$$

and

$$\frac{\|\mathbf{g}(\mathbf{x})\|}{\|\mathbf{x}\|} \leq M(\|\mathbf{x}\|) \to 0$$

as $\|\mathbf{x}\| \to 0$ and, $\mathbf{f}(\mathbf{y}^0) = 0$, the lemma is proved. ■

The linear system

$$\mathbf{x}' = \mathcal{A}\mathbf{x}$$

is called the linearization of (3.5.31) around the equilibrium point $\mathbf{y}^0$.

Theorem 4.1 *Suppose that* $\mathbf{f}$ *is a differentiable function in some neighbourhood of the equilibrium point* $\mathbf{y}^0$*. Then,*

1. *The equilibrium point* $\mathbf{y}^0$ *is asymptotically stable if all the eigenvalues of the matrix* $\mathcal{A}$ *have negative real parts, that is, if the equilibrium solution* $\mathbf{x}(t) = \mathbf{0}$ *of the linearized system is asymptotically stable.*
2. *The equilibrium point* $\mathbf{y}^0$ *is unstable if at least one eigenvalue has a positive real part.*
3. *If all the eigenvalues of* $\mathcal{A}$ *have non-negative real part but at least one of them has real part equal to* 0*, then the stability of the equilibrium point* $\mathbf{y}^0$ *of the nonlinear system (3.5.31) cannot be determined from the stability of its linearization.*

Proof. To prove 1) we use the variation of constants formula (2.8.61) applied to (3.5.31) written in the form of Lemma 4.1 for $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{y}^0$:

$$\mathbf{x}' = \mathbf{y}' = \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x} + \mathbf{y}^0) = \mathcal{A}\mathbf{x} + \mathbf{g}(\mathbf{x}).$$

Thus

$$\mathbf{x}(t) = e^{t\mathcal{A}}\mathbf{x}(0) + \int_0^t e^{(t-s)\mathcal{A}}\mathbf{g}(\mathbf{x}(s))ds.$$

Denoting by α' the maximum of real parts of eigenvalues of $\mathcal{A}$ we observe that for any $\alpha > \alpha'$

$$\|e^{t\mathcal{A}}\mathbf{x}(0)\| \leq Ke^{-\alpha t}\|\mathbf{x}(0)\|, \quad t \geq 0,$$

for some constant $K \geq 1$. Note that in general we have to take $\alpha > \alpha'$ to account for possible polynomial entries in $e^{t\mathcal{A}}$. Thus, since $\alpha' < 0$, then we can take also $\alpha < 0$ keeping the above estimate satisfied. From the assumption on $\mathbf{g}$, for any ϵ we find $\delta > 0$ such that if $\|\mathbf{x}\| \leq \delta$, then

$$\|\mathbf{g}(\mathbf{x})\| \leq \epsilon\|\mathbf{x}\|. \tag{3.4.17}$$

Assuming for a moment that for $0 \leq s \leq t$ we can keep $\|\mathbf{x}(s)\| \leq \delta$, we can write

$$\begin{aligned}\|\mathbf{x}(t)\| &\leq \|e^{\mathcal{A}t}\mathbf{x}(0)\| + \int_0^t \|e^{\mathcal{A}(t-s)}\mathbf{g}(\mathbf{x}(s))\|ds \\ &\leq Ke^{-\alpha t}\mathbf{x}(0) + K\epsilon\int_0^t e^{-\alpha(t-s)}\|\mathbf{x}(s)\|ds\end{aligned}$$

or, multiplying both sides by $e^{\alpha t}$ and setting $z(t) = e^{\alpha t}\|\mathbf{x}(t)\|$,

$$z(t) \leq K\|\mathbf{x}(0)\| + K\epsilon\int_0^t z(s)ds. \tag{3.4.18}$$

Using Gronwall's lemma we obtain thus

$$\|\mathbf{x}(t)\| = e^{-\alpha t}z(t) \leq K\|\mathbf{x}(0)\|e^{(K\epsilon-\alpha)t},$$

providing $\|\mathbf{x}(s)\| \leq \delta$ for all $0 \leq s \leq t$. Let us take $\epsilon \leq \alpha/2K$, then the above can be written as

$$\|\mathbf{x}(t)\| \leq K\|\mathbf{x}(0)\|e^{-\frac{\alpha t}{2}}. \tag{3.4.19}$$

Assume now that $\|\mathbf{x}(0)\| < \delta/K \leq \delta$ where δ was fixed for $\epsilon \leq \alpha/2K$. Then $\|\mathbf{x}(0)\| < \delta$ and, by continuity, $\|\mathbf{x}(t)\| \leq \delta$ for some time. Let $\mathbf{x}(t)$ be defined on some interval I and $t_1 \in I$ be the first time for which $\|\mathbf{x}(t)\| = \delta$. Then for $t \leq t_1$ we have $\|\mathbf{x}(t)\| \leq \delta$ so that for all $t \leq t_1$ we can use (3.4.19) getting, in particular,

$$\|\mathbf{x}(t_1)\| \leq \delta e^{-\frac{\alpha t_1}{2}} < \delta,$$

that is a contradiction. Thus $\|\mathbf{x}(t)\| < \delta$ if $\|\mathbf{x}(0)\| < \delta_1$ in the whole interval of existence but then, if the interval was finite, then we could extend the solution to a larger interval as the solution is bounded at the endpoint and the same procedure would ensure that the solution remains bounded by δ on the larger interval. Thus, the extension can be carried out for all the values of $t \geq 0$ and the solution exists for all t and satisfies $\|\mathbf{x}(t)\| \leq \delta$ for all $t \geq 0$. Consequently, (3.4.19) holds for all t and the solution $\mathbf{x}(t)$ converges exponentially to 0 as $t \to \infty$ proving the asymptotic stability of the stationary solution $\mathbf{y}^0$.

Statement 2 follows either from Example 5.11 or Stable Manifold theorem, which will be commented upon later.

To prove 3, it is enough to display two systems with the same linear part and different behaviour of solutions. Let us consider

$$\begin{aligned} y_1' &= y_2 - y_1(y_1^2 + y_2^2) \\ y_2' &= -y_1 - y_2(y_1^2 + y_2^2) \end{aligned}$$

with the linearized system given by

$$\begin{aligned} y_1' &= y_2 \\ y_2' &= -y_1 \end{aligned}$$

The eigenvalues of the linearized system are $\pm i$. To analyze the behaviour of the solutions to the non-linear system, let us multiply the first equation by y_1 and the second by y_2 and add them together to get

$$\frac{1}{2}\frac{d}{dt}(y_1^2 + y_2^2) = -(y_1^2 + y_2^2)^2.$$

Solving this equation we obtain

$$y_1^2 + y_2^2 = \frac{c}{1 + 2ct}$$

where $c = y_1^2(0) + y_2^2(0)$. Thus $y_1^2(t) + y_2^2(t)$ approaches $\mathbf{0}$ as $t \to \infty$ and $y_1^2(t) + y_2^2(t) < y_1^2(0) + y_2^2(0)$ for any $t > 0$ and we can conclude that the equilibrium point $\mathbf{0}$ is asymptotically stable.

Consider now the system

$$\begin{aligned} y_1' &= y_2 + y_1(y_1^2 + y_2^2) \\ y_2' &= -y_1 + y_2(y_1^2 + y_2^2) \end{aligned}$$

with the same linear part and thus with the same eigenvalues. As above we obtain that

$$y_1^2 + y_2^2 = \frac{c}{1 - 2ct}$$

with the same meaning for c. Thus, any solution with non-zero initial condition blows up at the time $t = 1/2c$ and therefore the equilibrium solution $\mathbf{0}$ is unstable. ■

Example 4.1 Find all equilibrium solutions of the system of differential equations

$$\begin{aligned} y_1' &= 1 - y_1 y_2, \\ y_2' &= y_1 - y_2^3, \end{aligned}$$

and determine, if possible, their stability.

Solving equation for equilibrium points $1 - y_1 y_2 = 0, y_1 - y_2^3 =$ we find two equilibria: $y_1 = y_2 = 1$ and $y_1 = y_2 = -1$. To determine their stability we have to reduce each case to the equilibrium at $\mathbf{0}$. For the first case we put $u(t) = y_1(t) - 1$ and $v(t) = y_2 - 1$ so that

$$\begin{aligned} u' &= -u - v - uv, \\ v' &= u - 3v - 3v^2 - v^3, \end{aligned}$$

so that the linearized system has the form

$$\begin{aligned} u' &= -u - v, \\ v' &= u - 3v, \end{aligned}$$

and the perturbing term is given by $\mathbf{g}(u, v) = (-uv, -3v^2 + v^3)$ and, as the right-hand side of the original system is infinitely differentiable at $(0, 0)$ the assumptions of the stability theorem are satisfied. The eigenvalues of the linearized system are given by $\lambda_{1,2} = -2$ and therefore the equilibrium solution $\mathbf{y}(t) \equiv (1, 1)$ is asymptotically stable.

For the other case we set $u(t) = y_1(t) + 1$ and $v(t) = y_2 + 1$ so that

$$\begin{aligned} u' &= u + v - uv, \\ v' &= u - 3v + 3v^2 - v^3, \end{aligned}$$

so that the linearized system has the form

$$\begin{aligned} u' &= u + v, \\ v' &= u - 3v, \end{aligned}$$

and the perturbing term is given by $\mathbf{g}(u, v) = (-uv, 3v^2 - v^3)$. The eigenvalues of the linearized system are given by $\lambda_1 = -1 - \sqrt{5}$ and $\lambda_2 = -1 + \sqrt{5}$and therefore the equilibrium solution $\mathbf{y}(t) \equiv (-1, -1)$ is unstable.

5 Stability through the Lyapunov function

Consider again the system (2.4.26) in $\mathbb{R}^n$. Suppose that it has an isolated equilibrium at $\mathbf{x}_0$. Then, by writing (2.4.26) as

$$\mathbf{y}' = (\mathbf{y} - \mathbf{x}_0)' = \mathbf{x}' = \mathbf{f}(\mathbf{x} + \mathbf{x}_0) = \tilde{\mathbf{f}}(\mathbf{x}),$$

we obtain an equivalent system for which $\mathbf{x} = 0$ becomes an isolated equilibrium. Thus there is no loss of generality to consider (2.4.26) with $\mathbf{x} = 0$ as its equilibrium.

Let Ω be an open neighbourhood of 0 and let $V : \Omega \to \mathbf{R}$ be a continuously differentiable function. We define derivative of the derivative of V along trajectories of (2.4.26) by the chain rule

$$V' = \frac{dV}{dt} = \mathbf{x}' \cdot \nabla V = \mathbf{f} \cdot \nabla V = \sum_{i=1}^{n} f_i \frac{\partial V}{\partial x_i} \tag{3.5.20}$$

Example 5.1 Let us consider a system

$$\mathbf{y}' = \mathbf{f}(\mathbf{y})$$

with $\mathbf{f}$ being a potential field; that is, there is a scalar function V satisfying

$$\mathbf{f}(\mathbf{x}) = -\text{grad}V(x).$$

In general, not every vector field has a potential. An exception is offered by one dimensional fields when we have

$$V(x) = -\int_a^x f(z)dz$$

for some fixed a (the potential of a field is determined up to a constant). We note that since $dV/dx = -f$, the stationary points of V correspond to equilibria of f. Furthermore, if $t \to x(t)$ is any solution of the equation $x' = f(x)$ then we have

$$V'(x(t)) = \frac{dV(x(t))}{dt} = \frac{dV}{dx}(x(t))\frac{dx}{dt} = -f(x(t))f(x(t)) < 0,$$

so that $V(x(t))$ strictly decreases along trajectories. In other words, the point $x(t)$ moves always in the direction of decreasing V and thus equilibria corresponding to minima of V are asymptotically stable and corresponding to maxima are unstable.

In this section we shall discuss a generalization of the above concept.

Definition 5.1 *A continuously differentiable function V on $\Omega \ni 0$ is called a Lyapunov function for (2.4.26) if*

1. $V(0) = 0$ *and* $V(\mathbf{x}) > 0$ *on* Ω*;*

2. $V' \leq 0$ *on* Ω*.*

Theorem 5.1 *Assume that there exists a Lyapunov function defined on a neighbourhood* Ω *of an equilibrium* $\mathbf{x} = 0$ *of system (2.4.26). Then* $\mathbf{x} = 0$ *is stable.*

Proof. There is a ball $B(0, r) \subset \Omega$ (centred at 0 with radius r) such that $0 < V(\mathbf{x})$ on $B(0, r) \setminus 0$ and $V' \geq 0$ on $B(0, r)$. Let us take $0 \neq \mathbf{x}_0 \in B(0, r)$ and consider the flow $\phi(t, \mathbf{x}_0)$. Let $[0, t_{max})$ be the maximal interval of existence of the trajectory. We do not know whether t_{max} is finite or not. Since V is decreasing along trajectories, we have

$$0 < V(\phi(t, \mathbf{x}_0)) \leq V(\mathbf{x}_0), \quad t \in [0, t_{max}),$$

where the left-hand side inequality follows from the fact that $\phi(t, \mathbf{x}_0) \neq 0$ (by Theorem 6.1(i)) and strict positivity of V away from 0). Let $\mu = \min_{\|\mathbf{y}\|=r} V(\mathbf{y})$. Since $V(\mathbf{x}) \to 0$ as $\|\mathbf{x}\| \to 0$, we can find ball $B(0, \delta)$ with $\delta < r$ such that $V(\mathbf{x}) < \mu$ for $\mathbf{x} \in B(0, \delta)$. Then, for $\|\mathbf{y}_0\| < \delta$ we have

$$0 < V(\phi(t, \mathbf{y}_0)) \leq V(\mathbf{y}_0) < \mu, \quad t \in [0, t_{max}),$$

(with t_{max} is not necessarily the same as above). By the definition of μ and continuity of the flow, $\|\phi(t, \mathbf{y}_0)\| \leq r$ for $[0, t_{max})$. Indeed, otherwise there would be $t' > 0$ with $\|\phi(t', \mathbf{y}_0)\| > r$ and, by continuity, for some t'' we would have $\|\phi(t'', \mathbf{y}_0)\| = r$ so that $V(\phi(t'', \mathbf{y}_0)) \geq \mu$.

By the n-dimensional version of Theorem 2.2, this means that $t_{max} = \infty$ and, at the same time, yields stability, as r was arbitrary. ■

Example 5.2 Consider the equation

$$u'' + g(u) = 0$$

where g is a continuously differentiable function for $|u| < k$, with some constant $k > 0$, and $ug(u) > 0$ for $u \neq 0$. Thus, by continuity, $g(0) = 0$. Particular examples include $g(u) = \omega^2 u$ which gives harmonic oscillator of frequency ω, or $g(u) = \sin u$: the undamped simple pendulum. Writing the equation as a system, we get

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -g(x_1). \end{aligned} \tag{3.5.21}$$

It is clear that $(0, 0)$ is an isolated equilibrium point. To construct Lyapunov function we employ mechanical interpretation of the model to find the energy of the system. If we think of g as the restoring force of a spring, the potential energy of the particle at a displacement $u = x_1$ from equilibrium is given by

$$\int_0^{x_1} g(\sigma) d\sigma.$$

On the other hand, the kinetic energy is

$$\frac{1}{2} x_2^2$$

as $x_2 = u'$ which is the velocity of the particle. This suggests to take the total energy of the system as a Lyapunov function

$$V(x_1, x_2) = \frac{1}{2} x_2^2 + \int_0^{x_1} g(\sigma) d\sigma.$$

This function is defined on the region

$$\Omega = \{(x_1, x_2);\ |x_1| < k, x_2 \in \mathbb{R}\}.$$

Clearly, V is positive definite on Ω. Let us calculate the derivative of V along trajectories. We have

$$V'(x_1, x_2) = x_2 x_2' + g(x_1)x_1' = -g(x_1)x_2 + g(x_1)x_2 = 0.$$

Thus, V is a Lyapunov function for (3.5.21) and the equilibrium at $(0, 0)$ is stable.

Actually, we have proved more. For any $\mathbf{x}_0 = (x_{1,0}, x_{2,0}) \in \Omega$, we obtain

$$V(\phi(t, \mathbf{x}_0)) = V(\mathbf{x}_0)$$

for any t. Thus, the orbits are given by implicit equation

$$\frac{1}{2}x_2^2 + \int_0^{x_1} g(\sigma)d\sigma = V(\mathbf{x}_0).$$

Because of the hypotheses on g the integral is positive for both $x_1 > 0$ and $x_1 < 0$; moreover it is an increasing function of $|x_1|$, which is zero at $x_1 = 0$. On the other hand, $V(\mathbf{x}_0) \to 0$ as $\|\mathbf{x}_0\| \to 0$. This means that, for sufficiently small $\|\mathbf{x}_0\|$ ($V(\mathbf{x}_0) < \sup_{|x_1|<k} \int_0^{x_1} g(\sigma)d\sigma$) the orbits are closed orbits symmetric with respect to the y_2 axis and thus solutions are periodic. Hence, $(0, 0)$ is stable but not asymptotically stable.

It is rare to be able to find a Lyapunov function in one go. For left hand side of polynomial type, the method of undetermined coefficients is often employed.

Example 5.3 Consider the system

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -cx_2 - ax_1 - bx_1^3, \end{aligned} \tag{3.5.22}$$

where a, b, c are positive constants. We are looking for a Lyapunov function as a polynomial in two variables. Let us try

$$V(x_1, x_2) = \alpha x_1^2 + \beta x_1^4 + \gamma x_1^3,$$

with $\alpha, \beta, \gamma > 0$. Clearly, $V(\mathbf{x}) > 0$ for $\mathbf{x} \neq 0$. Differentiating V along trajectories, we have

$$\begin{aligned} V'(\mathbf{x}) &= (2\alpha x_1 + 4\beta x_1^3)x_1' + 2\gamma x_2 x_2' \\ &= (2\alpha x_1 + 4\beta x_1^3)x_2 + 2\gamma x_2(-cx_2 - ax_1 - bx_1^3) = (2\alpha - 2\gamma a)x_1 x_2 + (4\beta - 2\gamma b)x_1^3 x_2 - 2\gamma c x_2^2. \end{aligned}$$

Since $c, \gamma > 0$ the last term is non-positive. The first two terms are more difficult, but we have freedom to chose free parameters α and β. Fixing $\gamma > 0$ and setting

$$\alpha = a\gamma, \quad \beta = \frac{\gamma b}{2} = \frac{\alpha b}{2\alpha}$$

we obtain

$$V'(\mathbf{x}) = -2\gamma c x_2^2 \leq 0.$$

Hence $V(\mathbf{x})$ is a Lyapunov function on any open bounded set of $\mathbb{R}^2$ which contains $(0, 0)$ and hence $(0, 0)$ is a stable equilibrium point.

The first Liapunov theorem, Theorem 5.2, ensures stability but not asymptotic stability. From Example 5.2 we see that a possible problem is created by trajectories along which V is constant as these could give rise to periodic orbits which prevent asymptotic stability. The next theorem shows that indeed, preventing the possibility of V being constant in a neighborhood of zero solves the problem, at least partially. The full theory requires, however, a more sophisticated machinery and terminology related to the possible behaviour of the flow for large times.

We begin with introducing the concept of a limit set and to prove some properties of it.

Definition 5.2 *The ω-limit set of the trajectory $\Gamma_{\mathbf{x}_0}$ is the set of all points $\mathbf{p} \in \mathbb{R}^n$ such that there is a sequence $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to \infty$ as $n \to \infty$ for which*

$$\lim_{n\to\infty} \phi(t_n, \mathbf{x}_0) = \mathbf{p}.$$

Similarly, the α-limit set of the trajectory $\Gamma_{\mathbf{x}_0}$ is the set of all points $\mathbf{q} \in \mathbb{R}^n$ such that there is a sequence $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to -\infty$ as $n \to \infty$ for which

$$\lim_{n\to\infty} \phi(t_n, \mathbf{x}_0) = \mathbf{q}.$$

Since for a given equation (2.4.26)) any trajectory is uniquely determined by any point on it (and conversely), sometimes we shall use the notation $\omega(\mathbf{x}_0)$ instead of $\omega(\Gamma_{\mathbf{x}_0})$ (and the same for α-limit sets).

Example 5.4 If $\mathbf{v}_0$ is an equilibrium, then $\Gamma_{\mathbf{v}_0} = \{\mathbf{v}_0\} = \omega(\Gamma_{\mathbf{v}_0}) = \alpha(\Gamma_{\mathbf{v}_0})$. The only ω and α limit sets of scalar equations are equilibria.

Example 5.5 Consider the system in polar coordinates

$$\begin{aligned} r' &= r(1-r^2), \\ \theta' &= 1. \end{aligned}$$

Since $r' > 0$ if $r \in (0,1)$ and $r' < 0$ if $r > 1$, trajectories which start with $0 < r < 1$ and $r > 1$ tend to $r = 1$ which, since $\theta' \neq 0$, is a periodic orbit. The origin $r = 0$ is a stationary point and so $\omega(\{r = 0\}) = \alpha(\{r = 0\}) = (0, 0)$. If $r \neq 0$, then

$$\omega(\Gamma_{(r,\theta)}) = \{(r, \theta);\ r = 1\},$$

and

$$\omega(\Gamma_{(r,\theta)}) = \begin{cases} \{(r, \theta);\ r = 0\} & \text{for} \quad r < 1, \\ \text{does not exist} & \text{for} \quad r > 1 \end{cases}$$

We start with two observations often used in the sequel. They are seemingly obvious but require some reflection.

Remark 5.1 1. How do we prove that $\phi(t, \mathbf{x}_0) \to \mathbf{x}$ as $t \to \infty$? We take arbitrary sequence $(t_n)_{n\in\mathbb{N}}$ and show that it contains a subsequence with $\phi(t_{n_k}, \mathbf{x}_0) \to \mathbf{x}$. In fact, assume that the above holds but $\phi(t, \mathbf{x}_0) \nrightarrow \mathbf{x}$. Then there must be a sequence $(t_n)_{n\in\mathbb{N}}$ for which $\|\phi(t_n, \mathbf{x}_0) - \mathbf{x}\| \geq r$ for some r. But such a sequence cannot contain a subsequence converging to $\mathbf{x}$, so we proved the thesis.

2. Consider a bounded sequence $\mathbf{y}_n = \phi(t_n, \mathbf{x}_0)$ with $t_n \to \infty$. Then, as we know, we have a subsequence $\mathbf{y}_{n_k}$ converging to, say, $\mathbf{y}$. Can we claim that $\mathbf{y} \in \omega(\Gamma_{\mathbf{x}_0})$; that is, is there a sequence $(t'_n)_{n\in\mathbb{N}}$ converging to ∞ such that $\phi(t'_n, \mathbf{x}_0)) \to \mathbf{y}$? The reason it is not obvious is that selecting $\mathbf{y}_{n_k} = \phi(t_{n_k}, \mathbf{x}_0)$ could create a sequence $\{t_{n_1}, \ldots, t_{n_k} \ldots\}$ which does not diverge to ∞ as we do not have any control on how the latter is ordered with respect to the former. First, we note that we can assume that $t_n - t_{n-1} \geq 1$ for any n. Indeed, $(t_n)_{n\in\mathbb{N}}$ must contain a subsequence with this property so, if necessary, we can select such a subsequence for further analysis. Next, we note that we can assume that $\mathbf{y}_n \neq \mathbf{y}_m$ for $n \neq m$ as otherwise, by Theorem **??**, the orbit would be periodic since $t_n \neq t_m$ and in the case of periodic orbit with period T we can take $t_n = nT$. Thus, $\mathbf{y}_{n_k} \neq \mathbf{y}_{n_l}$ and hence $t_{n_k} \neq t_{n_l}$ for $k \neq l$. Now, $(n_k)_{k\in\mathbb{N}}$ is an infinite sequence of mutually different natural numbers and thus we can select a monotonic subsequence $(n_{k_l})_{l\in\mathbb{N}}$.

Consider a nested sequence of balls $B(\mathbf{y}, 1/N)$. For each N there is n_N such that $\mathbf{y}_{n_k} = \phi(t_{n_k}, \mathbf{x}_0)$ for all $n_k \geq n_N$. In particular, each set $\{t_{n_k};\ n_k \geq n_N\}$ is infinite. Now, the crucial observation is that an infinite subset of a sequence $(t_n)_{n\in\mathbb{N}}$ converging to ∞ must be unbounded and thus must contain a subsequence converging to ∞. Indeed, from the definition of convergence to ∞, for each M only finite number of elements of the sequence is smaller that M. If so, having selected $\phi(t_N, \mathbf{x}_0)$ in $B(\mathbf{y}, N^{-1})$, we select $\phi(t_{N+1}, \mathbf{x}_0)$ in $B(\mathbf{y}, (N+1)^{-1})$ with $t_{N+1} > t_N + 1$ as an infinite collection of $t_{n_k}s$ corresponding to $\mathbf{y}_{n_k} \in B(\mathbf{y}, (N+1)^{-1})$ must contain arbitrary large $t_{n_k}s$. This shows that we have $\phi(t_N, \mathbf{x}_0) \to \mathbf{y}_0$ with $t_N \to \infty$ as $N \to \infty$.

Lemma 5.1 *ω-limit sets have the following properties:*

1. *If Γ is bounded, then $\omega(\Gamma)$ is non-empty;*
2. *An ω-limit set is closed;*
3. *If $\mathbf{y} \in \omega(\Gamma)$, then $\Gamma_{\mathbf{y}} \subset \omega(\Gamma)$; that is, $\omega(\Gamma)$ is invariant.*
4. *If Γ is bounded, then $\omega(\Gamma)$ is connected;*
5. *If $\mathbf{z} \in \omega(\Gamma)$ and $\Gamma \cap \omega(\Gamma') \neq \emptyset$, then $\mathbf{z} \in \omega(\Gamma')$.*
6. *If $\Gamma_{\mathbf{y}}$ is bounded, then $\phi(t, \mathbf{y}) \to \omega(\Gamma_{\mathbf{y}})$ as $t \to \infty$ in the sense that for each $\epsilon > 0$ there is $t > 0$ such that for every $t' > t$ there is $\mathbf{p} \in \omega(\Gamma_{\mathbf{y}})$ (possibly depending on t) which satisfies $\|\phi(t, \mathbf{y}) - \mathbf{p}\| < \epsilon$.*

The same properties are valid for α-limit sets.

Proof. ad 1) Let $\mathbf{x} \in \Gamma$. Taking e.g. $\phi(n, \mathbf{x})$ we obtain a bounded sequence of points which must have a converging subsequence with the limit in $\omega(\Gamma)$ by Remark 5.1.

ad 2) Let $(\mathbf{p}_n)_{n\in\mathbb{N}}$ be a sequence of points of $\omega(\Gamma_{\mathbf{x}})$ converging to $\mathbf{p} \in \mathbb{R}^n$. We must prove that there is a sequence $t_n \to \infty$ such that $\phi(t_n, \mathbf{x}) \to \mathbf{p}$. We proceed as follows. For $\epsilon = 1/2$ we can find $\mathbf{p}_{n_1}$ and t_{n_1} such that

$$\|\mathbf{p} - \phi(t_{n_1}, \mathbf{x})\| \leq \|\mathbf{p} - \mathbf{p}_{n_1}\| + \|\mathbf{p}_{n_1} - \phi(t_{n_1}, \mathbf{x})\| \leq 1/2 + 1/2 = 1$$

Similarly, for $\epsilon = 1/4$ we can find $\mathbf{p}_{n_2}$ and $t_{n_2} > t_{n_1} + 1$ such that

$$\|\mathbf{p} - \phi(t_{n_2}, \mathbf{x})\| \leq \|\mathbf{p} - \mathbf{p}_{n_2}\| + \|\mathbf{p}_{n_2} - \phi(t_{n_2}, \mathbf{x})\| \leq 1/4 + 1/4 = 1/2,$$

and, by induction, for any given k we can find $\mathbf{p}_{n_k}$ and $t_{n_k} > t_{n_{k-1}} + 1$ such that

$$\|\mathbf{p} - \phi(t_{n_k}, \mathbf{x})\| \leq \|\mathbf{p} - \mathbf{p}_{n_k}\| + \|\mathbf{p}_{n_k} - \phi(t_{n_k}, \mathbf{x})\| \leq 1/2k + 1/2k = 1/k.$$

Since the sequence $(t_{n_k})_{k\in\mathbb{N}}$ is infinite and increasing by 1 at each step, it must diverge to infinity. Thus, $\mathbf{p} \in \omega(\Gamma_{\mathbf{x}})$.

ad 3) Let $\mathbf{y} \in \omega(\Gamma_{\mathbf{x}})$ and consider arbitrary $\mathbf{q} \in \Gamma_{\mathbf{y}}$. Thus $\mathbf{q} = \phi(t', \mathbf{y})$. Since $\mathbf{y}$ is in the ω-limit set,

$$\mathbf{y} = \lim_{t_n \to \infty} \phi(t_n, \mathbf{x}).$$

But

$$\phi(t_n + t', \mathbf{x}) = \phi(t', \phi(t_n, \mathbf{x})) \to \phi(t', \mathbf{y}) = \mathbf{q}$$

by continuity of the flow with respect to the initial value. Since $t' + t_n \to \infty$, we obtain that $\mathbf{q} \in \omega(\Gamma_{\mathbf{y}})$ and since $\mathbf{q} \in \omega(\Gamma_{\mathbf{x}})$ was arbitrary, the thesis follows.

ad 4.) By 1.), the set $\omega(\Gamma)$ is bounded and closed. Thus, if it was not connected, then $\omega(\Gamma) = A \cup B$ with A, B closed bounded and a distance $d > 0$ apart. Fix $\mathbf{x} \in \Gamma$ and chose $\mathbf{v} \in A$. There is $t_n \to \infty$ such that $\mathbf{x_n} := \phi(t_n, \mathbf{x}) \to \mathbf{v}$. Thus, for sufficiently large n, the distance between $\mathbf{x}_n$ and B is not less than $3d/4$. Let $\mathbf{w} \in B$. For each $\mathbf{x}_n$ we can find $t'_n > 0$ such that the distance between $\mathbf{y}_n = \phi(t'_n, \mathbf{x}_n) = \phi(t_n + t'_n, \mathbf{x})$ and $\mathbf{w}$ is not greater than $d/4$, thus the distance between A and $\mathbf{y}_n$ is at least $3d/4$. On the other hand, from the continuity of the flow, for each n there is a point $\mathbf{z}_n = \phi(t''_n, \mathbf{x})$ whose distance from A is $d/2$. The set of such points is bounded and thus have a convergent subsequence whose limit $\mathbf{z} \in \omega(\Gamma)$ is at the distance $d/2$ from A, which contradicts the assumption that the components of $\omega(\Gamma)$ are d apart.

ad 5.) Let $\mathbf{y} \in \omega(\Gamma)$ and $\mathbf{y} = \lim_{t'_n \to \infty} \phi(t'_n, \mathbf{x})$ with $\mathbf{x} \in \Gamma'$. Let us fix $\epsilon > 0$. From continuity of the flow with respect to the initial condition (Lemma 2.1), we know that for a given ϵ and T, there is $\delta_{T,\epsilon}$ such that $\|\phi(t, \mathbf{x}_1) - \phi(t, \mathbf{x}_2)\| < \epsilon$ provided $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta_{T,\epsilon}$ for all $0 \leq t \leq T$.

For this given $\epsilon > 0$ we find t_n such that $\|\mathbf{z} - \phi(t_n, \mathbf{y})\| < \epsilon$ and also t'_n such that $\|\mathbf{y} - \phi(t'_n, \mathbf{x}_0)\| < \delta_{t_n,\epsilon}$ (since $\mathbf{y} \in \omega(\Gamma')$). Hence

$$\|\mathbf{z} - \phi(t_n, \phi(t'_n, \mathbf{x}_0))\| \leq \|\mathbf{z} - \phi(t_n, \mathbf{y})\| + \|\phi(t_n, \mathbf{y}) - \phi(t_n, \phi(t'_n, \mathbf{x}_0))\| < 2\epsilon$$

but since $\phi(t_n, \phi(t'_n, \mathbf{x}_0)) = \phi(t_n + t'_n, \mathbf{x}_0)$, we see that $\mathbf{z} \in \omega(\Gamma')$ (we note that the sequence $\tau_n := t_n + t'_n$ can be made increasing with n and thus convergent to ∞ as t_n and t'_n are).

ad 6.) Suppose that the statement is false; that is, there is $\epsilon > 0$ and a sequence $t_n \to \infty$ with $\|\phi(t_n, \mathbf{y}) - \mathbf{p}\| > \epsilon$ for all $\mathbf{p} \in \omega(\Gamma_\mathbf{y})$. This means that the sequence $\phi(t_n, \mathbf{y})$ stays at least ϵ away from $\omega(\Gamma_\mathbf{y})$. But the orbit is bounded so, by Remark 5.1, there is a converging subsequence $\phi(t_{n_k}, \mathbf{y})$ of this sequence which, by the definition of $\omega(\Gamma_\mathbf{y})$ must belong to it. ■

Theorem 5.2 *Assume that there exists a Lyapunov function defined on a neighbourhood Ω of an equilibrium $\mathbf{x} = 0$ of system (2.4.26), which additionally satisfies*

$$V' < 0 \quad in \quad \Omega \setminus \{0\}. \tag{3.5.23}$$

Then $\mathbf{x} = 0$ is asymptotically stable.

Definition 5.3 *A Lyapunov function satisfying assumptions of this theorem is called* strict Lyapunov function.

Proof. Using the notation of the previous proof, we consider $\phi(t, \mathbf{y}_0)$ with $\|\mathbf{y}_0\| < \delta$ so that the solution stays in the closed ball $\overline{B(0,r)}$. Since this set is compact, we have sequence $(t_n)_{n\in\mathbb{N}}$, $t_n \to \infty$ as $n \to \infty$ such that $\phi(t_n, \mathbf{y}_0) \to \mathbf{z} \in \overline{B(0,r)}$. We have to prove that $\mathbf{z} = 0$. To this end, first observe that $V(\phi(t, \mathbf{y}_0)) > V(\mathbf{z})$ for all t as V decreases along trajectories and $V(\phi(t_n, \mathbf{y}_0) \to V(\mathbf{z})$ by continuity of V. If $\mathbf{z} \neq 0$ (and there are no other equilibria in $\overline{B(0,r)}$), we consider $\phi(t, \mathbf{z})$ which must satisfy $V(\phi(t, \mathbf{z}) < V(\mathbf{z})$. By continuity of the flow with respect to the initial condition and continuity of V, if $\mathbf{x}$ is close enough to $\mathbf{z}$, then for some $t > 0$ (not necessarily all) we will have also $V(\phi(t, \mathbf{x})) < V(\mathbf{z})$. We take $\mathbf{x} = \phi(t_n, \mathbf{y}_0)$ for t_n large enough. But then we obtain

$$V(\mathbf{z}) > V(\phi(t, \phi(t_n, \mathbf{y_0}))) = V(\phi(t + t_n, \mathbf{y_0})) > V(\mathbf{z})$$

which is a contradiction. Thus $\mathbf{z} = 0$. This shows asymptotic stability. Indeed, if there was a sequence $(t_n)_{n\in\mathbb{N}}$ converging to infinity, for which $\phi(t, \mathbf{y}_0)$ was not converging to zero (that is, staying outside some ball $B(0, r_0)$), then as above we could pick a subsequence of $\phi(t_n, \mathbf{y}_0)$ converging to some $\mathbf{z}$ which, by the above, must be 0. ■

Example 5.6 Consider the system

$$\begin{aligned} x'_1 &= -x_1 + x_1^2 - 2x_1x_2, \\ x'_2 &= -2x_2 - 5x_1x_2 + x_2^2, \end{aligned} \tag{3.5.24}$$

The point $(0, 0)$ clearly is a stationary point. Let us investigate its stability. We try the simplest Lyapunov function

$$V(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2).$$

We obtain

$$\begin{aligned} V'(x_1, x_2) &= x_1x'_1 + x_2x'_2 = x_1(-x_1 + x_1^2 - 2x_1x_2) + x_2(-2x_2 - 5x_1x_2 + x_2^2) \\ &= -x_1^2(1 - x_1 + 2x_2) - x_2^2(2 + 5x_1 - x_2). \end{aligned}$$

Hence V is a strict Lyapunov function provided $2 + 5x_1 - x_2 > 0$ and $1 - x_1 + 2x_2 > 0$ in some neighbourhood of $(0, 0)$. We see that this set, say Ω, is a sector containing $(0, 0)$. Hence, the origin is asymptotically stable.

Let us consider another example which, in conjunction with the previous one, provide background for a refinement of the theory developed so far.

Example 5.7 Consider a more realistic version of the nonlinear oscillator equation, which includes resistance of the medium proportional to the velocity.

$$u'' + u' + g(u) = 0.$$

This equation is called the Liénard equation. We adopt the same assumptions as before: g is a continuously differentiable function for $|u| < k$, with some constant $k > 0$, and $ug(u) > 0$ for $u \neq 0$. Thus, by continuity, $g(0) = 0$. Writing the equation as a system, we get

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -g(x_1) - x_2. \end{aligned} \tag{3.5.25}$$

Again, $(0,0)$ is an isolated equilibrium point. Since this equation differs from the previously considered one by a dissipative term, the total energy of the system is a good initial guess for a Lyapunov function. Hence, we take

$$V(x_1, x_2) = \frac{1}{2}x_2^2 + \int_0^{x_1} g(\sigma)d\sigma.$$

This function is defined and positive on the region

$$\Omega = \{(x_1, x_2);\ |x_1| < k, x_2 \in \mathbb{R}\},$$

with the derivative along trajectories given by

$$V'(x_1, x_2) = -x_2^2 + x_2x_2' + g(x_1)x_1' = -x_2^2 - g(x_1)x_2 + g(x_1)x_2 = -x_2^2.$$

Thus, again V is a Lyapunov function for (3.5.25) and the equilibrium at $(0,0)$ is stable. However, it fails to be a strict Lyapunov function as there is no neighbourhood of $(0,0)$ on which V' is strictly positive. Now, if we look closer at this example, we see that we should be able to prove something more. Namely, if we can ensure that a trajectory stays away from $L = \{(x_1, x_2);\ x_2 = 0, x_1 \neq 0\}$ then, following the proof of Theorem 5.2, we obtain that it must converge to $(0,0)$. On the other hand, at any point of L the vector field is transversal to L so the trajectory cannot stay on L as then the field would have to be tangent to L. Thus, it is to be expected that the trajectory must eventually reach $(0,0)$. We shall provide a rigorous and more general result of this type below.

Theorem 5.3 *(La Salle invariance principle) Let $\mathbf{y} = 0$ be a stationary point of (2.4.26) and let V be a Lyapunov function on some neighbourhood $\Omega \ni 0$. If, for $\mathbf{x} \in \Omega$, $\Gamma_{\mathbf{x}}^+$ is bounded with limit points in Ω and M is the largest invariant set of*

$$E = \{\mathbf{x} \in \Omega;\ V'(\mathbf{x}) = 0\}, \tag{3.5.26}$$

then

$$\phi(t, \mathbf{x}) \to M, \quad t \to \infty. \tag{3.5.27}$$

Proof. By assumptions, for any $\mathbf{x} \in \Omega$ satisfying the assumption of the theorem, $\emptyset \neq \omega(\Gamma_{\mathbf{x}}) \subset \Omega$. Since V is a Lyapunov function, $V(\phi(t, \mathbf{x}))$ in a non-increasing function of t which is bounded below by zero. Hence, there is $c \geq 0$ such that

$$\lim_{t\to\infty} V(\phi(t, \mathbf{x})) = c.$$

Now let $\mathbf{y} \in \omega(\Gamma_{\mathbf{x}})$. Then, for some $(t_n)_{n\in\mathbb{N}}$ converging to ∞ as $n \to \infty$ we have $\phi(t_n, \mathbf{x}) \to \mathbf{y}$. On the other hand, by continuity of V, we have

$$\lim_{t\to\infty} V(\phi(t_n, \mathbf{x})) = c.$$

Consequently, V is constant on $\omega(\Gamma_{\mathbf{x}})$.

Now, by Lemma 5.1(3), $\omega(\Gamma_{\mathbf{x}})$ is invariant so that if $\mathbf{y} \in \omega(\Gamma_{\mathbf{x}})$, then $\phi(t, \mathbf{y}) \in \omega(\Gamma_{\mathbf{x}})$ for all t. Thus, $V(\phi(t, \mathbf{y})) = c$ for all t and $\mathbf{y} \in \omega(\Gamma_{\mathbf{x}})$. Thus,

$$V'(\mathbf{y}) = 0, \quad for\ \mathbf{y} \in \omega(\Gamma_{\mathbf{x}})$$

and so

$$\omega(\Gamma_{\mathbf{x}}) \subset M \subset E.$$

But $\phi(t, \mathbf{x}) \to \omega(\Gamma_{\mathbf{x}})$ as $t \to \infty$, so $\phi(t, \mathbf{z}) \to M$ as $t \to \infty$ for all $\mathbf{z} \in \Omega$ satisfying $\omega(\Gamma_{\mathbf{z}}) \subset \Omega$. ■

Remark 5.2 *A class of sets in Ω with forward trajectories in Ω is given by*

$$V_k = \{\mathbf{x};\ V(\mathbf{x}) < k\}.$$

Indeed, since V is non-increasing along trajectories, $V(\phi(t, \mathbf{x})) \leq V(\mathbf{x}) < k$ provided $\mathbf{x} \in V_k$.

Remark 5.3 *La Salle principle gives immediate proof of Theorem 5.2. Indeed, in the domain of applicability of this theorem, $M = E = \{0\}$.*

Corollary 5.1 *Assume that there is a Lyapunov function for (2.4.26) defined on the whole $\mathbb{R}^n$ which satisfies additionally $V(\mathbf{y}) \to \infty$ as $\|\mathbf{y}\| \to \infty$. If 0 is the only invariant set of $E = \{\mathbf{x} \in \mathbb{R}^n;\ V'(\mathbf{x}) = 0\}$, then 0 is globally asymptotically stable.*

Proof. From the properties of the Lyapunov function, we have

$$V(\phi(t, \mathbf{y})) \leq V(\mathbf{y}), \quad \mathbf{y} \in \mathbb{R}^n,$$

independently of t. From the assumption on the behaviour of V at infinity, $\phi(t, \mathbf{y})$ must stay bounded and thus exist for all t. But then the limit points must belong to the set $\Omega = \mathbb{R}^n$ and the La Salle principle can be applied. ■

Example 5.8 Consider the so-called van der Pol equation

$$z'' - az'(z^2 - 1) + z = 0, \tag{3.5.28}$$

where $a > 0$ is a constant. In this case it is easier to work with the so-called Lièнard coordinates which are applicable to any equation of the form

$$x'' + f(x)x' + g(x) = 0.$$

Let us define $F(x) = \int^x f(\xi)d\xi$. Then

$$dF/dt = f(x)x'.$$

Hence, if we define $x_1 = x$ and $x_2 = x_1' + F(x_1)$, then $x_2' = x_1'' + f(x_1)x_1' = -g(x_1)$. The differential equation can then be written as

$$\begin{aligned} x_1' &= x_2 - F(x_1), \\ x_2' &= -g(x_1). \end{aligned}$$

In our case, we obtain

$$\begin{aligned} x_1' &= x_2 + a\left(\frac{1}{3}x_1^3 - x_1\right), \\ x_2' &= -x_1. \end{aligned}$$

Let us use the standard Lyapunov function

$$V(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2).$$

With this choice we get

$$V'(x_1,x_2) = x_1\left(x_2 + a\left(\frac{1}{3}x_1^3 - x_1\right)\right) - x_1x_2 = ax_1^2\left(\frac{1}{3}x_1^2 - 1\right).$$

Thus, $V' \leq 0$ for $x_1^2 < 3$. The largest domain of the form

$$V_k = \{(x_1,x_2);\ V(x_1,x_2) < k\}$$

which lies entirely in the region $\{(x_1,x_2);\ V' \leq 0\}$ is given by $V_{3/2}$. Furthermore, $V' = 0$ on $x_1 = 0$ and on this line $x_1' = x_2$. Hence, the trajectories will stay on this line only if the x_1-coordinate of the tangent is zero; that is, only of $x_1' = x_2 = 0$. Thus, the largest invariant subset of $V_{3/2} \cap \{(x_1,x_2);\ V' = 0\}$ is $(0,0)$. Thus, by the La Salle principle, $(0,0)$ is asymptotically stable and $V_{3/2}$ is a basin of attraction.

Example 5.9 Consider the equation

$$z'' + 2az' + z + z^3 = 0, \tag{3.5.29}$$

where a is a constant satisfying $0 < a < 1$. Equivalent system is given by

$$\begin{aligned} x_1' &= x_2, \\ x_2' &= -x_1 - 2ax_2 - x_1^3. \end{aligned} \tag{3.5.30}$$

The origin $(0,0)$ is the only equilibrium. If $a = 0$, then (3.5.29) is the same as in the Example 6.2 and thus we know that

$$V(x_1,x_2) = \frac{x_2^2}{2} + \frac{x_1^2}{2} + \frac{x_1^4}{4}$$

is the first integral (energy) and $V' = 0$ on the trajectories. The addition of $2az'$ makes the system dissipative and thus it is reasonable to take V as the trial Lyapunov function for (3.5.30). We get

$$V'(x_1,x_2) = -2ax_2^2 \leq 0$$

for any (x_1,x_2). Let us apply Corollary 5.1. It is clear that $V(x_1,x_2) \to \infty$ as $\|(x_1,x_2)\| \to \infty$. Furthermore, $V' = 0$ on $x_2 = 0$. We have to find the largest invariant subset of this set. To stay on $x_2 = 0$ we must have $x_2' = 0$. But, if $x_2 = 0$, then $x_1' = 0$ hence $x_1 = \textit{constant}$. This, from the second equation of (3.5.30) we obtain $x_1 = 0$. Consequently, $(0,0)$ is the largest invariant subset of $\{(x_1,x_2);\ V' = 0\}$ and thus $(0,0)$ is globally asymptotically stable.

Example 5.10 Stability by linearization. W shall give a proof of Theorem 4.1 1 using the Lyapunov function method. Consider again

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{3.5.31}$$

If $\mathbf{f}$ has continuous partial derivatives of the first order in some neighbourhood of $\mathbf{y}^0$, then

$$\mathbf{f}(\mathbf{x} + \mathbf{y}^0) = \mathbf{f}(\mathbf{y}^0) + \mathcal{A}\mathbf{x} + \mathbf{g}(\mathbf{x}) \tag{3.5.32}$$

where

$$\mathcal{A} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{y^0}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{y^0}) \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n}(\mathbf{y^0}) & \dots & \frac{\partial f_n}{\partial x_n}(\mathbf{y^0}) \end{pmatrix},$$

and $\mathbf{g}(\mathbf{x})/\|\mathbf{x}\|$ is continuous in some neighbourhood of $\mathbf{y}^0$ and vanishes at $\mathbf{x} = \mathbf{y}^0$. This follows from the Taylor expansion theorem. Note that if $\mathbf{y}^0$ is an equilibrium of (3.5.31), then $f(\mathbf{y}^0) = 0$ and we can write

$$\mathbf{x}' = \mathcal{A}\mathbf{x} + o(\|\mathbf{x}\|). \tag{3.5.33}$$

where $\mathbf{x} = \mathbf{y} - \mathbf{y}^0$.

Theorem 5.4 *Suppose that* $\mathbf{f}$ *is differentiable function in some neighbourhood of the equilibrium point* $\mathbf{y}^0$. *Then, the equilibrium point* $\mathbf{y}^0$ *is asymptotically stable if all the eigenvalues of the matrix* $\mathcal{A}$ *have negative real parts, that is, if the equilibrium solution* $\mathbf{x}(t) = \mathbf{0}$ *of the linearized system is asymptotically stable.*

Proof. We shall give the proof for the case when $\mathcal{A}$ has distinct eigenvalues. The general case also can be proved using Lyapunov function method but it is much more involved.

Let $\{\lambda_1, \ldots, \lambda_n\}$ be distinct eigenvalues of $\mathcal{A}$ with $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ being the corresponding eigenvalues (since the eigenvalues are distinct, they must be simple so that to each there corresponds exactly one eigenvector). Now, denoting by $< \cdot, \cdot >$ the dot product in $\mathbb{C}^n$

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{n} x_i y_i,$$

we have

$$< \mathbf{y}, \mathcal{A}\mathbf{x} >=< \mathcal{A}^T\mathbf{y}, \mathbf{x} >$$

where A^T is the transpose of $A = \{a_{ij}\}_{1\le i,j\le n}$, $A^T\{a_{ji}\}_{1\le i,j\le n}$ (thanks to the fact that entries of A are real). A^T has the same eigenvalues as A; denote by $\{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ the corresponding eigenvectors. It is easy to see that $< \mathbf{f}_j, \mathbf{e}_i >= 0$ provided $j \neq i$. Indeed,

$$\lambda_i < \mathbf{f}_i, \mathbf{e}_j >=< A^T\mathbf{f}_i, \mathbf{e}_j >=< \mathbf{f}_i, A\mathbf{e}_j >= \lambda_j < \mathbf{f}_i, \mathbf{e}_j >$$

so that $(\lambda_i - \lambda_j) < \mathbf{f}_i, \mathbf{e}_j >= 0$ and the statement follows if $\lambda_i \neq \lambda_j$. It is then possible to normalize the eigenvectors so that

$$< \mathbf{f}_i, \mathbf{e}_j >= \delta_{ij},$$

(that is, 1 for $i = 1$ and 0 for $i \neq j$). We can expand any $\mathbf{x}$ as

$$\mathbf{x} = \sum_{i=1}^{n} < \mathbf{f}_i, \mathbf{x} > \mathbf{e}_i.$$

Then

$$\mathbf{x}' = \sum_{i=1}^{n} \frac{d}{dt} < \mathbf{f}_i, \mathbf{x} > \mathbf{e}_i,$$

and

$$A\mathbf{x} = \sum_{i=1}^{n} \lambda_i < \mathbf{f}_i, \mathbf{x} > \mathbf{e}_i$$

so that

$$\frac{d}{dt} < \mathbf{f}_i, \mathbf{x} >= \lambda_i < \mathbf{f}_i, \mathbf{x} > +o(\|\mathbf{x}\|)$$

(as multiplying $o(\|\mathbf{x}\|)$ by $\mathbf{e}$ does not change the asymptotic behaviour of the '$o(\cdot)$' symbol).

This allows to define a Lyapunov function. Let $\alpha_1, \ldots, \alpha_n$ be positive numbers and put

$$V(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i \overline{< \mathbf{f}_i, \mathbf{x} >} < \mathbf{f}_i, \mathbf{x} > .$$

This is clearly differentiable function which is positive for $\mathbf{x} \neq 0$. Differentiating along trajectories, we get

$$\begin{aligned} V'(\mathbf{x}) &= \sum_{i=1}^{n} \alpha_i \left(\frac{d}{dt} \overline{< \mathbf{f}_i, \mathbf{x} >} < \mathbf{f}_i, \mathbf{x} > \right) \\ &= +\sum_{i=1}^{n} \alpha_i \left(\left(\frac{d}{dt} \overline{< \mathbf{f}_i, \mathbf{x} >} \right) < \mathbf{f}_i, \mathbf{x} > + \overline{< \mathbf{f}_i, \mathbf{x} >} \left(\frac{d}{dt} < \mathbf{f}_i, \mathbf{x} > \right) \right) \\ &= \sum_{i=1}^{n} \alpha_i \left(\bar{\lambda}_i + \lambda_i \right) \overline{< \mathbf{f}_i, \mathbf{x} >} < \mathbf{f}_i, \mathbf{x} > +o(\|\mathbf{x}\|^2). \end{aligned}$$

Since $\bar{\lambda_i} + \lambda_i = 2\Re\lambda_i < 0$, the first term is negative of second order and the other term is of higher order than 2,and thus for sufficiently small $\|\mathbf{x}\|$ the derivative $V'(\mathbf{x})$ is strictly negative in some neighbourhood Ω of 0. Hence, 0 is asymptotically stable. Constants α_i can be changed to fine-tune the estimate a basin of attraction.

Example 5.11 Instability from linearization Under assumption of Theorem 5.4, if at least one eigenvalue has positive real part, then the zero equilibrium solution is unstable.

We prove the theorem for $n = 2$. There are two cases. If both eigenvalues have negative real parts then the result follows by reversing the direction of time. In this case, the system $\tilde{\mathbf{y}}_\tau = -\mathbf{f}(\mathbf{y}(-t)) =: (\tilde{\mathbf{f}}(\tau)$, where $\tau = -t$ has both eigenvalues with real parts negative and, given $\mathbf{y}_0$ with $\|\mathbf{y}_0\| = r$, for r sufficiently small so that $\mathbf{y}_0$ is in the basin of attraction of $(0, 0)$, for any $\delta > 0$ we find τ_δ with $\|\tilde{\phi}(\tau_\delta, \mathbf{y}_0) < \delta\|$. But the flow $\phi(t, \mathbf{x_0})$ of the original system is related to $\tilde{\phi}(\tau, \mathbf{x}_0)$ by $\tilde{\phi}(\tau, \mathbf{x}_0) = \phi(-t, \mathbf{x}_0)$ so that

$$\mathbf{y}_0 = \phi(\tau_\delta, \tilde{\phi}(\tau_\delta, \mathbf{y}_0))$$

and we see that a fixed distance from the origin can be attained starting from arbitrarily small neighbourhood of the origin.

If one has positive, one negative real part, then they must be real (as for a matrix with real entries eigenvalues come as pairs of elements complex conjugate to each other. So, let $\lambda_1 \leq 0 < \lambda_2$. By diagonalization, the system can be written as

$$\begin{aligned} y_1' &= \lambda_1 y_1 + g_1(y_1, y_2), \\ y_2 &= \lambda_2 y_2 + g_2(y_1, y_2), \end{aligned}$$

where the function $\mathbf{g} = (g_1, g_2)$ has the same properties as in Theorem 5.4.

Consider the function

$$V(y_1, y_2) = \frac{1}{2}(y_2^2 - y_1^2).$$

Suppose $\mathbf{y}(t) = (y_1(t), y_2(t))$ is a solution satisfying $\|\mathbf{y}(t)\| < \epsilon$, where ϵ is the radius of the ball in which $\|\mathbf{g}(\mathbf{y})\| \leq m\|\mathbf{y}\|$. Evaluating the derivative of V along this trajectory, we get

$$\begin{aligned} V'(\mathbf{y}) &= \lambda_2 y_2^2 + y_2 g_2(y_1, y_2) - \lambda_1 y_1^2 - y_1 g_1(y_1, y_2) \\ &\geq \lambda_2 y_2^2 - m\|\mathbf{y}\|(|y_1| + |y_2|) - \lambda_1 y_1^2 \\ &\geq (\lambda_2 - m)y_2^2 - 2m|y_1||y_2| - (\lambda_1 + m)y_1^2. \end{aligned}$$

Next, consider the set

$$\Omega = \{(y_1, y_2);\ y_2 > |y_1|\}.$$

In this region we have

$$V'(\mathbf{y}) = (\lambda_2 - 4m)y_2^2 - \lambda_1 y_1^2.$$

where the second term is nonnegative. Now, we can chose ϵ so small that $4m < \lambda_2$, hence on $\Omega \cap \{\|y\| < \epsilon\}$ we have $V' > 0$. Suppose $\mathbf{y}^0 \in \Omega \cap \{\|y\| < \epsilon\}$. Then since $V' > 0$, the solution stays in Ω as V must increase keeping $y_2^2 > y_1^2$. But then V' along the trajectory is strictly positive (bounded away from zero by a positive constant). This means that the trajectory will eventually reach the boundary $\|\mathbf{y}\| = \epsilon$ showing that the zero equilibrium is unstable.

Chapter 4

The Poincaré-Bendixon Theory

The linearization theorem, Theorem 5.4, may suggest that locally nonlinear dynamics is the same as linear. This is, however, false even in 2 dimensions. The Poincaré-Bendixon theory provides a complete description of two-dimensional dynamics by giving full classification of possible limit sets for planar autonomous systems.

1 Preliminaries

Let $\phi(t, \mathbf{x})$ be the flow generated by the system

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}), \tag{4.1.1}$$

where $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$. Throughout this chapter we assume that $\mathbf{f}$ is a differentiable function. By a *local transversal* to ϕ we mean a line segment L which all trajectories cross from one side. In other words, the field $\mathbf{f}$ always forms either an acute or an obtuse angle with the normal to L.

Lemma 1.1 *If $\mathbf{x}_0$ is not a stationary point of (4.1.1), then there is always possible to construct a local transversal in a neighbourhood of $\mathbf{x}_0$.*

Proof. If $\mathbf{x}_0$ is not stationary point, then $\mathbf{v} = \mathbf{f}(\mathbf{x}_0) \neq 0$. Take coordinate system with origin at $\mathbf{x}_0$ and axes parallel to $\mathbf{v}$ and $\mathbf{w} \perp \mathbf{v}$. Let (v, w) be the coordinates of a point in this system. Then (4.1.1) can be written as

$$\begin{aligned} v' &= a + O(\|(v, w) - \mathbf{x}_0\|), \\ w' &= O(\|(v, w) - \mathbf{x}_0\|). \end{aligned}$$

Here $a \neq 0$ is the norm of $\mathbf{v}$. Now we can chose line L through $\mathbf{x}_0$ along $\mathbf{w}$. As long as we are close to $\mathbf{x}_0$, $v' \neq 0$ and the trajectories cross L in the same direction (positive if $a > 0$ and negative otherwise). □

Having found a local transversal L at $\mathbf{x}$ we can construct a *flow box* around $\mathbf{x}$ by taking a collection of trajectories with starting points at L and t running from $-\delta$ to δ with δ small enough that all trajectories exist over this time interval; that is

$$V_{\mathbf{x}} = \{\mathbf{y};\ \mathbf{y} = \phi(t, \mathbf{z}),\ \mathbf{z} \in L, t \in (-\delta, \delta)\}.$$

An important property of flow box is that if a trajectory starts close enough to $\mathbf{x}$, it crosses L (either forward or backward). Precisely speaking, we have

Lemma 1.2 *Let L be a transversal at $\mathbf{x}$. There is a neighbourhood D of $\mathbf{x}$ such that for any $\mathbf{y} \in D$ there are $t_1 < t_2$ such that the trajectory segment $\Gamma_{t_1,t_2} = \{\mathbf{z};\ \mathbf{z} = \phi(t, \mathbf{y}), t \in (t_1, t_2)\} \subset D$ and $\Gamma_{t_1,t_2} \cap L \neq \emptyset$.*

Proof. We chose coordinates (z_1, z_2) as in the proof of Lemma 1.1 so that $\mathbf{x} = (0,0)$ and $\mathbf{f}(\mathbf{x}) = (a, 0)$ where we select $a > 0$. Then L is on the x_2 axis. Let D_δ be the ball $\|(z_1, z_2)\| = |z_1| + |z_2| < \delta$ (that is a square with diagonals on L and $\mathbf{f}(\mathbf{x})$. We can chose δ small enough for the following condition to hold:

a) The slope of $\mathbf{f}(\mathbf{z})$ is strictly between -1 and 1; that is, $|f_2(\mathbf{z})/f_1(\mathbf{z})| < 1$,

b) $f_1(\mathbf{z}) > a/2$,

for $\mathbf{z} \in D_\delta$. This is possible by continuity of $\mathbf{f}$ in a neighbourhood of $\mathbf{x}$. Notice, in particular, that $\mathbf{f}$ is always transversal to the sides of D_δ and pointing outward at the righ-hand side and inward at the left-hand side of it. Hence, a trajectory can only leave D_δ at the right and enter at the left. Now, clearly, for any solution $\mathbf{z}(t)$

$$z_1(t) - z_1(0) = \int_0^t f_1(\mathbf{z}(s))ds > \frac{a}{2}t.$$

Since the maximum value of $z_1(t) - z_1(0)$ is 2δ, the solution starting from any point $\mathbf{z}_0 \in D_\delta$ must leave it in time shorter than $4\delta/a$. Let t_2 be the smallest value at which $\phi(t, \mathbf{z}_0)$ intersects the right-hand side of D_δ. Similarly, $\phi(t, \mathbf{z}_0)$ intersects the left-hand side of D_δ at some time $t \in (-4\delta/a, 0)$ nd hence there is $t_1 < 0$ at which this happens for the first time (going backward). Hence the segment $\Gamma_{t_1,t_2} = \{\mathbf{z};\ \mathbf{z} = \phi(t, \mathbf{y}), t \in (t_1, t_2)\} \in D$ and also $\Gamma_{t_1,t_2} \cap L \neq \emptyset$. □

Lemma 1.3 *If a trajectory $\Gamma_\mathbf{x}$ intersects a local transversal several times, the successive crossings points move monotonically along the transversal.*

Proof. Consider two successive crossings $\mathbf{y}_1 = \phi(t_1, \mathbf{x})$ and $\mathbf{y}_2 = \phi(t_2, \mathbf{x})$ with, say, $t_1 < t_2$ and a closed curve S composed of the piece Γ' of the trajectory between $\mathbf{y}_1$ and $\mathbf{y}_2$ and the piece L' of the transversal between these two points. Using Jordan's theorem, S divides $\mathbb{R}^2$ into two disjoint open sets with one, say D_1, bounded the other, say D_2, unbounded. We can assume that the flow through L is from D_1 into D_2. Consider $\mathbf{y}_3 = \phi(t_3, \mathbf{x}) \in L$ with $t_3 > t_2$ and first assume $\mathbf{y}_3 \in D_2$. Taking $t' = t_2 + \epsilon$ with ϵ sufficiently small we can be sure that $\phi(t', \mathbf{x})$ is outside D_1. If we assume that $\mathbf{y}_3 \in L'$, then there is ϵ' such that $\phi(t_3 - \epsilon', \mathbf{x}) \in D_1$. Hence, the trajectory between t' and $t_3 - \epsilon'$ joins points of D_1 and D_2. However, it cannot cross Γ' as trajectories cannot cross; also it cannot enter D_1 through L' by its definition. By similar argument, $\mathbf{y}_3$ cannot belong to the sub-segment of L which is inside D_1 (we note that this sub-segment cannot stick outside D_1 as this would require that a point moves along a piece of trajectory in two directions at once). Thus, it must belong to subsegment with $\mathbf{y}_2$ as the end-point. □

An important corollary is:

Corollary 1.1 *If $\mathbf{x} \in \omega(\Gamma_{\mathbf{x}_0})$ is not a stationary point and $\mathbf{x} \in \Gamma_{\mathbf{x}_0}$, then $\Gamma_\mathbf{x} (= \Gamma_{\mathbf{x}_0})$ is a closed curve.*

Proof. Since $\mathbf{x} \in \Gamma_{\mathbf{x}_0}$, $\omega(\Gamma_{\mathbf{x}_0}) = \omega(\Gamma_\mathbf{x})$ (the limit set depends on the trajectory and not on the initial point). Hence, $\mathbf{x} \in \omega(\Gamma_\mathbf{x})$. Chose L to be a local transversal at $\mathbf{x}$. By Lemma 1.2, for sufficiently large T we have an increasing sequence $t_i > t_{i-1} \geq T$ such that $\phi(t_i, \mathbf{x}) \to \mathbf{x}$ as $t_i \to \infty$ and $\phi(t_i, \mathbf{x}) \in L$ (see Lemma 1.2). Also, $\phi(0, \mathbf{x}) = \mathbf{x}$. Suppose $\phi(t_i, \mathbf{x}) \neq \mathbf{x}$, $t_i > T$, then successive intercepts are bounded away from $\mathbf{x}$ which contradicts the fact that $\mathbf{x} \in \omega(\Gamma_\mathbf{x})$. Hence $\phi(t_1, \mathbf{x}) = \mathbf{x}$ for some t and, by Theorem 6.1, the solution is periodic. □

Theorem 1.1 *If an orbit $\Gamma_{\mathbf{x}_0}$ enters and does not leave a closed bounded domain Ω which contains no equilibrium points, then $\omega(\Gamma_{\mathbf{x}_0})$ is a closed orbit.*

Proof. First we prove that $\omega(\Gamma_{\mathbf{x}_0})$ contains a closed orbit. Let $\mathbf{x} \in \omega(\Gamma_{\mathbf{x}_0})$. There are two possibilities.

(i) If $\mathbf{x} \in \Gamma_{\mathbf{x}_0}$, then by Corollary 1.1, $\Gamma_{\mathbf{x}_0}$ is a closed orbit.

(ii) If $\mathbf{x} \notin \Gamma_{\mathbf{x}_0}$, then since $\mathbf{x} \in \omega(\Gamma_{\mathbf{x}_0})$, the orbit $\Gamma_\mathbf{x} \subset \omega(\Gamma_{\mathbf{x}_0})$ by Lemma 5.1 (3) and, because $\omega(\Gamma_{\mathbf{x}_0})$ is closed,

$\omega(\Gamma_{\mathbf{x}}) \subset \omega(\Gamma_{\mathbf{x}_0})$. Let $\mathbf{x}^* \in \omega(\Gamma_{\mathbf{x}}) \subset \omega(\Gamma_{\mathbf{x}_0})$. If $\mathbf{x}^* \in \Gamma_{\mathbf{x}}$ then, again by Corollary 1.1, $\Gamma_{\mathbf{x}} \subset \omega(\Gamma_{\mathbf{x}_0})$ is a closed orbit. This leaves the only possibility that $\mathbf{x}^* \notin \Gamma_{\mathbf{x}}$. Consider a local transversal L at $\mathbf{x}^*$. Arguing as in the proof of Corollary 1.1 we have a sequence $(\mathbf{p}_i)_{i\in\mathbb{N}}$, $\mathbf{p}_i = \phi(t_i, \mathbf{x}) \in L$ with $\mathbf{p}_i \to \mathbf{x}^*$ as $t_i \to \infty$ in a monotonic way. On the other hand, $p_i \in \omega(\Gamma_{\mathbf{x}_0})$ and L is also a local transversal at each p_i. This means that there are sequences on the trajectory $\Gamma_{\mathbf{x}_0}$ converging monotonically to, say, $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$. Assume $\|\mathbf{p}_i - \mathbf{p}_{i+1}\| < \epsilon$ We have, say, $\|\phi(t_1, \mathbf{x}_0) - \mathbf{p}_i\| < \epsilon/4$, then there must be $t_2 > t_1$ such that $\|\phi(t_2, \mathbf{x}_0) - \mathbf{p}_{i+1}\| < \epsilon/4$ but then the next t_3 at which $\Gamma_{\mathbf{x}_0}$ intersects L must be closer than $\epsilon/4$ to both $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$, by Lemma 1.3. This is a contradiction, which shows that $\mathbf{x}^* \in \Gamma_{\mathbf{x}}$ and thus $\omega(\Gamma_{\mathbf{x}_0})$ contains a closed orbit.

The final step of the proof is to show that this orbit, say, Γ is equal to $\omega(\Gamma_{\mathbf{x}_0})$. To do this, we must show that $\phi(t, \mathbf{x}_0) \to \Gamma$ as $t \to \infty$ in the sense of Lemma 5.1(6); that is, that for each $\epsilon > 0$ there is $t > 0$ such that for every $t' > t$ there is $\mathbf{p} \in \Gamma$ (possibly depending on t) which satisfies $\|\phi(t, \mathbf{x}_0) - \mathbf{p}\| < \epsilon$. The argument again uses the properties of a local transversal. Let $\mathbf{z} \in \Gamma \subset \omega(\Gamma_{\mathbf{x}_0})$ and consider a local transversal L at $\mathbf{z}$. Using Lemmas 1.2 and 1.3 we find a sequence $t_0 < t_1 < \ldots t_n \to \infty$ such that $\mathbf{x}_n = \phi(t_k, \mathbf{x}_0) \in L$, $\mathbf{x}_n \to \mathbf{z}$ (in a monotonic way along L). Moreover, we select the sequence $(t_n)_{n\in\mathbb{N}}$ to be subsequent intercepts of $\Gamma_{\mathbf{x}_0}$ with L so that $\phi(t, \mathbf{x}_0) \notin L$ for $t_n < t < t_{n+1}$. Thus, we must estimate what happens to $\phi(t, \mathbf{x}_0)$ for $t \neq t_n$. For any δ there is n_0 such that for $n \geq n_0$ $\|\phi(t_n, \mathbf{x}_0) - \mathbf{z}\| < \delta$. Hence, We by continuity of the flow with respect to the initial condition, for any fixed T an any ϵ there is n_0 such that for any $n \geq n_0$ and $0 \leq t' \leq T$

$$\|\phi(t', \mathbf{x}_n) - \phi(t', \mathbf{z})\| = \|\phi(t' + t_n, \mathbf{x}_0) - \phi(t', \mathbf{z})\| < \epsilon.$$

Now, we observe that if t' changes from 0 to $t_{n+1} - t_n$, the point $\mathbf{x}_n = \phi(t' + t_n, \mathbf{x}_0)$ moves from $\mathbf{x}_n$ to $\mathbf{x}_{n+1}$ which is even closer to $\mathbf{z}$ than $\mathbf{x}_n$ and the trajectory after $\mathbf{x}_{n+1}$ will again stay close to Γ for some finite time T. However, at each cycle the trajectory may wander off when $t' > T$ So, the problem is to determine whether it is possible for $t_{n+1} - t_n$ to become arbitrary large. We know that Γ is closed, hence it is an orbit of a periodic solution, say, $\phi(\lambda, \mathbf{z}) = \mathbf{z}$. Using again continuity of the flow with respect to the initial condition, for any $\mathbf{x}_n$ sufficiently close to $\mathbf{z}$ (that is, for all sufficiently large n), $\phi(\lambda, \mathbf{x}_n)$ will be in some some neigbourhood D_δ of $\mathbf{z}$ (described in Lemma 1.2). But then for all such n, there is $t'_n \in (-\delta, \delta)$ such that $\phi(\lambda + t'_n, \mathbf{x}_0) \in L$. This means that $t_{n+1} - t_n \leq \lambda + \delta$ and the time interval T can be chosen independently of n. Precisely, let us fix $\epsilon > 0$, then there is η such that for all $\|\mathbf{x}_n - \mathbf{z}\| \leq \eta$ and $|t| < \lambda + \delta$ we have

$$\|\phi(t', \mathbf{x}_n) - \phi(t', \mathbf{z})\| = \|\phi(t' + t_n, \mathbf{x}_0) - \phi(t', \mathbf{z})\| < \epsilon.$$

For given η and δ there is n_0 such that for all $n \geq n_0$ we have both $\|\mathbf{x}_n - \mathbf{z}\| < \eta$ and $t_{n+1} - t_n \leq \lambda + \delta$. Hence, taking $t > t_{n_0}$ and selecting n with $t_n \leq t \leq t_{n+1}$ we have, using $t = t' + t_n$ with $0 < t' < \lambda + \delta$

$$\|\phi(t, \mathbf{x}_0) - \phi(t - t_n, \mathbf{z}\| = \|\phi(t - t_n, \mathbf{x}_n) - \phi(t - t_n, \mathbf{z}\| < \epsilon$$

and the proof of the theorem is complete. □

Remark 1.1 The proof of Theorem 1.1 actually shows that a stronger result is valid. Namely, if $\mathbf{x} \in \omega(\Gamma_{\mathbf{x}_0})$ is such that $\omega(\Gamma_{\mathbf{x}})$ contains a non-stationary point, then $\omega(\Gamma_{\mathbf{x}})$ is a closed orbit. This fact will be used in the proof of the following corollary.

Corollary 1.2 *Let $\Gamma_{\mathbf{x}}$ be a bounded trajectory of the system $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ with C^1 planar field for which equilibria are isolated. Then the following three possibilities hold:*

1. *$\omega(\Gamma_{\mathbf{x}})$ is an equilibrium,*
2. *$\omega(\Gamma_{\mathbf{x}})$ is a periodic orbit,*
3. *$\omega(\Gamma_{\mathbf{x}})$ consists of a finite number of equilibria $\mathbf{p}_1, \ldots, \mathbf{p}_k$ connected by trajectories Γ with $\alpha(\Gamma) = \mathbf{p}_i$ and $\omega(\Gamma) = \mathbf{p}_j$.*

Proof. By Lemma 5.1(1), the set $\omega(\Gamma_{\mathbf{x}})$ is bounded an closed and thus can contain only finite number of equilibria. If it contains only equilibria, then it contains only one of them, since $\omega(\Gamma_{\mathbf{x}})$ is connected by

Lemma 5.1(4), which covers the first possibility. If this is not the case, then $\omega(\Gamma_{\mathbf{x}})$ contains non-equilibria and, by Lemma 5.1(3) (invariance), it contains trajectories through these points. Let $\mathbf{u} \in \omega(\Gamma_{\mathbf{x}})$ be an arbitrary non-equilibrium. If $\omega(\Gamma_{\mathbf{u}})$ and $\alpha(\Gamma_{\mathbf{u}})$ are equilibria, then case 3. holds. Assume then that $\omega(\Gamma_{\mathbf{u}})$ contains a non-equilibrium point, say, $\mathbf{z}$. Then, arguing as in the proof of Theorem 1.1 (which is possible as $\mathbf{z}$ is a non-equilibrium, $\Gamma_{\mathbf{u}}$ is a periodic orbit, which means that $\omega(\Gamma_{\mathbf{x}})$ contains a periodic orbit. But then, by the second part of the proof of Theorem 1.1, the whole $\omega(\Gamma_{\mathbf{x}})$ is a periodic orbit. The same argument applies if $\alpha(\Gamma_{\mathbf{u}})$ contains a non-equilibrium point. □

Remark 1.2 Case 3 of the previous corollary can be fine-tuned. Namely, if $\mathbf{p}_1$ and $\mathbf{p}_2$ are two equilibria in $\omega(\Gamma_{\mathbf{x}})$, then there is at most one trajectory $\Gamma \subset \omega(\Gamma_{\mathbf{x}})$ with $\alpha(\Gamma) = \mathbf{p}_1$ and $\omega(\Gamma) = \mathbf{p}_2$. Indeed, assume to the contrary that we have two trajectories Γ_1, Γ_2 with this property and take points $\mathbf{q}_1 \in \Gamma_1$ close to $\mathbf{p}_1$ and $\mathbf{q}_2 \in \Gamma_2$ close to $\mathbf{p}_2$. Since $\mathbf{q}_i$ are not equilibria, there are local transversals L_1 and L_2 at these points. Since Γ_i, $i = 1, 2$ are in $\omega(\Gamma_{\mathbf{x}})$ the trajectory $\Gamma_{\mathbf{x}}$ crosses L_1 in the same direction as Γ_1 and L_2 in the direction of Γ_2. Since the direction of the field along the transversals is the same as of the above trajectories, the region bounded by L_1, the segment of $\Gamma_{\mathbf{x}}$ between L_1 and L_2, L_2, and corresponding segments of Γ_2 and Γ_1, is positively invariant (that is the trajectory $\Gamma_{\mathbf{x}}$ must stay inside). But this is a contradiction as the segments of trajectories Γ_1 and Γ_2 outside this region form a part of the limit set for $\Gamma_{\mathbf{x}}$.

Hence, if $\omega(\Gamma_{\mathbf{x}})$ contains two equilibria, then they must be joined either by a single trajectory, or two trajectories running in opposite directions (and thus forming, together with the equilibria, a loop).

We say that a closed orbit Γ is a *limit cycle* if for some $\mathbf{x} \notin \Gamma$ we have $\Gamma \subset \omega(\Gamma_{\mathbf{x}})$ (*ω-limit cycle*) or $\Gamma \subset \alpha(\Gamma_{\mathbf{x}})$ (*α-limit cycle*).

In the proof of Theorem 1.1 we showed that a limit cycle Γ enjoys the following property: there is $\mathbf{x} \notin \Gamma$ such that

$$\phi(t, \mathbf{x}) \to \Gamma, \qquad t \to \infty.$$

Geometrically it means that some trajectory spirals towards Γ as $t \to \infty$ (or $t \to -\infty$). It follows that limit cycles have certain (one-sided) stability property.

Proposition 1.1 *Let Γ be an ω-limit cycle such that $\Gamma \subset \omega(\Gamma_{\mathbf{x}})$, $\mathbf{x} \notin \Gamma$. Then there is a neighourhood V of $\mathbf{x}$ such that for any $\mathbf{y} \in V$ we have $\Gamma \in \omega(\Gamma_{\mathbf{y}})$.*

Proof. Let Γ be an ω-limit cycle and let $\phi(t, \mathbf{x})$ spirals toward Γ as $t \to \infty$. Take $\mathbf{z} \in \Gamma$ and a transversal L at $\mathbf{z}$. Take $t_0 < t_1$ such that $\mathbf{x}_0 = \phi(t_0, \mathbf{x}), \mathbf{x}_1 = \phi(t_1, \mathbf{x}) \in L$ with $\phi(t, \mathbf{x}) \notin L$ for $t_0 < t < t_1$. For sufficiently large t_0 the segment $L_{\mathbf{x}_0, \mathbf{x}_1}$ of L between $\mathbf{x}_0$ and $\mathbf{x}_1$ does not intersect Γ. Then the region A bounded by Γ, the part of $\Gamma_{\mathbf{x}}$ between $\mathbf{x}_0$ and $\mathbf{x}_1$ and $L_{\mathbf{x}_0, \mathbf{x}_1}$ is forward invariant, as is the set $B = A \setminus \Gamma$. For sufficiently large $t > 0$, $\phi(t, \mathbf{x})$ is in the interior of A. But than, the same is true for $\phi(t, \mathbf{y})$ for $\mathbf{y}$ sufficiently close to $\mathbf{x}$. Such $\phi(t, \mathbf{y})$ stays in A and by the Poincareé-Bendixon theorem, must spiral towards Γ. □

Corollary 1.3 *Let Γ be a closed orbit enclosing an open set Ω (contained in the domain of $\mathbf{f}$). Then Ω contains an equilibrium.*

Proof. Assume Ω contains no equilibrium. Our first step is to show that there must be the 'smallest' closed orbit. We know tat orbits don't intersect so if we have a collection of closed orbits Γ_n, then the regions Ω_n enclosed by them form a nested sequence with decreasing areas A_n. We also note that if we have a sequence $\mathbf{x}_n$ of points on Γ_n converging to $\mathbf{x} \in \Omega$, then $\mathbf{x}$ also is on a closed orbit. Otherwise, by Poincaré-Bendixon theorem, $\phi(t, \mathbf{x})$ would spiral towards a limit cycle but then this would be true for $\phi(t, \mathbf{x}_n)$ for sufficiently large n, by virtue of the previous result, which contradict the assumption that $\mathbf{x}_n$ is on a closed orbit.

Let $0 \leq A$ be an infimum of all areas of regions enclosed by closed orbits. Then $A = \lim_{n \to \infty} A_n$ for some sequence of Γ_n. Let $\mathbf{x}_n \in \Gamma_n$; since $\Omega \cup \Gamma$ is compact, we may assume $\mathbf{x}_n \to \mathbf{x}$ and, by the previous part, $\mathbf{x}$ belongs to a closed orbit Γ_0. By using the standard argument with transversal at $\mathbf{x}$ we find that Γ_n get arbitrarily close to Γ_0 so A is the area enclosed by Γ_0. Since Γ_0 does not reduce to a point, $A > 0$. But

then the region enclosed by Γ_0 contains neither equilibria nor closed orbit, which contradicts the Poincaré-Bendixon theorem. □

A difficult part in applying the Poincareé-Bendixon theorem is to find the *trapping region*; that is, the closed and bounded region which contains no equilibria and which trajectories cannot leave. Quite often this will be an annular region with the equilibrium (source) in the hole, so that the trajectories can enter through the inner boundary and the outer boundary chosen in such a way that the vector field there always points inside the region. We illustrate this in the following example:

Example 1.1 Show that the second order equation

$$z'' + (z^2 + 2(z')^2 - 1)z' + z = 0$$

has a nontrivial periodic solution. We start with writing this equation as the system

$$\begin{aligned} x' &= y, \\ y' &= -x + y(x^2 + 2y^2 - 1) \end{aligned}$$

Clearly, $(0, 0)$ is an equilibrium and by linearization we see that this is an unstable equilibrium hence there is a chance that the flow in a small neighbourhood of the origin will be outward. Thus, let us write the equation in polar coordinates. We get

$$\frac{d}{dt}(x^2 + y^2) = 2xx' + 2yy' = 2y^2(1 - x^2 - 2y^2).$$

We observe that $1 - x^2 - 2y^2 > 0$ for $x^2 + y^2 < 1/2$ and $1 - x^2 - 2y^2 < 0$ for $x^2 + y^2 > 1$. Hence, any solution which starts in the annulus $1/2 < x^2 + y^2 < 1$ must stay there. Since the annulus does not contain an equilibrium, there must be a periodic orbit inside it.

Let us consider a more sophisticated example.

Example 1.2 Glycolysis is a fundamental biochemical process in which living cells obtain energy by breaking down sugar. In many cells, such as yeast cells, glycolysis can proceed in an oscillatory fashion.

A simple model of glycolysis presented by Sel'kov reads

$$\begin{aligned} x' &= -x + ay + x^2 y, \\ y' &= b - ay - x^2 y, \end{aligned} \tag{4.1.2}$$

where x and y are, respectively, concentrations of ADP (adenosine diposphate) and F6P (fructose-6-phosphate), and $a, b > 0$ are kinetic parameters. We shall show that under certain assumptions there are, indeed, periodic solutions to the system.

First we find *nullclines* (that is, curves along which one or the other time derivative is zero). Note that an equilibria are the intersection points of nullclines. We obtain that $x' = 0$ along the curve $y = x/(a + x^2)$ and $y' = 0$ along $y = b/(a + x^2)$. The nullclines are sketched in Figure 1, along with some representative vectors.

Fig 1. Isoclines of (4.1.2).

Our first step is to construct a trapping region. First we observe that the trajectory cannot escape across any coordinate axis and also, since $y' < 0$ above both nullclines, no trajectory can cross a horizontal line in this region going upward. This leaves only the question of closing this region from the right. We note that $x' + y' = b - x < 0$ as long as $x > b$. Thus, in this region $y' < -x'$ and, since $x' > 0$, we have $y'/x' < -1$ which geometrically means that the slope is steeper than any straight line with slope -1. In other words, any trajectory in this region will cross a line with the slope -1 from above-right to below-left. Hence, if we draw such a line crossing the horizontal line $y = b/a$ at $x = b$, then trajectories from the left will not be able to escape through it. Finally, we continue the line till it intersects with the nullcline $y = x/(a+x^2)$ and note that below this nullcline $x' < 0$ so that we can close the region by a vertical segment to obtain a trapping region, see Fig 2:

Fig 2. Trapping region for (4.1.2).

Can we conclude that there is a periodic trajectory? Well, no as there is an equilibrium $(b, b/(a+b^2))$ inside the region. However, this is not necessarily a bad news. If the equilibrium is a repeller (real parts of all eigenvalues are negative), then there is a neighbourhood V of the equilibrium such that any trajectory starting from V will eventually reach a sphere of a fixed radius (see the first part of Example 5.11). Thus, the boundary of V can be taken as the inner boundary of the required set. So, we need to prove that $(b, b/(a+b^2))$ is a repeller. The Jacobi matrix at for (4.1.2) at (x_0, y_0) is

$$J = \begin{pmatrix} -1+2x_0y_0 & a+x_0 \\ -2x_0y_0 & -(a+x_0^2) \end{pmatrix}.$$

The Jacobian at the equilibrium is $a+b^2$ and the trace τ is

$$\tau = -\frac{b^2+(2a-1)b^2+(a+a^2)}{a+b^2}.$$

Hence, the equilibrium is unstable $\tau > 0$, and stable $\tau > 0$.

Fig 3. Repelling character of equilibrium (4.1.2).

The dividing line $\tau = 0$ occurs when

$$b^2 = \frac{1}{2}(1-2a \pm \sqrt{1-4a})$$

and this curve is shown in Fig. 4.

Fig 4. Bifurcation curve for (4.1.2).

For parameters in the region corresponding to $\tau > 0$ we are sure that there is a closed orbit of the system.

Fig 5. Closed orbit for (4.1.2).

2 Other criteria for existence and non-existence of periodic orbit

When we were discussing the Liapunov function, we noted the cases such that if V was constant on tracjectories, then the trajectories are closed so that the solutions are periodic. It turns out that this is a general

situation for *conservative systems*. To be more precise, let us consider the system

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}). \tag{4.2.3}$$

We say that there is a *conservative quantity* for this system if there exists a scalar continuous function $E(\mathbf{x})$ which is constant along trajectories; that is $\frac{d}{dt}E(\phi(t,\mathbf{x}_0)) = 0$. To avoid trivial examples, we also require $E(\mathbf{x})$ to be nonconstant on every open set. We have the following result.

Proposition 2.1 *Assume that there is a conserved quantity of the system (4.2.3), where $\mathbf{y} \in \mathbb{R}^2$, and that $\mathbf{y}^*$ is an isolated fixed point of (4.2.3). If $\mathbf{y}^*$ is alocal minimum of E, then there is a neighbourhood of $\mathbf{y}^*$ in which all trajectories are closed.*

Proof. Let $B_\delta = B(\mathbf{y}^*, \delta)$ be a closed neighbourhood of $\mathbf{y}^*$ in which $\mathbf{y}^*$ is absolute minimum. We can assume $E(\mathbf{y}^*) = 0$. Since E is constant on trajectories, each trajectory is contained in a level set of E:

$$E_c = \{(y_1, y_2);\ E(y_1, y_2) = c\} \cap B_\delta.$$

E_c is a nonempty bounded closed set for all sufficiently small $0 < c \le c_0$. We use Theorem 1.1. For each c there are two possibilities: either the trajectory stays in E_c in which case also its ω-limit set is in E_c and thus it is a closed orbit or the trajectory leaves E_c so that there is a point $\mathbf{y}$ on the trajectory with $E(\mathbf{y}) = c$ and $\mathbf{y} = \delta$.

If the second case happens for some $c_n \to 0$, then we have a sequence $\mathbf{y}_n$ with $\|\mathbf{y}_n\| = \delta$ for which $E(\mathbf{y}_n)c_n \to 0$. But then, by continuity of E we find a point $\mathbf{y}_0$ with $E(\mathbf{y}_0) = 0$, contradicting the assumption that $\mathbf{y}^*$ is an absolute minimum in B_δ. □

Another case is concerned with 'time reversible' systems. We say that the system

$$\begin{aligned} y_1' &= f_1(y_1, y_2), \\ y_2' &= f_2(y_1, y_2), \end{aligned} \tag{4.2.4}$$

is time reversible, if it is invariant under $t \to -t$ and $y \to -y$. If f is odd in y_2 and f_2 is even in y_2, then such system is 'time-reversible'.

Proposition 2.2 *Suppose $\mathbf{y}^* = (0,0)$ is the center for the linearization of a reversible system (4.2.4).Then sufficiently close to the origin all trajectories are closed curves.*

Proof. Change coordinates we can write (4.2.4) as

$$\begin{aligned} x_1' &= -x_2 + \omega_1(x_1, x_2), \\ x_2' &= x_1 + \omega_2(x_1, x_2), \end{aligned} \tag{4.2.5}$$

Let A denotes the matrix of the linearization. Then

$$\mathbf{x}(t) = e^{tA}\mathbf{x}_0 + e^{tA}\int_0^t e^{-sA}\omega(\mathbf{x}(s))ds,$$

where $\omega = (\omega_1, \omega_2)$. The trajectories of the linear part are circles traversed anticlockwise and so $\|e^{tA}\mathbf{x}_0\| = \|\mathbf{x}_0\|$ for any t. Hence

$$\|\mathbf{x}(t)\| \le \|\mathbf{x}_0\| + \int_0^t \|\omega(\mathbf{x}(s))\|ds,$$

Our aim is to show that the trajectory of the nonlinear system follows closely the trajectory of the linear system and thus starting from $\mathbf{x}_0$ on the positive horizontal semiaxis, it will cross the negative positive semiaxis. First, let us fix $\xi > 1$ satisfying $\ln \xi < 1/\xi$. Next, from the properties of linearization, for any ϵ we

find δ such that $\|\omega(\mathbf{x})\| \le \epsilon\|\mathbf{x}\|$ as long as $\|\mathbf{x}\| \le \delta$. Next, select $\mathbf{x}_0$ with $\|\mathbf{x}_0\| = \delta/\xi$. Then, by the Gronwall inequality

$$\|\mathbf{x}(t)\| \le \|\mathbf{x}_0\|e^{\epsilon t} \le \delta$$

as long as $t \le \ln\xi/\epsilon > 0$. Next, calculate the norm of the difference $\mathbf{z}(t)$ between the nonlinear and linear solution starting from $\mathbf{x}_0$ over the time interval $[0, \ln\xi/\epsilon]$. We obtain

$$\|z(t)\| \le \int_0^t \|\omega(\mathbf{y}(s)\|ds \le \epsilon t\delta \le \delta\ln\xi < \delta/\xi < \delta.$$

Hence, $\mathbf{x}(t)$ stays in the annulus surrounding the circle of radius δ/ϵ with the inner radius $\delta(1/\xi - \ln\xi)$. In particular, taking ϵ sufficiently small, we can take $t = 3\pi/2$ in which case $\mathbf{x}(t)$ must be necessarily below the horizontal axis. Thus, there mast be a point $t' > 0$ with $x_1(t') < 0$ and $x_2(t') = 0$. Now, trajectories are invariant with respect to the change $t \to -t$ and $y \to -y$ and hence we can complete the constructed trajectory to a closed one.

Since δ can be taken arbitrarily small, we see that all trajectories close to $(0,0)$ are closed. □

Sometimes it is equally important to show that there are no periodic orbits in certain regions. We have already seen one such criterion related to Lyapunov functions: if there is a strict Lyapunov function in certain region surrounding an equilibrium, then there are no periodic orbits in this region. Here we consider two more 'negative' criteria.

Consider a system $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ where $\mathbf{f} = -\nabla V$ in some region Ω for a scalar function V. Such systems are called *gradient systems* with *potential function* V. We have

Proposition 2.3 *Closed orbits are impossible in gradient systems with $V \neq const$.*

Proof. Suppose that there is a closed orbit Γ corresponding to a periodic solution with period T. Define ΔV to be the change of V along Γ. Clearly, $\Delta V = 0$ as the the orbit is closed and V is continuous. On the other hand

$$\Delta V = \int_0^T \frac{dV}{dt}dt = \int_0^T \nabla V \cdot \mathbf{y}'dt = -\int_0^T \|\mathbf{y}'\|^2 dt < 0$$

as the trajectory is not an equilibrium due to $V \neq const$. This contradiction proves the statement. □

To illustrate this result, consider the system

$$\begin{aligned} y_1' &= \sin y_2, \\ y_2' &= y_1 \cos y_2. \end{aligned}$$

This is a gradient system with $V(y_1, y_2) = -y_1 \sin y_2$, so there are no periodic solutions.

Note. The above criterion works in arbitrary dimension.

Next we consider the so-called *Dulac criterion.*

Proposition 2.4 *Consider a planar system $\mathbf{y}' = \mathbf{f}(\mathbf{y})$ with continuously differentiable $\mathbf{f}$. Suppose that there is a scalar differentiable function $g : \mathbb{R}^2 \to \mathbb{R}$ such that $\mathrm{div}(g\mathbf{f}) \neq 0$ on some simply connected domain Ω. Then there are no periodic orbits of this system which lie in Ω.*

Proof. Suppose $\Gamma \subset \Omega$ is a periodic orbit. Using Green's theorem and the assumption

$$0 \neq \int\int_A \mathrm{div}(g\mathbf{f})dx_1dx_2 = \oint_\Gamma g\mathbf{f}\mathbf{n}d\sigma$$

where $\mathbf{n}$ is the normal to Γ and $d\sigma$ the line element along Γ. However, since $\mathbf{n}$ is normal to Γ, we have $\mathbf{n} \cdot \mathbf{f} = 0$ as $\mathbf{f}$ is tangent to any trajectory. Thus the left hand side is zero and we obtain a contradiction. □

We can illustrate this criterion on a variant of Lotka-Volterra model

$$\begin{aligned} y_1' &= y_1(A - a_1 y_1 + b_1 y_2), \\ y_2' &= y_2(B - a_2 y_2 + b_2 y_1). \end{aligned}$$

with $a_i > 0$ (to model the effect of overcrowding. Using the function $g = 1/y_1 y_2$ we can prove that there are no periodic orbits in the first quadrant.

Chapter 5

Comments on the Stable Manifold Theorem

In Theorem 5.4 (ii) we stated that if at least one eigenvalue of the linearized system has positive real part, then the corresponding nonlinear system is unstable. In this subsection we shall concentrate on planar systems though the main theorem can be proved in an arbitrary dimension; the proof is practically the same as in two dimensions provided we have sufficient background in differential geometry to operate with manifolds rather than with curves.

In two dimensions the only possible situation described in Theorem 5.4 (ii) is that $\lambda_1 < 0$ and $\lambda_2 > 0$ (recall that the case of zero eigenvalue has been ruled out by the non-degeneracy assumption on the matrix $\mathcal{A}$. To simplify notation let us denote $\lambda_1 = -\mu$, $\lambda_2 = \lambda$, $\mu, \lambda > 0$. For the linear system this case corresponds to the origin $\mathbf{0}$ being a saddle, that is, solutions tend to zero as $t \to \pm\infty$ if and only if the initial conditions are on one of the two half-lines determined by the eigenvectors corresponding to the negative (resp. positive) eigenvalue. For any other initial conditions the solutions become unbounded for both $t \to \pm\infty$.

Our aim here is to show that for a nonlinear system for which the origin is a saddle of its linearization, the phase portrait close to the origin is similar to a saddle in the sense that there exist two curves defined in some neighbourhood of $\mathbf{0}$ such that the solutions with initial conditions at $t = 0$ on one of these curves will remain on it for all $t \geq 0$ and with the initial condition on the other will be on it for all $t \leq 0$ and in both cases the solution will converge to $\mathbf{0}$ as $t \to \infty$ (resp. $t \to -\infty$). For any other initial condition, however small but non-zero, from this neighbourhood, the solution will become unbounded for both $t \to \pm\infty$. Let us first illustrate this idea on an example.

Example 0.1 Consider

$$\begin{aligned} x' &= -x, \\ y' &= y + x^2 \end{aligned}$$

Solving the first equation and inserting the solution to the second we obtain the solution in the form

$$\begin{aligned} x(t) &= x_0 e^{-t}, \\ y(t) &= \left(y_0 + \frac{x_0^2}{3}\right) e^t - \frac{x_0^2}{3} e^{-2t}. \end{aligned}$$

From this solution we clearly see that only the initial conditions satisfying

$$y_0 = -\frac{x_0^2}{3}$$

yield the solution tending to zero as $t \to \infty$. Any solution of this type will have the form

$$\begin{aligned} x(t) &= x_0 e^{-t}, \\ y(t) &= -\frac{x_0^2}{3} e^{-2t}. \end{aligned}$$

so that the orbit will be on the parabola $y = -\frac{1}{3}x^2$ (this parabola will consists of three orbits: the right-hand branch for $x_0 > 0$, the left-hand branch for $x_0 < 0$ and the stationary point $\mathbf{0}$. Thus, as in the linear case, the curve consisting of the initial conditions giving solutions converging to zero consists itself of solutions.

On the other hand, only the initial conditions $x_0 = 0$ with arbitrary y_0 produce solutions tending to zero as $t \to -\infty$ and of course, again, this curve (straight line) consists of orbits of the solutions.

Example 0.2 To prepare the ground for the proof of the main result, let us consider the following system

$$\begin{aligned} x' &= -\mu x + f(t), \\ y' &= \lambda y + g(t) \end{aligned}$$

where f and g are known bounded continuous functions. The question is: can we find initial condition for this system so that the solution will be bounded. We can immediately find the general solution to this system:

$$\begin{aligned} x(t) &= e^{-\mu t} x_0 + e^{-\mu t} \int_0^t e^{\mu s} f(s) ds, \\ y(t) &= e^{\lambda t} y_0 + e^{\lambda t} \int_0^t e^{-\lambda s} g(s) ds, \end{aligned}$$

with $x(0) = x_0, y(0) = y_0$. Clearly, $x(t)$ is bounded if $f(t)$ is bounded as $t \to \infty$ (why?) but $y(t)$ in general is unbounded. Let us then write $y(t)$ in the following form

$$y(t) = e^{\lambda t} y_0 + e^{\lambda t} \int_0^\infty e^{-\lambda s} g(s) ds - e^{\lambda t} \int_t^\infty e^{-\lambda s} g(s) ds.$$

The last term is well-defined as

$$\left| \int_t^\infty e^{-\lambda s} g(s) ds \right| \leq \max |g(t)| \lambda^{-1} e^{-\lambda t}$$

and bounded as $t \to \infty$. Thus, if $y_0 = -\int_0^\infty e^{-\lambda s} g(s) ds$, then the first term will vanish and the solution $y(t)$ will be also bounded. Hence, we found that the initial conditions lying on the straight line $(x_0, -\int_0^\infty e^{-\lambda s} g(s) ds)$ are the only initial conditions producing bounded solutions that can be written in the form

$$\begin{aligned} x(t) &= e^{-\mu t} x_0 + e^{-\mu t} \int_0^t e^{\mu s} f(s) ds, \\ y(t) &= -e^{\lambda t} \int_t^\infty e^{-\lambda s} g(s) ds. \end{aligned}$$

After these preliminaries, we are ready to formulate the theorem for nonlinear systems. As before, we consider the system in the form

$$\mathbf{x}' = \mathcal{A}\mathbf{x} + \mathbf{g}(\mathbf{x}), \tag{5.0.1}$$

where we assume that $-\mu$ and λ, $\mu, \lambda > 0$, are eigenvalues of the matrix $\mathcal{A}$. We can assume then that

$$\mathcal{A} = \begin{pmatrix} -\mu & 0 \\ 0 & \lambda \end{pmatrix},$$

otherwise any matrix with two distinct real eigenvalues can be reduced to such a form by a linear change of variables called diagonalization and a linear change of variables does not alter the properties of $\mathbf{g}$ that are relevant in the proof below.

Theorem 0.1 *Let* $\mathbf{g}$ *be continuously differentiable for* $\|\mathbf{x}\| < k$ *for some constant* $k > 0$*, with* $\mathbf{g}(\mathbf{0}) = \mathbf{0}$ *and*

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial \mathbf{g}}{\partial x_j}(\mathbf{x}) = \mathbf{0}.$$

If the eigenvalues of $\mathcal{A}$ *are* $\lambda, -\mu$ *with* $\lambda, \mu > 0$*, then there exists in the space* $\mathbf{x}$ *a curve* C*:* $x_2 = \phi(x_1)$*, passing through the origin such that if* $\mathbf{x}(t)$ *is any solution with* $\mathbf{x}(0)$ *on* C *and* $\|\mathbf{x}(0)\|$ *sufficiently small, then* $\mathbf{x}(t) \to \mathbf{0}$ *as* $t \to \infty$*. Similarly, there is a curve* C' *such that if* $\mathbf{x}(t)$ *is any solution with* $\mathbf{x}(0)$ *on* C' *and* $\|\mathbf{x}(0)\|$ *sufficiently small, then* $\mathbf{x}(t) \to \mathbf{0}$ *as* $t \to -\infty$*. Moreover, no solution* $\mathbf{x}(t)$ *with* $\mathbf{x}(0)$ *small enough, but not on* C *(resp.* C'*) can remain bounded as* $t \to \infty$ *(resp.* $t \to -\infty$*).*

Proof. Let us denote

$$\mathcal{U}(t) = \begin{pmatrix} -\mu & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathcal{V}(t) = \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix} \tag{5.0.2}$$

so that

$$e^{t\mathcal{A}} = \mathcal{U}(t) + \mathcal{V}(t).$$

Then, for any $\mathbf{a} \in \mathbb{R}^2$

$$\|\mathcal{U}(t)\mathbf{a}\| \le e^{-\mu t}\|\mathbf{a}\|, \qquad \|\mathcal{V}(t)\mathbf{a}\| \le e^{\lambda t}\|\mathbf{a}\|.$$

In the proof it will be convenient to write these estimates as

$$\begin{aligned} \|\mathcal{U}(t)\mathbf{a}\| &\le e^{-(\alpha+\sigma)t}\|\mathbf{a}\|, \quad t \ge 0 \\ \|\mathcal{V}(t)\mathbf{a}\| &\le e^{\sigma t}\|\mathbf{a}\|, \quad t \le 0, \end{aligned}$$

where $\alpha, \sigma > 0$ are chosen so that $\sigma \le \lambda$ and $\sigma + \alpha \le \mu$.

The next preliminary step is the observation that due to the assumptions on $\mathbf{g}$ (smallness of the partial derivatives in a neighbourhood of $\mathbf{0}$), for any $\epsilon > 0$ there is $\delta > 0$ such that if $\|\mathbf{x}\|, \|\mathbf{x}^*\| < \delta$, then

$$\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{x}^*)\| \le \epsilon\|\mathbf{y} - \mathbf{y}^*\|.$$

Passing to the main part of the proof, we use the considerations of Example 0.2 and will consider, for $\mathbf{a} = (a_1, 0)$ the integral equation

$$\mathbf{x}(t, \mathbf{a}) = \mathcal{U}(t)\mathbf{a} + \int_0^t \mathcal{U}(t-s)\mathbf{g}(\mathbf{x}(s, \mathbf{a}))ds - \int_t^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{x}(s, \mathbf{a}))ds, \tag{5.0.3}$$

or in the expanded form

$$\begin{aligned} x_1(t, a_1) &= e^{-\mu t}a_1 + e^{-\mu t}\int_0^t e^{\mu s} g_1(x_1(s, a_1), x_2(s, a_1))ds, \\ x_2(t, a_1) &= -e^{\lambda t}\int_t^\infty e^{-\lambda s} g_2(x_1(s, a_1), x_2(s, a_1))ds. \end{aligned} \tag{5.0.4}$$

Let us first observe that, as in Example 0.2, any bounded continuous solution of (5.0.3), (5.0.4) is a solution to the original system (5.0.1) satisfying the initial condition $x_1(0) = a_1$ and $x_2(0) = -\int_0^\infty e^{-\lambda s} g_2(x_1(s,a_1), x_2(s,a_1))ds$. If we recall that $\mathbf{x}(t)$ depends on a_1 in a continuous way we see that in fact $x_2(0) = \phi(a_1)$ so that there is a curve $x_2(0) = \phi(x_1(0))$ such that the solutions starting from points of this curve remain bounded. The problem is to show that the integral equation has solutions that are bounded. We shall show it by successive approximations. Let us define, for a moment formally,

$$\begin{aligned}\mathbf{x}^0(t,\mathbf{a}) &= \mathbf{0},\\ \mathbf{x}^n(t,\mathbf{a}) &= \mathcal{U}(t)\mathbf{a} + \int_0^t \mathcal{U}(t-s)\mathbf{g}(\mathbf{x}^{n-1}(s,\mathbf{a}))ds - \int_t^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{x}^{n-1}(s,\mathbf{a}))ds.\end{aligned}$$

Since by the assumption $\mathbf{g}(\mathbf{0}) = \mathbf{0}$, we have

$$\|\mathbf{x}^1(t,\mathbf{a}) - \mathbf{x}^0(t,\mathbf{a})\| = \|\mathcal{U}(t)\mathbf{a}\| \le e^{-(\alpha+\sigma)t}\|\mathbf{a}\| \le e^{-\alpha t}\|\mathbf{a}\|.$$

Clearly, if $\|\mathbf{a}\| < k$, then $\|\mathbf{x}^1(t,\mathbf{a})\| < k$ for all times and we can consider the second iterate

$$\mathbf{x}^2(t,\mathbf{a}) = \mathcal{U}(t)\mathbf{a} + \int_0^t \mathcal{U}(t-s)\mathbf{g}(\mathbf{x}^1(s,\mathbf{a}))ds - \int_t^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{x}^1(s,\mathbf{a}))ds.$$

and estimate, provided $\|\mathbf{x}^1(t,\mathbf{a}))\| < \delta$,

$$\begin{aligned}\|\mathbf{x}^2(t,\mathbf{a}) - \mathbf{x}^1(t,\mathbf{a})\| &\le \int_0^t \|\mathcal{U}(t-s)\|\|\mathbf{g}(\mathbf{x}^1(t,\mathbf{a})) - \mathbf{g}(\mathbf{x}^0(t,\mathbf{a}))\|ds\\ &\quad + \int_t^\infty \|\mathcal{V}(t-s)\|\|\mathbf{g}(\mathbf{x}^1(s,\mathbf{a})) - \mathbf{g}(\mathbf{x}^0(s,\mathbf{a}))\|ds\\ \le \epsilon\|\mathbf{a}\| \int_0^t e^{-(\sigma+\alpha)(t-s)}e^{-\alpha s}ds &+ \epsilon\|\mathbf{a}\| \int_t^\infty e^{\sigma(t-s)}e^{-\alpha s}ds\\ = \epsilon\|\mathbf{a}\| \left(\frac{e^{-(\sigma+\alpha)t}}{\sigma}(e^{\sigma t}-1)\right. &+ \left. e^{\sigma t}\frac{e^{-(\sigma+\alpha)t}}{\alpha+\sigma}\right)\\ &\le \frac{2\epsilon\|\mathbf{a}\|}{\sigma}e^{-\alpha t}\end{aligned}$$

where we used $e^{\sigma t} - 1 \le e^{\sigma t}$ and $1/(\alpha+\sigma) \le 1/\sigma$. Let us assume that $\epsilon/\sigma < 1/4$, then

$$\|\mathbf{x}^2(t,\mathbf{a}) - \mathbf{x}^1(t,\mathbf{a})\| \le \frac{\|\mathbf{a}\|}{2}e^{-\alpha t}.$$

Let us conjecture the induction assumption

$$\|\mathbf{x}^j(t,\mathbf{a}) - \mathbf{x}^{j-1}(t,\mathbf{a})\| \le \frac{\|\mathbf{a}\|}{2^{j-1}}e^{-\alpha t}, \tag{5.0.5}$$

for $j \le n$. Note, that if this assumption is satisfied for all $j \le n$, then

$$\|\mathbf{x}^n(t,\mathbf{a})\| \le \|\mathbf{x}^n(t,\mathbf{a}) - \mathbf{x}^{n-1}(t,\mathbf{a})\| + \ldots + \|\mathbf{x}^1(t,\mathbf{a})\| = \|a\|e^{-\alpha t}(2^{-(n-1)} + \ldots + 1) \le 2\|a\|e^{-\alpha t}$$

so that if $\|\mathbf{a}\| < \delta/2$, then $\|\mathbf{x}^n(t,\mathbf{a})\| < \delta$ and the estimates for iterates can be carried for the next step. Thus, we have

$$\|\mathbf{x}^{n+1}(t,\mathbf{a}) - \mathbf{x}^n(t,\mathbf{a})\| \le \int_0^t \|\mathcal{U}(t-s)\|\|\mathbf{g}(\mathbf{x}^n(s,\mathbf{a})) - \mathbf{g}(\mathbf{x}^{n-1}(s,\mathbf{a}))\|ds$$

$$+\int_t^\infty \|\mathcal{V}(t-s)\|\|\mathbf{g}(\mathbf{x}^n(s,\mathbf{a}))-\mathbf{g}(\mathbf{x}^{n-1}(s,\mathbf{a}))\|ds$$

$$\begin{aligned}\leq \epsilon\|\mathbf{a}\|\int_0^t e^{-(\sigma+\alpha)(t-s)}2^{-(n-1)}e^{-\alpha s}ds &+ \epsilon\|\mathbf{a}\|\int_t^\infty e^{\sigma(t-s)}2^{-(n-1)}e^{-\alpha s}ds\\ = \frac{\epsilon\|\mathbf{a}\|}{2^{n-1}}\left(\frac{e^{-(\sigma+\alpha)t}}{\sigma}(e^{\sigma t}-1)\right. &+ \left. e^{\sigma t}\frac{e^{-(\sigma+\alpha)t}}{\alpha+\sigma}\right)\\ \leq \frac{\epsilon\|\mathbf{a}\|}{2^{n-2}\sigma}e^{-\alpha t} &\leq \frac{\|\mathbf{a}\|}{2^n}e^{-\alpha t}\end{aligned}$$

where we used $\epsilon/\sigma < 1/4$ in the last step.

Therefore, as in the Picard theorem, the sequence $\mathbf{x}^n(t,\mathbf{a})$ converges uniformly on $[0,\infty)$ to some continuous function $\mathbf{x}(t,\mathbf{a})$ satisfying

$$\|\mathbf{x}(t,\mathbf{a})\| \leq 2\|\mathbf{a}\|e^{-\alpha t}$$

for $t \geq 0$ and $\|\mathbf{a}\| < \delta/2$. As in Picard's theorem we find that this $\mathbf{x}(t,\mathbf{a})$ is the solution of the integral equation (5.0.4) and therefore it is also a unique solution of the differential equation (5.0.1) satisfying the initial condition $x_1(0) = a_1, x_2(0) = -\int_0^\infty e^{-\lambda s}g_2(x_1(s,a_1),x_2(s,a_1))ds$ so that we found a curve C, defined in some neighbourhood of the origin, with the property that any solution emanating from C tends to zero as $t \to \infty$.

Let us consider a solution $\mathbf{y}(t)$ of (5.0.1) with $\mathbf{y}(0) = \mathbf{b}$, where $\mathbf{b}$ is small but not on C. Assume that $\|\mathbf{y}(t)\| \leq \delta$, where δ is defined as above. The solution satisfies

$$\mathbf{y}(t) = e^{t\mathcal{A}}\mathbf{b} + \int_0^t e^{(t-s)\mathbf{A}}\mathbf{g}(\mathbf{y}(s))ds,$$

and, as before, we write this solution as

$$\mathbf{y}(t) = \mathcal{U}(t)\mathbf{b} + \mathcal{V}(t)\mathbf{b} + \int_0^t \mathcal{U}(t-s)\mathbf{g}(\mathbf{y}(s))ds + \int_0^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{y}(s))ds - \int_t^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{y}(s))ds \tag{5.0.6}$$

where all the integrals exist due to the bound on $\mathcal{V}(t)$ and since $\mathbf{g}(\mathbf{y}(s))$ is bounded whenever $\|\mathbf{y}(s)\| \leq \delta$. Since $\mathcal{V}(t)$ is just a multiplication of the second coordinate by $e^{\lambda t}$, we can rewrite the above as

$$\mathbf{y}(t) = \mathcal{U}(t)\mathbf{b} + \mathcal{V}(t)\mathbf{c} + \int_0^t \mathcal{U}(t-s)\mathbf{g}(\mathbf{y}(s))ds - \int_t^\infty \mathcal{V}(t-s)\mathbf{g}(\mathbf{y}(s))ds$$

where

$$\mathbf{c} = \mathbf{b} + \int_0^\infty \mathcal{V}(-s)\mathbf{g}(\mathbf{y}(s))ds.$$

Since $\|\mathbf{y}(t)\| \leq \delta$ and all the terms on the right-hand side, except possibly $\mathcal{V}(t)\mathbf{c}$, are clearly bounded. Thus, $\mathcal{V}(t)\mathbf{c}$ must be also bounded but $\mathcal{V}(t)\mathbf{c} = (0, e^{\lambda t c_2})$, where c_2 is the second component of $\mathbf{c}$. Hence $c_2 = 0$ but this implies $\mathcal{V}(t)\mathbf{c} = 0$ for all $t \geq 0$ so that $\mathbf{y}(0)$ is on the curve C, contrary to the assumption.

This result allows to establish that C consists of orbits. In fact, let $\mathbf{x}(t)$ be a solution emanating from C, converging to $\mathbf{0}$. If it does not stay on C, then for some time t_*, $\mathbf{x}(t_*) \notin C$. Considering $\hat{\mathbf{x}}(t) = \mathbf{x}(t+t_*)$, we see that $\hat{\mathbf{x}}(t)$ is a solution with initial condition not on C that converges to $\mathbf{0}$ which is a contradiction with the previous part of the proof.

The statement concerning the curve C' is obtained by changing the direction of time in (5.0.1) and noting that the properties of the system relevant to the proof do not change hance C for the system with reversed time becomes C' for the original system. ∎

Chapter 6

Origins of partial differential equations

1 Basic facts from Calculus

In the course we shall frequently need several facts from the integration theory. We list them here as theorems, though we shall not prove them during the lecture. Some easier proofs should be done as exercises.

Theorem 1.1 *Let f be a continuous function in $\bar{\Omega}$, where $\Omega \subset \mathbb{R}^d$ is a bounded domain. Assume that $\forall_{\boldsymbol{x}\in\bar{\Omega}}\ f(\boldsymbol{x}) \geq 0$ and $\int\limits_{\Omega} f(\boldsymbol{x})d\boldsymbol{x} = 0$. Then $f \equiv 0$ in $\bar{\Omega}$.*

Another "vanishing function" theorem reads as follows.

Theorem 1.2 *Let f be a continuous function in a domain Ω such that $\int\limits_{\Omega_0} f(\boldsymbol{x})d\boldsymbol{x} = 0$ for any $\Omega_0 \subset \Omega$. Then $f \equiv 0$ in Ω.*

In the proofs of both theorems the crucial role is played by the theorem on local sign preservation by continuous functions: if f is continuous at $\boldsymbol{x}_0$ and $f(\boldsymbol{x}_0) > 0$, then there exists a ball centered at $\boldsymbol{x}_0$ such $f(\boldsymbol{x}) > 0$ for $\boldsymbol{x}$ from this ball.

Next we shall consider perhaps the most important theorem of multidimensional integral calculus which is Green's-Gauss'-Divergence-.... Theorem. Before entering into details, however, we shall devote some time to the geometry of domains in space.

In one-dimensional space $\mathbb{R}^1$ typical sets with which we will be concerned are open intervals $]a, b[$, where $-\infty \leq a < b \leq +\infty$. For $-\infty < a < b < +\infty$, by $[a, b]$ we will denote the closed interval with endpoints a, b. In this case, we say that $]a, b[$ is the interior of the interval, $[a, b]$ is its closure and the two-point set consisting of $\{a\}$ and $\{b\}$ constitutes the boundary.

In general, for a set Ω, we denote by $\partial\Omega$ its boundary.

The situation in $\mathbb{R}^2$ and $\mathbb{R}^3$ is much more complicated. Let us consider first the two-dimensional situation. Then, in most cases, the boundary $\partial\Omega$ of a two-dimensional region Ω is a closed curve. The two most used analytic descriptions of curves in $\mathbb{R}^2$ are:

a) as a level curve of a function of two variables

$$F(x_1, x_2) = c,$$

b) using two functions of a single variable

$$\begin{aligned} x_1(t) &= f(t), \\ x_2(t) &= g(t), \end{aligned}$$

where $t \in [t_0, t_1]$ (parametric description). Note that since the curve is to be closed, we must have $f(t_0) = f(t_1)$ and $g(t_0) = g(t_1)$.

In many cases the boundary is composed of a number of arcs so that it is impossible to give a single analytical description applicable to the whole boundary.

Example 1.1 Let us consider the elliptical region $x_1^2/a^2 + x_2^2/b^2 \leq 1$. The boundary is then the ellipse

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1.$$

This is the description of the curve (ellipse) as a level curve of the function $F(x_1, x_2) = x_1^2/a^2 + x_2^2/b^2$ (with the constant $c = 1$). Equivalently, the boundary can be written in parametric form as

$$x_1(t) = a\cos t, \qquad x_2(t) = b\sin t,$$

with $t \in [0, 2\pi]$. ■

In three dimensions the boundary of a solid Ω is a two-dimensional surface. This surface can be analytically described as a level surface of a function of three variables

$$F(x_1, x_2, x_3) = c,$$

or parametrically by, this time, three functions of two variables each:

$$\begin{aligned} x_1(t,s) &= f(t,s), \\ x_2(t,s) &= g(t,s), \\ x_3(t,s) &= h(t,s), \qquad t \in [t_0, t_1],\ s \in [s_0, s_1]. \end{aligned}$$

As in two dimensions, it is possible that the boundary is made up of several patches having different analytic descriptions.

Example 1.2 Consider the domain Ω bounded by the ellipsoid

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1.$$

The boundary here is directly given as the level surface of the function $F(x_1, x_2, x_3) = x_1^2/a^2 + x_2^2/b^2 + x_3^2/c^2$:

$$F(x_1, x_2, x_3) = 1.$$

The same boundary can be described parametrically as $\boldsymbol{r}(t,s) = (f(t,s), g(t,s), h(t,s))$, where

$$\begin{aligned} f(t,s) &= a\cos t \sin s \\ g(t,s) &= b\sin t \sin s \\ h(t,s) &= c\cos s \end{aligned}$$

where $t \in [0, 2\pi], s \in [0, \pi]$. ■

One of the most important concepts in partial differential equations is that of the *unit outward normal vector* to the boundary of the set. For a given point $\boldsymbol{p} \in \partial\Omega$ this is the vector $\boldsymbol{n}$, normal (perpendicular) to the boundary at p, pointing outside Ω, and having unit length.

If the boundary of (two or three dimensional) set Ω is given as a level curve of a function F, then the vector given by

$$\boldsymbol{N}(\boldsymbol{p}) = \nabla F|_{\boldsymbol{p}},$$

is normal to the boundary at $\boldsymbol{p}$. However, it is not necessarily unit, nor outward. To make it a unit vector, we divide $\boldsymbol{N}$ by its length; then the unit outward normal is either $\boldsymbol{n} = \boldsymbol{N}/\|\boldsymbol{N}\|$, or $\boldsymbol{n} = -\boldsymbol{N}/\|\boldsymbol{N}\|$ and the proper sign must be selected by inspection.

Example 1.3 *Find the unit outward normal to the ellipsoid*

$$x_1^2 + \frac{x_2^2}{4} + \frac{x_3^2}{9} = 1,$$

at the point $\boldsymbol{p} = (1/\sqrt{2}, 0, 3/\sqrt{2})$.

Since $F(x_1, x_2, x_3) = x_1^2 + \frac{x_2^2}{4} + \frac{x_3^2}{9}$, we obtain $\nabla F = (2x_1, x_2/2, 2x_3/9)$. Therefore

$$\boldsymbol{N}(\boldsymbol{p}) = \nabla F|_{\boldsymbol{p}} = (2/\sqrt{2}, 0, 2/3\sqrt{2}).$$

Furthermore

$$\|\boldsymbol{N}(\boldsymbol{p})\| = \sqrt{2 + 2/9} = \sqrt{20/9} = 2\sqrt{5}/3,$$

thus

$$\boldsymbol{n}(\boldsymbol{p}) = \pm(3/\sqrt{10}, 0, 1/\sqrt{10}).$$

To select the proper sign let us observe that the vector pointing outside the ellipsoid must necessarily point away from the origin. Since the coordinates of the point $\boldsymbol{p}$ are nonnegative, a vector rooted at this point and pointing away from the origin must have positive coordinates, thus finally

$$\boldsymbol{n}(\boldsymbol{p}) = (3/\sqrt{10}, 0, 1/\sqrt{10}).$$

■

If the boundary is given in a parametric way, then the situation is more complicated and we have to distinguish between dimensions 2 and 3.

Let us first consider the boundary $\partial\Omega$ of a two-dimensional domain Ω, described by $\boldsymbol{r}(t) = (x_1(t), x_2(t)) = (f(t), g(t))$. It is known that the derivative vector

$$\frac{d}{dt}\boldsymbol{r}(t_p) = (f'(t_p), g'(t_p))$$

is tangent to $\partial\Omega$ at $\boldsymbol{p} = (f(t_p), g(t_p))$. Since any normal vector at $\boldsymbol{p} \in \Omega$ is perpendicular to any tangent vector at this point, we immediately obtain that

$$\boldsymbol{N}(\boldsymbol{p}) = (-g'(t_p), f'(t_p)). \tag{6.1.1}$$

Therefore the unit outward normal is given by

$$\boldsymbol{n}(\boldsymbol{p}) = \pm\frac{\boldsymbol{N}(\boldsymbol{p})}{\|\boldsymbol{N}(\boldsymbol{p})\|},$$

where the sign must be decided by inspection so that $\boldsymbol{n}$ points outside Ω.

If the domain Ω is 3-dimensional, then its boundary $\partial\Omega$ is a surface described by the 3-dimensional vector function of 2 variables:

$$\boldsymbol{r}(t, s) = (x_1(t, s), x_2(t, s), x_3(t, s)) = (f(t, s), g(t, s), h(t, s)).$$

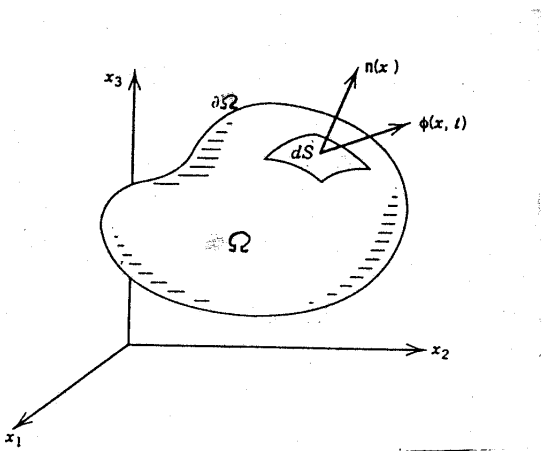

Fig. 3.1 Domain Ω with boundary $\partial\Omega$ showing a surface element dS with the outward normal $\mathbf{n}(\mathbf{x})$ and flux $\phi(\mathbf{x}, t)$ at point $\mathbf{x}$ and time t

In this case, at each point $\partial\Omega \ni \boldsymbol{p} = \boldsymbol{r}(t_p)$, we have two derivative vectors $\boldsymbol{r}'_s(t_p)$ and $\boldsymbol{r}'_t(t_p)$ which span the two dimensional tangent plane to $\partial\Omega$ at $\boldsymbol{p}$. Any normal vector must be thus perpendicular to both these vectors and the easiest way to find such a vector is to use the cross-product:

$$\boldsymbol{N}(\boldsymbol{p}) = \boldsymbol{r}'_t(t_p) \times \boldsymbol{r}'_s(t_p).$$

Again, the unit outward normal is given by

$$\boldsymbol{n}(\boldsymbol{p}) = \pm\frac{\boldsymbol{N}(\boldsymbol{p})}{\|\boldsymbol{N}(\boldsymbol{p})\|} = \pm\frac{\boldsymbol{r}'_t(t_p) \times \boldsymbol{r}'_s(t_p)}{\|\boldsymbol{r}'_t(t_p) \times \boldsymbol{r}'_s(t_p)\|},$$

where the sign must be decided by inspection.

Example 1.4 *Find the outward unit normal to the ellipse $\boldsymbol{r}(t) = (\cos t, 2\sin t)$ at the point $\boldsymbol{p} = \boldsymbol{r}(\pi/4)$* Differentiating, we obtain $\boldsymbol{r}'_t = (-\sin t, 2\cos t)$; this is a tangent vector to ellipse at $\boldsymbol{r}(t)$. Thus,

$$\boldsymbol{N} = (-2\cos t, -\sin t).$$

Next, $\|\boldsymbol{N}\| = \sqrt{4\cos^2 t + \sin^2 t}$ and

$$\boldsymbol{n} = \pm\frac{1}{\sqrt{4\cos^2 t + \sin^2 t}}(2\cos t, \sin t).$$

At the particular point $\boldsymbol{p}$ we have $\cos\pi/2 = \sin\pi/2 = \frac{\sqrt{2}}{2}$, thus $\|\boldsymbol{N}\| = \sqrt{5/2}$ and

$$\boldsymbol{n}(\boldsymbol{p}) = \pm 2/\sqrt{5}(1, 1/2).$$

Since the normal must point outside the ellipse, we must chose the "+" sign and finally

$$\boldsymbol{n}(\boldsymbol{p}) = 2/\sqrt{5}(1, 1/2).$$

■

Another important concept related to the normal is the *normal derivative* of a function. Let us recall that if $\boldsymbol{u}$ is any unit vector and f is a function, then the derivative of f at a point $\boldsymbol{p}$ in the direction of $\boldsymbol{u}$ is defined as

$$f_{\boldsymbol{u}}(\boldsymbol{p}) = \lim_{t\to 0^+} \frac{f(\boldsymbol{p} + t\boldsymbol{u}) - f(\boldsymbol{p})}{t}.$$

Application of the Chain Rule produces the following handy formula for the directional derivative:

$$f_{\boldsymbol{u}}(\boldsymbol{p}) = \nabla f|_{\boldsymbol{p}} \cdot \boldsymbol{u}.$$

Let now f be defined in a neighbourhood of a point $\boldsymbol{p} \in \partial\Omega$. The *normal derivative* of f at $\boldsymbol{p}$ is defined as the derivative of f in the direction of $\boldsymbol{n}(\boldsymbol{p})$:

$$\frac{\partial f}{\partial n}(\boldsymbol{p}) = f_{\boldsymbol{n}}(\boldsymbol{p}) = \nabla f|_{\boldsymbol{p}} \cdot \boldsymbol{n}(\boldsymbol{p}).$$

Example 1.5 Let us consider the spherical coordinates

$$\begin{aligned} x_1 &= r\cos\theta\sin\phi, \\ x_2 &= r\sin\theta\sin\phi, \\ x_3 &= r\cos\phi \end{aligned}$$

and let $f(x_1, x_2, x_3)$ be a function of three variables. This function can be expressed in the spherical coordinates as the function of (r, θ, ϕ)

$$F(r, \theta, \phi) = f(r\cos\theta\sin\phi, r\sin\theta\sin\phi, r\cos\phi) = f(x_1, x_2, x_3).$$

Using the Chain Rule we have

$$\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial r} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial r} + \frac{\partial f}{\partial x_3}\frac{\partial x_3}{\partial r}.$$

Since, for $i = 1, 2, 3$, $\partial x_i/\partial r = x_i/r$, we can write

$$\frac{\partial F}{\partial r} = \frac{1}{r}\nabla f \cdot \boldsymbol{r}. \tag{6.1.2}$$

Assume now that f (and thus F) be given in some neighbourhood of the sphere

$$F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = R^2.$$

To find the outward unit normal to this sphere we note that $\nabla F = (2x_1, 2x_2, 2x_3)$ and $\|\nabla F\| = 2\sqrt{x_1^2 + x_2^2 + x_3^2} = 2R$. Thus,

$$\boldsymbol{n} = \frac{1}{R}(x_1, x_2, x_3) = \frac{1}{R}\boldsymbol{r}. \tag{6.1.3}$$

Geometrically, $\boldsymbol{n}$ is parallel to the radius of the sphere but has unit length.

Combining (6.1.2) with (6.1.3), we see that the normal derivative of f at any point of the sphere is given by

$$\frac{\partial f}{\partial n} = \nabla f \cdot \boldsymbol{n} = \frac{\partial F}{\partial r}.$$

In other words, the normal derivative of any function at the surface of a sphere is equal to the derivative of this function (expressed in spherical coordinates) with respect to r. ■

Next we shall discuss yet another important concept: the flux of a vector field. Let us recall that a vector field is a function $\boldsymbol{f} : \Omega \to \mathbb{R}^d$, where $\Omega \subset \mathbb{R}^d$, where $d = 1, 2, 3....$ In other words, a vector field assigns a vector to each point of a subset of the space.

Definition 1.1 *The flux of the vector field $\boldsymbol{f}$ across the boundary $\partial\Omega$ of a domain $\Omega \subset \mathbb{R}^d$, $d \geq 2$ is*

$$\int\limits_{\partial\Omega} \boldsymbol{f} \cdot \boldsymbol{n} d\sigma.$$

Here, if $d = 2$, then $\partial\Omega$ is a closed curve and the integral above is the line integral (of the second kind). The arc length element $d\sigma$ is to be calculated according to the description of $\partial\Omega$. The easiest situation occurs if $\partial\Omega$ is described in a parametric form by $\boldsymbol{r}(t) = (f(t), g(t))$, $t \in [t_0, t_1]$; then $d\sigma = \sqrt{(f')^2 + (g')^2}dt$ and

$$\int\limits_{\partial\Omega} \boldsymbol{f}\cdot\boldsymbol{n}d\sigma = \int\limits_{t_0}^{t_1} \boldsymbol{f}\cdot\boldsymbol{n}(t)\sqrt{(f')^2(t) + (g')^2(t)}dt.$$

When $d = 3$, then $\partial\Omega$ is a closed surface and the integral above is the surface integral (of the second kind). The surface element $d\sigma$ is again the easiest to calculate if $\partial\Omega$ is given in a parametric form $\boldsymbol{r}(t) = (f(t,s), g(t,s), h(t,s))$, $t \in [t_0, t_1], s \in [s_0, s_1]$. Then $d\sigma = |\boldsymbol{r}_t \times \boldsymbol{r}_s|dtds$ and

$$\int\limits_{\partial\Omega} \boldsymbol{f}\cdot\boldsymbol{n}d\sigma = \int\limits_{t_0}^{t_1}\int\limits_{s_0}^{s_1} \boldsymbol{f}\cdot\boldsymbol{n}(t,s)|\boldsymbol{r}_t \times \boldsymbol{r}_s|dsdt.$$

Remark 1.1 With a little imagination the definition of the flux can be used also in one dimensional case. To do this, we have to understand that the integration is something like a summation of the integrand over all the points of the boundary. In one-dimensional case we have $\Omega = [a, b]$ with $\partial\Omega = \{a\} \cup \{b\}$. A vector field in one-dimension is just a scalar function. The unit outward normal at $\{a\}$ is -1, and at $\{b\}$ is 1. Thus $\boldsymbol{f}\boldsymbol{n}(a) = f(a)(-1)$ and $\boldsymbol{f}\boldsymbol{n}(b) = f(b)(1)$ and the flux across the boundary of Ω is

$$\boldsymbol{f}\cdot\boldsymbol{n}(a) + \boldsymbol{f}\cdot\boldsymbol{n}(b) = f(b) - f(a). \tag{6.1.4}$$

■

Example 1.6 To understand the meaning of the flux let us consider a fluid moving in a certain domain of space. The standard way of describing the motion of the fluid is to associate with any point $\boldsymbol{p}$ of the domain filled by the fluid its velocity $\boldsymbol{v}(\boldsymbol{p})$. In this way we have the *velocity field* of the fluid.

Consider first the one-dimensional case (one can think about a thin pipe). If at a certain point x we have $v(x) > 0$, then the fluid flows to the right, and if $v(x) < 0$, then it flows to the left. Let the points $x = a$ and $x = b$ be the end-points of a section of the pipe and consider the new field $f(x) = \rho(x)v(x)$, where ρ is the (linear) density of the fluid in point x. The flux of f, as defined by (6.1.4), is

$$f(b) - f(a) = \rho(b)v(b) - \rho(a)v(a).$$

For instance, if both $v(b)$ and $v(a)$ are positive, then at $x = b$ the fluid leaves the pipe with velocity $v(b)$ and at $x = a$ it enters the pipe with velocity $v(a)$. In a small interval of time Δt mass of fluid which left the segment through $x = b$ is equal $\rho(b)v(b)\Delta t$ and the mass which entered the segment through $x = a$ is $\rho(a)v(a)\Delta t$, thus, the net rate at which the mass leaves (enters) the segment is equal to $\rho(b)v(b) - \rho(a)v(a)$, that is, exactly to the flux of the field f across the boundary.

This idea can be generalized to more realistic case of three dimensional space. Let us consider a fluid with possibly variable density $\rho(\boldsymbol{p})$ filling a portion of space and moving with velocity $\boldsymbol{v}(\boldsymbol{p})$. We define the mass-velocity field $\boldsymbol{f} = \rho\boldsymbol{v}$. Let us consider now the domain Ω with the boundary Ω. Imagine a small portion ΔS of $\partial\Omega$, which could be considered flat, having the area $\Delta\sigma$. Let $\boldsymbol{n}$ be the unit normal to this surface and consider the rate at which the fluid crosses ΔS. We decompose the velocity $\boldsymbol{v}$ into the component parallel to $\boldsymbol{n}$, given by $(\boldsymbol{v}\cdot\boldsymbol{n})\boldsymbol{n}$ and the tangential component. It is clear that the crossing of the surface can be only due to the normal component (if in time Δt you make two steps perpendicular to the boundary and two parallel, then you will find yourself only two steps from the boundary despite having done four steps). Therefore the mass of fluid crossing ΔS in time Δt is given by

$$\Delta m = \rho(\boldsymbol{v}\cdot\boldsymbol{n})\Delta t\Delta\sigma.$$

Thus, the rate at which the fluid is crossing the whole boundary $\partial\Omega$ (that is, the net rate at which the fluid is filling/leaving the domain Ω) can be approximated by summing up all the contributions Δm over all

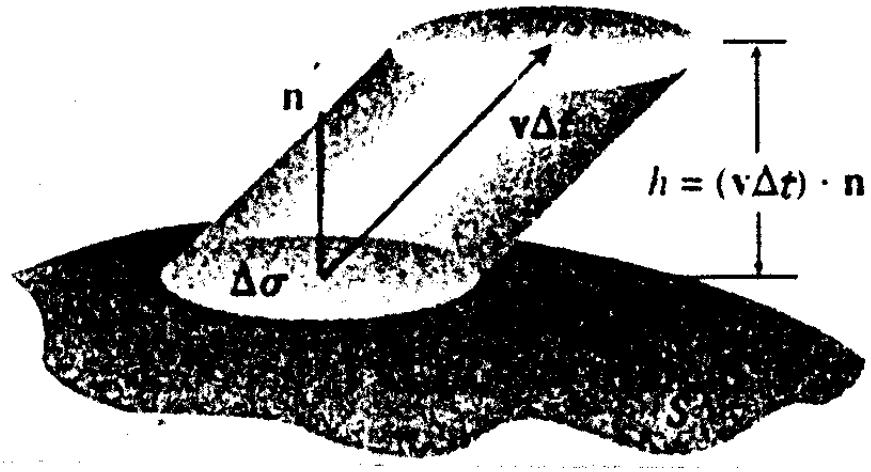

Fig. 3.2 The fluid that flows through the patch ΔS in a short time Δt fills a slanted cylinder whose volume is approximately equal to the base times height - $\mathbf{v} \cdot \mathbf{n}\Delta\sigma\Delta t$. The mass of the fluid in the cylinder is then $\rho(\mathbf{v} \cdot \mathbf{n})Delta\sigma\Delta t$.

patches ΔS of the boundary which, in the limit as ΔS go to zero, is nothing but the flux of $\boldsymbol{f}$:

$$\int_{\partial\Omega} (\rho \boldsymbol{v} \cdot \boldsymbol{n}) d\sigma.$$

Thus again we have the identity

Flux of $\rho\boldsymbol{v}$ across $\partial\Omega$ = the net rate at which the mass of fluid is leaving Ω.

■

Let us return now to the Green-Gauss-.... Theorem. This is a theorem which relates the behaviour of a vector field on the boundary of a domain (flux) with what is happening inside the domain.

Suppose that somewhere inside the domain there is a source of the field (fluid). How can we check this? We can construct a small box around the source and measure whether there is a net outflow, inflow or whether the outflow balances the inflow. If we make this box really small, then we can be quite sure that in the first case there is a source, in the second there is a sink, and in the third case that there is neither sink nor source.

Let us then put this idea into mathematical formulae. Assume that the box B with one vertex at (x_1^0, x_2^0, x_3^0) has the edges parallel to the axes of the coordinate system and of lengths Δx_1, Δx_2 and Δx_3.

We calculate the net rate of flow of the vector field $\boldsymbol{f} = (f_1, f_2, f_3)$ from the box. Following the calculations given earlier, the flow through the top side is given by

$$\boldsymbol{f} \cdot \boldsymbol{n}(x_1^0, x_2^0, x_3^0 + \Delta x_3)\Delta x_1 \Delta x_2 = \boldsymbol{f} \cdot \boldsymbol{j}(x_1^0, x_2^0, x_3^0 + \Delta x_3)\Delta x_1 \Delta x_2 = f_3(x_1^0, x_2^0, x_3^0 + \Delta x_3)\Delta x_1 \Delta x_2,$$

and through the bottom

$$\boldsymbol{f} \cdot \boldsymbol{n}(x_1^0, x_2^0, x_3^0)\Delta x_1 \Delta x_2 = \boldsymbol{f} \cdot (-\boldsymbol{j})(x_1^0, x_2^0, x_3^0)\Delta x_1 \Delta x_2 = -f_3(x_1^0, x_2^0, x_3^0)\Delta x_1 \Delta x_2,$$

thus the net flow through the horizontal sides is

$$\left(f_3(x_1^0, x_2^0, x_3^0 + \Delta x_3) - f_3(x_1^0, x_2^0, x_3^0)\right) \Delta x_1 \Delta x_2 \approx \frac{\partial f_3}{\partial x_3}(x_1^0, x_2^0, x_3^0)\Delta x_1 \Delta x_2 \Delta_3.$$

Similar calculations can be done for the two remaining pairs of the sides and the total flow from the box can be approximated by

$$\left(\sum_{i=1}^{3}\frac{\partial f_i}{\partial x_i}(x_1^0, x_2^0, x_3^0))\right)\Delta x_1\Delta x_2\Delta_3.$$

This expression can be considered to be the net rate of the production of the field (fluid) in the box B. The expression

$$\operatorname{div}\boldsymbol{f} = \sum_{i=1}^{3}\frac{\partial f_i}{\partial x_i}$$

is called the divergence of the vector field $\boldsymbol{f}$ and can considered to be the rate of the production per unit volume (density). To obtain the total net rate of the production in the domain Ω we have to add up contributions coming from all the (small) boxes. Thus using the definition of the integral we obtain that the total net rate of the production is given by

$$\iiint\limits_{\Omega}\operatorname{div}\boldsymbol{f}(x_1, x_2, x_3)d\boldsymbol{v}.$$

Using some common sense reasoning it is easy to arrive at the identity

$$\text{The net rate of production in } \Omega = \text{The net flow across the boundary}$$

The Green–Gauss–... Theorem is the mathematical expression of the above law.

Theorem 1.3 *Let Ω be a bounded domain in $\mathbb{R}^d$, $d \geq 1$, with a piecewise C^1 boundary $\partial\Omega$. Let $\boldsymbol{n}$ be the unit outward normal vector on $\partial\Omega$. Let $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x}))$ be any C^1 vector field on $\bar{\Omega} = \Omega \cup \partial\Omega$. Then*

$$\int\limits_{\Omega}\operatorname{div}\boldsymbol{f}d\boldsymbol{x} = \oint\limits_{\partial\Omega}\boldsymbol{f}\cdot\boldsymbol{n}d\sigma, \tag{6.1.5}$$

where $d\sigma$ is the element of surface of $\partial\Omega$.

Remark 1.2 In one dimension this theorem is nothing but the fundamental theorem of calculus:

$$\int\limits_{a}^{b}\frac{df}{dx}(x)dx = f(b) - f(a). \tag{6.1.6}$$

In fact, for a function of one variable $\operatorname{div}\boldsymbol{f} = \frac{df}{dx}$ and the right-hand side of this equation represents the outward flux across the boundary of $\Omega = [a, b]$, as discussed in Remark 1.1.

Remark 1.3 In two dimensions the most popular form of this theorem is known as the Green Theorem which apparently differs form the one given above. To explain this, let the boundary $\partial\Omega$ of Ω be a curve given by parametric equation $\boldsymbol{r}(t) = (x_1(t), x_2(t))$, $t \in [t_0, t_1]$ and suppose that if t runs from t_0 to t_1, then $\boldsymbol{r}(t)$ traces $\partial\Omega$ in the positive (anticlockwise) direction. Then it is easy to check that the unit outward normal vector $\boldsymbol{n}$ is given by

$$\boldsymbol{n}(t) = \frac{1}{\boldsymbol{r}'(t)}(x_2'(t), -x_1'(t)),$$

where $\boldsymbol{r}'(t) = (x_1(t), x_2(t))$. Thus, if $\boldsymbol{f} = (f_1, f_2)$, then Eq. (6.1.5) takes the form

$$\iint\limits_{\Omega}\left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2}\right)dx_1dx_2 = \int\limits_{t_0}^{t_1}\left(f_1(t)x_2'(t) - f_1(t)x_1'(t)\right)dt \tag{6.1.7}$$

On the other hand, the standard version of Green's theorem reads

$$\iint\limits_{\Omega} \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right) dx_1 dx_2 = \oint\limits_{\partial\Omega} f_1 dx_1 + f_2 dx_2 = \int\limits_{t_0}^{t_1} \left(f_1(t) x_1'(t) + f_2(t) x_2'(t) \right) dt. \tag{6.1.8}$$

To see that these forms are really equivalent, let us define the new vector field $\boldsymbol{g} = (g_1, g_2)$ by : $g_1 = -f_2, g_2 = f_1$. Then

$$\operatorname{div} \boldsymbol{f} = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = \frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2}.$$

The boundary of $\partial\Omega$ is, as above, parameterized by the function $\boldsymbol{r}(t)$. Thus, if we assume that (6.1.8) holds (for an arbitrary vector field), then

$$\begin{aligned} \iint\limits_{\Omega} \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} \right) dx_1 dx_2 &= \iint\limits_{\Omega} \left(\frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2} \right) dx_1 dx_2 \\ = \int\limits_{t_0}^{t_1} \left(g_1(t) x_1'(t) + g_2(t) x_2'(t) \right) dt &= \int\limits_{t_0}^{t_1} \left(f_1(t) x_2'(t) - f_2(t) x_1'(t) \right) dt, \end{aligned}$$

which is (6.1.7). The converse is analogous. ■

2 Conservation laws

Many of the fundamental equations that occurr in the natural and physical sciences are obtained from *conservation laws*. Conservation laws express the fact that some quantity is balanced throughout the process.

In thermodynamics, for example, the First Law of Thermodynamics states that the change in the internal energy in a given system is equal to, or is balanced by, the total heat added to the system plus the work done on the system. Therefore the First Law of Thermodynamics is really an energy balance, or conservation, law.

As an another example, consider a fluid flowing in some region of space that consists of chemical species undergoing chemical reaction. For a given chemical species, the rate of change in time of the total amount of that species in the region must equal the rate at which the species flows into the region, minus the rate at which the species flows out, plus the rate at which the species is created or consumed, by the chemical reactions. This is a verbal statement of a conservation law for the amount of the given chemical species.

Similar balance or conservation laws occur in all branches of science. We can recall the population equation the derivation of which was based on the observation that the rate of change of a population in a certain region must equal the birth rate, minus the death rate, plus the migration rate into or out of the region.

Such conservation laws in mathematics are translated usually into differential equations and it is surprising how many different processes of real world end up with the same mathematical formulation. These differential equations are called the *governing equations* of the process and dictate how the process evolves in time. Here we discuss the derivation of some governing equations starting from the first principles. We begin with a basic one-dimensional model.

2.1 One-dimensional conservation law

Let us consider a quantity $u = u(x,t)$ that depends on a single spatial variable x in an interval $R \subset \mathbb{R}$, and time $t > 0$. We assume that u is a density, or concentration, measured in an amount per unit volume, where the amount may refer to the population, mass, energy or any other quantity. The fact that u depends only on one spatial variable x can be physically justified by assuming e.g. that we consider a flow in a tube,

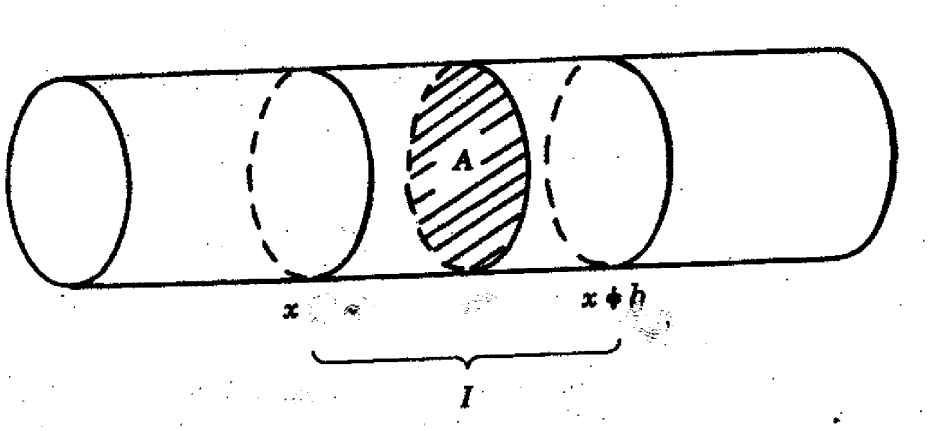

Fig. 3.3 Tube $\mathcal{I}$.

which is uniform (doesn't change in radial direction), or that the tube is very thin so that any change in the radial direction is negligible (later we shall derive the equation for the same conservation law in an arbitrary dimensional space). Note that the discussion here is related to Remark 1.1 and Example 1.6. We consider the subinterval $I = [x, x + h] \subset R$. The total amount of the quantity u contained at time t in the section $\mathcal{I}$ of the tube between x and $x + h$ is given by the integral

$$\text{total amount of quantity in } \mathcal{I} = A \int_x^{x+h} u(s,t)ds,$$

where A is the area of the cross section of $\mathcal{I}$.

Assume now that there is motion of the quantity in the tube in the axial direction. We define the *flux* of u at time t at x to be the scalar function $\phi(x, t)$ which is equal to the amount of the quantity u passing through the cross section at x at time t, per unit area, per unit time. By convention, the flux density at x is positive if the flow at x is in the positive x direction. Therefore, at time t the net rate that the quantity is flowing into the section is the rate that it is flowing in at x and minus the rate that it is flowing out at $x+h$, that is

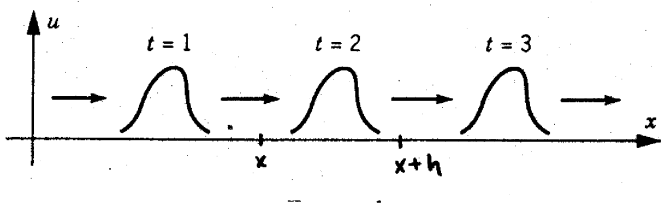

Fig. 3.4 One dimensional flow through cross-sections x and $x+h$.

$$\text{net rate that the quantity flows into } \mathcal{I} = A(\phi(x,t) - \phi(x+h,t)).$$

This equation should be compared with Example 1.6, where the flux density was easy to understand: it was the rate at which the mass of fluid flows through the boundary. Here we are considering the flux density of an arbitrary quantity (arbitrary one-dimensional vector field).

Finally, the quantity u may be destroyed or created inside $\mathcal{I}$ by some internal or external source (e.g. by a chemical reaction if we consider chemical kinetics equations, or by birth/death processes in mathematical biology). We denote this *source function*, which is assumed to be local (acts at each x and t), by $f(x,t,u)$. This function gives the rate at which u is created or destroyed at x at time t, per unit volume. Note that f may depend on u itself (e.g. the rate of chemical reactions is determined by concentration of the chemicals). Given f, we may calculate the total rate that u is created/destroyed in $\mathcal{I}$ by integration:

$$\text{rate that quantity is produced in } \mathcal{I} \text{ by sources} = A \int_x^{x+h} f(s,t,u(s,t))ds.$$

The fundamental conservation law can be formulated as follows: for any section $\mathcal{I}$

$$\begin{aligned}&\text{the rate of change of the total amount in } \mathcal{I}\\ &= \text{net rate that the quantity flows into } \mathcal{I}\\ &+\text{rate that the quantity is produced in } \mathcal{I}\end{aligned}$$

Using the mathematical formulas obtained above we get, having simplified the area A

$$\frac{d}{dt}\int_x^{x+h} u(s,t)ds = \phi(x,t) - \phi(x+h,t) + \int_x^{x+h} f(s,t,u)ds. \tag{6.2.1}$$

The equation above is called a *conservation law in integral form* and holds even if u, f, ϕ are not smooth functions. This form is useful in many cases but rather difficult to handle, therefore it is convenient to reduce it to a differential equation. This requires assuming that all the functions (including the unknown u) are continuously differentiable. Using basic facts from calculus:

(i) $\int_x^{x+h} \phi_s(s,t)ds = \phi(x+h,t) - \phi(x,t)$,

(ii) $\frac{d}{dt}\int_x^{x+h} u(s,t)ds = \int_x^{x+h} u_t(s,t)ds$,

we can rewrite (6.2.1) in the form

$$\int_I (u_t(s,t) + \phi_s(s,t) - f(s,t,u))ds = 0. \tag{6.2.2}$$

Since this equation is valid for any interval I we can use Theorem 1.2 to infer that the integral must vanish identically; that is, changing the independent variable back into x we must have

$$u_t(x,t) + \phi_x(x,t) = f(x,t,u) \tag{6.2.3}$$

for any $x \in R$ and $t > 0$. Note that in (6.2.3) we have two unknown functions: u and ϕ; function f is assumed to be given. Function ϕ is usually to be determined from empirical considerations. Equations resulting from such considerations, which specify ϕ, are often called *constitutive relations* or *equations of state*.

Before we proceed in this direction, we shall show how the procedure described above works in higher dimensions.

2.2 Conservation laws in higher dimensions

It is relatively straightforward to generalize the discussion above to multidimensional space. Let $\boldsymbol{x} = (x_1, x_2, x_3)$ denote a point in $\mathbb{R}^3$ and assume that $u = u(\boldsymbol{x}, t)$ is a scalar density function representing the amount per unit volume of some quantity of interest distributed throughout some domain of $\mathbb{R}^3$. In this domain, let $\Omega \subset \mathbb{R}^3$ be an arbitrary region with a smooth boundary $\partial\Omega$. As in one-dimensional case, the total amount of the quantity in Ω at time t is given by the integral

$$\int_\Omega u(\boldsymbol{x}, t) d\boldsymbol{x},$$

and the rate that the quantity is produced in Ω is given by

$$\int_\Omega f(\boldsymbol{x}, t, u) d\boldsymbol{x},$$

where f is the rate at which the quantity is being produced in Ω.

Some changes are required however when we want to calculate the flow of the quantity in or out Ω. Because now the flow can occur in any direction, the flux is given by a vector $\boldsymbol{\Phi}$ (in Example 1.6, the flux density was given by $\rho(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x})$). However, as we noted in this example, not all the flowing quantity which is close to the boundary will leave or enter the region Ω – the component of $\boldsymbol{\Phi}$ which is tangential to the boundary $\partial\Omega$ will not contribute to the total loss (see also Fig. 2.1). Therefore, if we denote by $\boldsymbol{n}(\boldsymbol{x})$ the outward unit normal vector to $\partial\Omega$, then the net outward flux of the quantity u through the boundary $\partial\Omega$ is given by the surface integral

$$\int_{\partial\Omega} \boldsymbol{\Phi}(\boldsymbol{x}, t) \cdot \boldsymbol{n}(\boldsymbol{x}) d\sigma,$$

where $d\sigma$ denotes a surface element on $\partial\Omega$. Finally, the balance law for u is given by

$$\frac{d}{dt} \int_\Omega u d\boldsymbol{x} = - \oint_{\partial\Omega} \boldsymbol{\Phi} \cdot \boldsymbol{n} d\sigma + \int_\Omega f d\boldsymbol{x}. \tag{6.2.4}$$

The minus sign at the flux term occurs because the outward flux decreases the amount of u in Ω.

As before, the integral form can be reformulated as a local differential equation, provided $\boldsymbol{\Phi}$ and u are sufficiently smooth in $\boldsymbol{x}$ and t. To this end we must use the Gauss theorem (Theorem 1.3) which gives

$$\int_{\partial\Omega} \boldsymbol{\Phi} \cdot \boldsymbol{n} d\sigma = \int_\Omega \operatorname{div} \boldsymbol{\Phi} d\boldsymbol{x}.$$

Using this formula, Equation (6.2.4) can be rewritten as

$$\int_\Omega \left(u_t + \operatorname{div} \boldsymbol{\Phi} - f\right) d\boldsymbol{x} = 0$$

for any subregion Ω. Using the vanishing theorem (Theorem 1.2) we finally obtain the differential form of the general conservation law in 3 (or higher) dimensional space

$$u_t + \operatorname{div} \mathbf{\Phi} = f(\boldsymbol{x}, t, u). \tag{6.2.5}$$

In the next section we discuss a number of constitutive relations which provide examples of the flux function.

3 Constitutive relations and examples

In this section we describe a few constitutive relation which often appear in the applied sciences. It is important to understand that a constitutive relations appear on a different level than conservation law: the latter is is a fundamental law of nature – the mathematical expression of the fact that books should balance (at least in a decent(?) enterprise like the Universe), whereas the former is often an approximate equation having its origins in empirics.

3.1 Transport equation

Probably the simplest nontrivial constitutive relation is when the quantity of interest moves together with the surrounding medium with a given velocity, say, $\boldsymbol{v}$. This can be a simple model for a low density pollutant in the air moving only due to wind with velocity $\boldsymbol{v}$ or, as we discussed earlier, fluid with density u moving with the velocity $\boldsymbol{v}$. Then the flux is given by

$$\mathbf{\Phi} = \boldsymbol{v}u$$

and, if there are no sources or sinks, the transport equation is given by

$$u_t + \operatorname{div} (\boldsymbol{v}u) = 0. \tag{6.3.1}$$

We can rewrite this equation in a more explicit form

$$u_t + \boldsymbol{v} \cdot \nabla u + u \operatorname{div} \boldsymbol{v} = 0,$$

or, if the velocity is constant or, more general, divergence (or source) free,

$$\partial_t u + \boldsymbol{v} \cdot \nabla u = 0. \tag{6.3.2}$$

In more general cases, $\boldsymbol{v}$ may depend on the solution u itself. For example, (6.3.1) may describe a motion of a herd of animals over certain area with u being (suitably rounded) density of animals per, say, square kilometer. Then the speed of the herd will depend on the density - the bigger the squeeze, the slower the animals move. This constitutes an example of a *quasilinear* equation in which the coefficient of the highest derivative depend on the solution:

$$u_t + \operatorname{div} (\boldsymbol{v}(u)u) = 0. \tag{6.3.3}$$

3.2 McKendrick partial differential equation

The derivation of the conservation law was based on a physical intuition about a fluid flow in a physical space. Yet the argument can be used in a much broader context. Consider an age-structured population, described by the density of the population $n(a, t)$ with respect to age $n(a, t)$, and look at the population as if it was 'transported' through stages of life. Taking into account that $n(a, t)\Delta a$ is the number of individuals (females) in the age group $[a, a + \Delta a)$ at time t, we may write that the rate of change of this number

$$\frac{\partial}{\partial t}[n(a, t)\Delta a]$$

equals rate of entry at a minus rate of exit at $a+\Delta a$ and minus deaths. Denoting per capita mortality rate for individuals by $\mu(a,t)$, the last term is simply $-\mu(a,t)n(a,t)\Delta t$. The first two terms require introduction of the 'flux' of individuals J describing the 'speed' of ageing. Thus, passing to the limit $\Delta a \to 0$, we get

$$\frac{\partial n(a,t)}{\partial t} + \frac{\partial J(a,t)}{\partial a} = -\mu(a,t)n(a,t).$$

Let us determine the flux in the simplest case when ageing is just the passage of time measured in the same units.

Here for consistency, we derive the full equation. If the number of individuals in the age bracket $[a, a+\Delta a)$ is $n(a,t)\Delta a$, then after Δt we will have $n(a,t+\Delta t)\Delta a$. On the other hand, $u(a-\Delta t,t)\Delta t$ individuals moved in while $u(a+\Delta a-\Delta t,t)\Delta t$ moved out, where we assumed, for simplicity, $\Delta t < \Delta a$. Thus

$$n(a,t+\Delta t)\Delta a - n(a,t)\Delta a = u(a-\Delta t,t)\Delta t - u(a+\Delta a-\Delta t)\Delta t - \mu(a,t)n(a,t)\Delta a\Delta t$$

or, using the Mean Value Theorem ($0 \le \theta, \theta' \le 1$)

$$n_t(a,t+\theta\Delta t)\Delta a\Delta t = -n_a(a-\Delta t+\theta'\Delta a,t)\Delta a\Delta t - \mu(a,t)n(a,t)\Delta a\Delta t$$

and, passing to the limit with $\Delta t, \Delta a \to 0$, we obtain

$$\frac{\partial n(a,t)}{\partial t} + \frac{\partial n(a,t)}{\partial a} = -\mu(a,t)n(a,t). \tag{6.3.4}$$

This equation is defined for $a > 0$ and the flow is to the right hence we need a boundary condition. In this model the birth rate enters here: the number of neonates ($a = 0$) is the number of births across the whole age range:

$$n(0,t) = \int_0^\omega n(a,t)m(a,t)da,$$

where m is the maternity function. Eq. (6.3.4) also must be supplemented by the initial condition

$$n(a,0) = n_0(a)$$

describing the initial age distribution.

3.3 Diffusion/heat equations

Assume at first that there are no sources, and write the basic conservation law in one-dimension as

$$u_t + \phi_x = 0 \tag{6.3.5}$$

In many physical (and not only) problems it was observed that the substance (represented here by the density u) moves from the regions of higher concentration to regions of lower concentration (heat, population, etc. offer a good illustration of this). Moreover, the larger difference the more rapid flow is observed. At a point, the large difference can be expressed as a large gradient of the concentration u, so it is reasonable to assume that

$$\phi(x,t) = -F(u_x(x,t)),$$

where F is an increasing function passing through $(0,0)$. The explanation of the minus sign comes from the fact that if the x-derivative of u at x is positive, that is, u is increasing, then the flow occurs in the negative direction of the x axis. Moreover, the larger the gradient, the larger (in magnitude) the flow. Now, the simplest increasing function passing through $(0,0)$ is a linear function with positive leading coefficient, and this assumption gives the so-called *Fick law*:

$$\phi(x,t) = -Du_x(x,t) \tag{6.3.6}$$

where D is a constant which is to be determined empirically. We can substitute Fick's law (6.3.6) into the conservation law (6.2.3) (assuming that the solution is twice differentiable and D is a constant) and get the one dimensional *diffusion equation*

$$u_t - Du_{xx} = 0, \tag{6.3.7}$$

which governs conservative processes when the flux is specified by Fick's law.

To understand why this equation governs also the evolution of the temperature distribution in a body, we apply the conservation law (6.2.3) to heat (thermal energy). Then the amount of heat in a unit volume can be written as $u = c\rho\theta$, where c is the specific heat of the medium, ρ the density of the medium and θ is the temperature. The heat flow is governed by Fourier's law which states that the heat flux is proportional to the gradient of the temperature, with the proportionality constant equal to $-k$, where k is the thermal conductivity of the medium. Assuming c, ρ, k to be constants, we arrive at (6.3.7) with $D = k/c\rho$.

Generalization of Equation (6.3.7) to multidimensional cases requires some assumptions on the medium. To simplify the discussion we assume that the medium is isotropic, that is, the process of diffusion is independent of the orientation in space. Then the Fick law states that the flux is proportional to the gradient of u, that is,

$$\boldsymbol{\Phi} = -D\nabla u$$

and the diffusion equation, obtained from the conservation law (6.2.5) in the absence of sources, reads

$$u_t = \operatorname{div}(D\nabla u) = D\Delta u, \tag{6.3.8}$$

where the second equality is valid if D is a constant.

3.4 Variations of diffusion equation

In many cases the evolution is governed by more then one process. One of the most widely occuring cases is when we have simultaneously both transport and diffusion. For example, when we have a pollutant in the air or water, then it is transported with the moving medium (due to the wind or water flow) and, at the same time, it is dispersed by diffusion. Without the latter, a cloud of pollution would travel without any change - the same amount that was released from the factory would eventually reach other places with the same concentration. Diffusion, as we shall see later, has a dispersing effect, that is, the pollution will be eventually deposited but in a uniform, less concentrated way, making life (perhaps) more bearable.

Mathematical expression of the above consideration is obtained by combining equations (6.3.1) and (6.3.8):

$$u_t + \operatorname{div}(\boldsymbol{v}u) = D\Delta u, \tag{6.3.9}$$

or, equivalently, by determining the expression for the flux taking both processes simultaneously into account. The resulting equation (6.3.9) is called the *drift-diffusion equation.*

When the sources are present and the constitutive relation is given by Fick's law, then the resulting equation

$$u_t - D\Delta u = f(\boldsymbol{x}, t, u) \tag{6.3.10}$$

is called the *reaction-diffusion equation.* If f is independent of u, then physically we have the situation when the sources are given a priori (like the injection performed by an observer); the equation becomes then a linear non-homogeneous equation. Sources, however, can depend on u. The simplest case is then when $f = cu$ with c constant. Then the source term describes spontaneous decay (or creation) of the substance at the rate that is proportional to the concentration. In such a case Equation (6.3.10) takes the form

$$u_t - D\Delta u = cu \tag{6.3.11}$$

and can be thought of as the combination of the diffusion and the law of exponential growth.

If f is nonlinear in u, then Equation (6.3.10) has many important applications, particularly in the combustion theory and mathematical biology.

Here we describe the so called *Fisher equation* which appears as a combination of the one-dimensional logistic process and the diffusion. It is reasonable to assume that the population is governed by the logistic law, which is mathematically expressed as the ordinary differential equation

$$\frac{du}{dt} = ru(N-u), \tag{6.3.12}$$

where u is the population in some region, $rN > 0$ is the *growth rate* and $N > 0$ is the carrying capacity of this region. If we are however interested is the spatial distribution of individuals, then we must introduce the density of the population $u(\boldsymbol{x}, t)$, that is, the amount of individuals per unit of volume (or surface) and the conservation law will take the form

$$u_t + \operatorname{div} \boldsymbol{\Phi} = ru\,(N-u)\,. \tag{6.3.13}$$

As a first approximation, it is reasonable to assume that within the region the population obeys the Fick law, that is, individuals migrate from the regions of higher density to regions of lower density. This is not always true (think about people migrating to cities) and therefore the range of validity of Fick's law is limited; in such cases $\boldsymbol{\Phi}$ has to be determined otherwise. However, if we assume that the resources are evenly distributed, then Fick's law can be used with good accuracy. In such a case, Equation (6.3.13) takes the form

$$u_t - D\Delta u = ru\,(N-u)\,, \tag{6.3.14}$$

which is called the Fisher equation. Note, that if the capacity of the region N is very large, then writing the Fisher equation in the form

$$u_t - D\Delta u = cu\left(1 - \frac{u}{N}\right),$$

where $c = rN$, we see that it is reasonable to neglect the small term u/N and approximate the Fisher equation by the linear equation (8.3.12) which describes a combination of diffusion-type migratory processes and exponential growth.

4 Systems of partial differential equations

As in the case of ODEs, a single equation describes an evolution of a quantity, which is governed by a non-responsive environment; that is, the environment acts on the described quantity but the latter does not have any impact on the environment. This is too crude an assumption in numerous applications: different species of animals interact with each other, and also with the environment changing it in a dramatic way, chemical substances react changing each other, etc. In physics we have conservation laws of many quantities, such as mass, energy, momentum, which are related to each other making the resulting equations coupled.

In this section we shall describe some examples of systems of partial differential equations.

4.1 Shallow water waves

The shallow water equations represent mass and momentum balance in a fluid and are in many ways typical for systems of hyperbolic conservation laws. The geometry of the problem is presented in Fig. 6.1. Water of constant density lies above a flat bottom. The vertical coordinate is denoted by y, with $y = 0$ at the bottom, while x is measured along the bottom. We assume that the problem is two-dimensional so that there is no variation in the direction perpendicular to the xy plane. Let H be the depth of undisturbed water and the profile of the free surface of water at time t over point x is given by $h(x,t)$. The pressure at the free surface is the ambient air pressure p_0. We can always assume $p_0 = 0$ by moving the reference point. Let a typical wave on the surface have length L. The shallow water assumption amounts to saying that H is very small compared to L: $H << L$. Because of this we ignore any vertical motion of water and assume that there is a flow velocity $u(x,t)$ in direction x that is an average velocity over the depth.

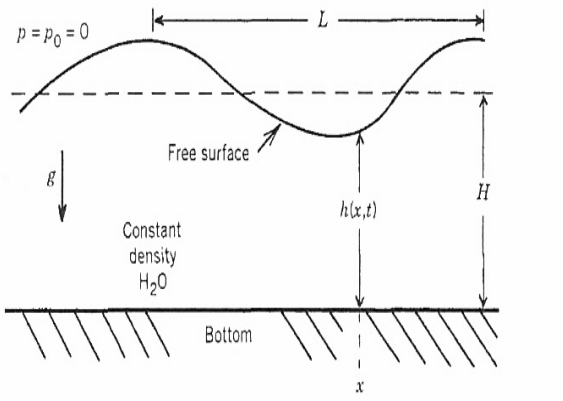

Figure 6.1: Waves in shallow water

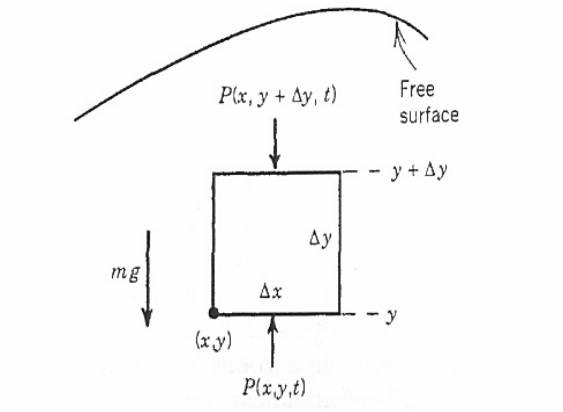

Figure 6.2: Hydrostatic balance

Since we ignored the vertical motion, each notional volume of water, which we consider to be a box with dimensions $\Delta x, \Delta y, \Delta z$, is in a hydrostatic balance in the vertical direction. This means that the upward pressure balances the downward pressure plus the weight mg of the volume, see Fig. 6.2. If we denote the pressure at (x, y) at time t by $P(x, y, t)$, the hydrostatic balance equation can be written as

$$P(x, y, t)\Delta x\Delta z = P(x, y + \Delta y, t)\Delta x\Delta z + \rho g\Delta x\Delta y\Delta z.$$

Dividing by $\Delta x\Delta y\Delta z$ and taking limits as $\Delta y \to 0$ yields

$$P_y(x, y, t) = -\rho g.$$

We can integrate this equation from a depth y to the free surface h gives

$$P(x, y, t) = p_0 + \rho g(h(x, t) - y) = \rho g(h(x, t) - y), \tag{6.4.15}$$

as we assumed $p_0 = 0$.

Next we derive an equation for mass balance. Consider a region bounded by $x = a$ and $x = b$. The rate of change of mass in the region must be equal to the net flux of mass into the region. The mass contained in

this region is $\Delta z\rho\int_a^b h(x,t)dx$ while the flux of mass at x is $\rho u(x,t)h(x,t)\Delta z$. This gives

$$\frac{d}{dt}\int_a^b h(x,t)dx = u(a,t)h(a,t) - u(b,t)h(b,t), \tag{6.4.16}$$

where we divided by $\rho\Delta z$. As before, assuming smoothness of the involved quantities, we arrive at the differential equation

$$h_t + (uh)_x = 0. \tag{6.4.17}$$

Next we have to derive the conservation of the momentum in the x direction (recall, there is no motion in the vertical direction). As before, the rate of change of the total momentum in the region between $x = a$ and $x = b$ must equal the net momentum flux into the region augmented by the forces (pressure) acting on the region. Total momentum is $\int_a^b \rho u(x,t)h(x,t)dx\Delta z$. Momentum flux through a section at x is the momentum multiplied by the velocity at x; that is, $\rho u^2 h\Delta z$. The force acting on the face of area $h(a,t)\Delta z$ is the pressure $P(a,y,t)$. However, the pressure changes with depth and therefore the force at $x = a$ is equal to

$$\int_0^{h(a,t)} P(a,y,t)\Delta z dy = \int_0^{h(a,t)} \rho g(h(a,t)-y)\Delta z dy = \frac{\rho g}{2}(h(a,t))^2\Delta z,$$

with an analogous expression at $x = b$. Consequently, the integral for of the momentum balance equation is

$$\frac{d}{dt}\int_a^b h(x,t)u(x,t)dx = u(a,t)h^2(a,t) - u(b,t)h^2(b,t) + \frac{g}{2}\left(h^2(a,t) - h^2(b,t)\right), \tag{6.4.18}$$

where again we divided each term by $\rho\Delta z$. Assuming, as before, that all terms are sufficiently smooth, we arrive at the differential equation for the momentum balance

$$(uh)_t + \left(u^2h + \frac{g}{2}h^2\right)_x = 0. \tag{6.4.19}$$

Equations (6.4.17), (6.4.19) are a coupled nonlinear system of PDEs for the velocity u and height h of the free surface. Clearly, (6.4.17) can be expanded as

$$h_t + uh_x + hu_x = 0$$

and, using the latter, (6.4.20) can be simplified to

$$\begin{aligned}
&(uh)_t + \left(u^2h + \frac{g}{2}h^2\right)_x \\
&= u(h_t + hu_x + h_x u) + h(u_t + uu_x + gh_x) = h(u_t + uu_x + gh_x) \\
&= h(u_t + uu_x + gh_x) = 0.
\end{aligned}$$

If we assume that $h \neq 0$, then the system can be written as

$$\begin{aligned}
h_t + uh_x + hu_x &= 0, \\
u_t + uu_x + gh_x &= 0.
\end{aligned} \tag{6.4.20}$$

This is still a system of quasilinear equations which can be simplified to a system of linear equations if we assume that the amplitude of the waves is small.

4.2 Small amplitude approximation

If we restrict analysis to waves of small amplitude; that is, which do not deviate much from the undisturbed depth H

$$h(x,t) = H + \eta(x,t)$$

and if the velocity is small

$$u(x,t) = 0 + v(x,t),$$

where η, v and their derivatives are small, then the system (6.4.20) can be linearized. Substituting the above into (6.4.20) still gives the nonlinear system

$$\begin{aligned} \eta_t + v\eta_x + (H+\eta)v_x &= 0, \\ v_t + vv_x + g\eta_x &= 0. \end{aligned} \tag{6.4.21}$$

However, the smallness assumption allows for discarding quadratic terms. Hence, retaining only the linear terms in (6.4.21) yields

$$\begin{aligned} \eta_t + Hv_x &= 0, \\ v_t + g\eta_x &= 0. \end{aligned} \tag{6.4.22}$$

We can eliminate one unknown in (6.4.22). Indeed, assuming that the solution is twice continuously differentiable, we differentiate the first equation with respect to t and the second with respect to x. Hence, using $v_{xt} = v_{tx}$, we obtain

$$\eta_{tt} - gH\eta_{xx} \tag{6.4.23}$$

This is the classical wave equation with wave speed $c^2 = gH$, and thus the d'Alembert formula gives the general solution in the form

$$\eta(x,t) = F_1(x - \sqrt{gH}t) + F_2(x + \sqrt{gH}t)$$

where F_1 and F_2 are arbitrary twice differentiable functions.

5 Systems obtained by coupling scalar conservation laws

In Subsection 2.1 we introduced a scalar one dimensional conservation law (6.2.3)

$$u_t(x,t) + \phi_x(x,t) = f(x,t,u)$$

In general, we can consider n interacting susbstances/species, represented by concentrations $u_i(x,t)$ where again, for simplicity, we consider one dimensional case: $x \in \mathbb{R}$. With each density u_i we associate its flux ϕ_i with usual convention that ϕ_i is positive if the net flow of u_i is to the right and let f_i be the rate of production of u_i. In general f_i depend on t, x and all densities. Then, the same argument as in the scalar case leads to the system of n differential equations

$$(u_i)_t + (\phi_i)_x = f_i, \quad i = 1, 2, \ldots, n. \tag{6.5.24}$$

Below we shall consider some special cases of (6.5.30).

5.1 An epidemiological system with age structure

In this case, the densities $u_i, i = 1, 2, 3$, (later denoted by s, i, r) are the age densities of individuals with different disease status. We assume that in each class the demographical processes are the same thus, in absence of the disease, each would follow the McKendrick equation (6.3.4). However, the presence of the disease introduces terms which we can treat as source terms and we begin with explaining them.

Standard epidemiological models treat the population as homogeneous apart from the differences due to the disease. Then, for the description of the epidemics the population is divided into three main classes: susceptibles (individuals who are not sick and can be infected), infectives (individuals who have the disease and can infect others) and removed (individuals who were infective but recovered and are now immune, dead or isolated). Depending on the disease, other classes can be introduces to cater for e.g. latent period of the disease. We denote by $S(t), I(t), R(t)$ the number of individuals in the classes above. By

$$S(t) + I(t) + R(t) = P(t)$$

we denote the total population size. In many models it is assumed that the population size is constant disregarding thus vital dynamics such as births and deaths. The conservation law can be written as

$$\begin{aligned} S' &= -\lambda S + \delta I, \\ I' &= \lambda S - (\gamma + \delta) I, \\ R' &= \gamma I \end{aligned} \tag{6.5.25}$$

with $S(0) = S_0, I(0) = I_0, R(0) = R_0$ and $S_0 + I_0 + R_0 = P$. The parameter λ is the force of infection, δ is the recovery rate and γ is the recovery/removal rate. While δ and γ are usually taken to be constant, the force of infection requires a constitutive law. The simplest is the law of mass action

$$\lambda = c\phi \frac{I}{P} \tag{6.5.26}$$

where c is the contact rate (the number of contacts that a single individual has with other individuals in the population per unit time, ϕ is the infectiveness; that is, the probability that a contact with an infective will result in infection and I/P is the probability that the contacted individual is infective. In what follows we shall denote $k = c\phi/P$.

There are many other assumptions underlying this model: that the population is homogenous, that no multiple infections are possible, that an infected individual immediately become infective.

Concerning the nature of the disease the basic distinction is between those which are not lethal and do not impart immunity (influenza, common cold) and those which could be caught only once (leading to death or immunity) such as measles or AIDS. In the first case, $\gamma = 0$ and the model is referred to as an SIS model and in the second $\delta = 0$ and the model is called an SIR model.

In many cases the rate of infection significantly varies with age and thus it is important to consider the age structure of the population. Thus we expect the interaction of the vital dynamics and the infection mechanism to produce a nontrivial behaviour.

To introduce the model we note again that, in absence of the disease, we would have the linear model introduced in (6.3.4). Because of the epidemics, the population is partitioned into the three classes of susceptibles, infectives and removed, represented by their respective age densities $s(a,t), i(a,t)$ and $r(a,t)$. Now, if we look at the population of susceptibles, than we see that it is losing individuals at the rate $\lambda(a,t)s(a,t)$ and gaining at the rate $\delta(a)i(a,t)$, where we have taken into account that the infection force and the cure rate are age dependent. Similarly, the source terms for the other two classes are given by the (age dependent) terms of the (6.5.25) model. This leads to the system

$$\begin{aligned} s_t(a,t) + s_a(a,t) + \mu(a)s(a,t) &= -\lambda(a,t)s(a,t) + \delta(a)i(a,t), \\ i_t(a,t) + i_a(a,t) + \mu(a)i(a,t) &= \lambda(a,t)s(a,t) - (\delta(a) + \gamma(a))i(a,t), \\ r_t(a,t) + r_a(a,t) + \mu(a)r(a,t) &= \gamma(a)i(a,t). \end{aligned} \tag{6.5.27}$$

$$\begin{aligned} s(0,t) &= \int_0^\omega m(a)(s(a,t) + (1-q)i(a,t) + (1-w)r(a,t))da, \\ i(0,t) &= q\int_0^\omega m(a)i(a,t)da, \end{aligned}$$

$$r(0,t) = w\int_0^\omega m(a)r(a,t)da, \tag{6.5.28}$$

where $q \in [0,1]$ and $w \in [0,1]$ are the vertical transmission coefficients of infectiveness and immunity, respectively. The system is complemented by initial conditions $s(a,0) = s_0(a), i(a,0) = i_0(a)$ and $r(a,0) = r_0(a)$. We remark that here we assumed that the death and birth coefficients are not significantly affected by the disease. In particular, if we add the equations together, then we obtain that the total population $p(a,t) = s(a,t) + i(a,t) + r(a,t)$ satisfies

$$\begin{aligned} p_t(a,t) + p_a(a,t) + \mu(a)p(a,t) &= 0, \\ p(0,t) &= \int_0^\omega m(a)p(a,t)da, \\ p(a,0) &= p_0(a) = s_0(a) + i_0(a) + r_0(a), \end{aligned}$$

that is, the disease does not change the global picture of the evolution of the population. Finally, we have to specify a constitutive relation for the force of infection λ. This usually is given by the equation

$$\lambda(a,t) = K_0(a)i(a,t) + \int_0^\omega K(a,s)i(s,t)ds, \tag{6.5.29}$$

where the two terms on the right hand side are called the intracohort and intercohort terms, respectively. The intracohort term describes the situation in which individuals only can be infected by those of their own age while the intercohort term describes the case in which they can be infected by individuals of any age.

5.2 Systems of reaction-diffusion equations

A prototype of a single reaction diffusion equation is the Fisher equation

$$u_t - Du_{xx} = ru\left(1 - \frac{u}{K}\right),$$

in which a substance (population) of density u moves along a line from regions of higher concentration to regions of lower concentration and, at the same time, increases according to the logistic growth.

The simplest constitutive relation for the fluxes ϕ_i in (6.5.30) is to use the Fick law; that is, to assume that ϕ_i only depends on the gradient of u_i: $\phi_i = -D_i u_{xx}$. Then we obtain

$$(u_i)_t - D_i u_{xx} = f_i(t, x, u_1, \ldots, u_n), \quad i = 1, 2, \ldots, n, \tag{6.5.30}$$

however, as we shall see later, ϕ_i may depend on other densities as well. If D_i is positive, the flow is from the region of higher concentration to the region of lower concentration, as in the Fisher equation. However, it may occur that organisms are attracted to their kind; that is, the flow is from low concentration regions to high concentration regions and then the diffusion coefficient should be negative. In the case of chemotaxis, the animals are attracted by chemicals (pheromones) which also requires negative coefficient of diffusion. Let us consider two typical examples.

Predator-prey models with space structure

Suppose that we have two populations: predators and prey distributed over certain (one-dimensional) area. The densities are respectively, u and v. As in the classical Lotka-Volterra model we assume that, in the

absence of predators, the prey grows exponentially while the predators, in the absence of prey, decays at an exponential rate. The effect of predation is modelled by the mass action law.

$$\begin{aligned} \text{The growth (production) rate of prey} &= av - buv, \\ \text{The growth (production) rate of predators} &= -cu + duv. \end{aligned}$$

where a, b, c, d are positive proportionality constants. Further, we assume that both predators and pray disperse over the area according to the Fick law. Then (6.5.30) takes the form

$$\begin{aligned} v_t &= D_1 v_{xx} + av - buv, \\ u_t &= D_2 u_{xx} - cv + duv. \end{aligned} \tag{6.5.31}$$

Chemotaxis

Another process leading to reaction-diffusion equations is the motion of organisms under the influence of certain chemicals. In normal conditions the organisms disperse following the usual diffusion process. However, sometimes (e.g. when some individuals find food) a certain chemical is secreted and the other organisms move up the concentration gradient. This process is called chemotaxis. We derive a one dimensional model of it. Let a denotes the population density of the organism and c the concentration of the chemical secreted by it. The conservation laws for a and c are

$$\begin{aligned} a_t + (\phi_1)_x &= 0, \\ c_t + (\phi_2)_x &= F(a, c), \end{aligned} \tag{6.5.32}$$

where, in the first equation, we neglected the vital processes of the organism so that the density only changes due to motion. On the other hand, the chemical is produced and a rate F. Next we specify the constitutive relations. The chemical is assumed to move only due to diffusion

$$\phi_2 = -\delta c_x$$

and its production rate is proportional to the concentration of the organism and decays at a constant rate:

$$F(a, c) = fa - kc$$

where f and k are positive constants. On the other hand, the organisms move due to diffusion and due to the chemical attraction

$$\phi_1 = \phi_{diff} + \phi_{chem} = -\mu a_x + \phi_{chem}$$

where μ, called the motility, is a positive constant. For ϕ_{chem} we assume, as noted before, that it is (positively) proportional to the concentration gradient. However, also it is experimentally established that the higher concentration of the chemical, the larger the flux of organisms. Again, a simple form of the flux satisfying these assumptions is

$$\phi_{chem} = \nu(ac_x)_x$$

and we arrive at the system

$$\begin{aligned} a_t &= \mu a_{xx} - \nu(ac_x)_x, \\ c_t &= \delta u_{xx} + fa - ac. \end{aligned} \tag{6.5.33}$$

The system is strongly coupled and the fact that the diffusion coefficient depends on the solution makes the analysis very complicated.

6 Partial differential equations through random walks

We have seen the diffusion operator appearing in a variety of cases though usually its derivation was specific to each process and was based on purely heuristic considerations. However, the reason of the abundance of the diffusion operator in nature is that it is a macroscopic effect of microscopic random motion of particles. Let us provide some justification for this.

6.1 Uncorrelated random walk

First we consider the unrestricted one-dimensional random walk. A particle starts at the origin of the x-axis executes a jump of length δ to the right or to the left. Let x_i be a random variable that assumes the value δ if the particle moves to the right at the ith step and $-\delta$ if it moves to the left. We assume that each step is independent of the others, so that the x_is are identically distributed independent random variables. Let the probabilities that the particle moves to the right or left equal p and q, respectively. Since the probabilities are the same for each step, we have $P(x_i = \delta) = p$ and $P(x_i = -\delta) = q$ for each i. If we assume that the particle cannot rest, we must have $p + q = 1$ (if we assumed that the particle can rest, then we would have to introduce the probability of it not moving $r = 1 - p - q$). The position of the particle after n jumps is given by the random variable

$$X_n = x_1 + x_2 + \ldots + x_n.$$

It is easy to see that that the expected value of x_i is

$$E(x_i) =< x_i >= (p - q)\delta$$

and, since the expectation is a linear function

$$E(X_n) =< X_n >= (p - q)n\delta.$$

For the variance we have

$$V(X_n) = E(X_n - E(X_n))^2) = E(X_n^2) - (E(X_n))^2.$$

Since $V(x_i^2) = \delta^2(p + q) = \delta^2$, again by linearity of E, we see that

$$\begin{aligned} V(X_n) &= \sum_{i=1}^{n}(< x_i >^2 - < x_i >^2) = \sum_{i=1}^{n}(\delta^2 - (p - q)^2\delta^2) \\ &= n\delta^2(1 - (p - q)^2) = 4pqn\delta^2 \end{aligned}$$

upon using $p + q = 1$.

In nature, random walk can is a model for a Brownian motion. Again we consider a one dimensional situation. Here, the motion is caused by collisions of the particle with the particles of the fluid and each collision results in a small jump by δ of the particle to the right or to the left. Many such collisions occur in a unit time.

For a particle immersed in fluid experimentally we can observe average displacement per unit time, denoted by c and the variance of the observed displacement around the average which we denote by $D > 0$. Assume that there are n collisions per unit time. Thus, approximately we should have

$$c \approx (p - q)\delta n, \tag{6.6.34}$$

and

$$D \approx 4pq\delta^2 n. \tag{6.6.35}$$

Since the motion appears to be continuous, we have to consider the limit of the above equations as $\delta \to 0$ while $n \to \infty$ in such a way that D and c given above remain fixed. Now, if $p \neq q$ and $p - q$ does not tend to zero as $\delta \to 0, n \to \infty$ we have

$$\delta n \to \frac{c}{p - q}$$

but then $4pq\delta^2 n \to 0$ yielding $D = 0$ in which case the motion would be deterministic. If we want $D \neq 0$, then $p - q \to 0$ yielding $p, q \to 1/2$. If $p = q = 1/2$ in the discrete case, then $c = 0$. However, if $p - q \neq 0$ then $c \neq 0$ and we have a drift. This conditions can be realized if $p = (1 + b\delta)/2, q = (1 - b\delta)/2$ for some yet unspecified b chosen so that $0 \leq p, q \leq 1$. These leads to $p \to 1/2, q \to 1/2$ and

$$(p - q)\delta n = b\delta^2 n$$

so that for this limit to be finite and non zero we need $\delta^2 n$ to converge to a finite limit. Combining this with (6.6.35), we find

$$\delta^2 n \to D \tag{6.6.36}$$

and thus $b = c/D$, yielding

$$(p-q)\delta n \to c. \tag{6.6.37}$$

Having fixed the constants, let us derive the equation governing the random walk in the continuous limit as $\delta \to 0, n \to \infty$ in such a way that (6.6.36) holds. For n steps to occur in a unit time, one step must occur in $\tau = 1/n$ units of time. We derive the formula for the probability that a particle starting at $x = 0$ at $t = 0$ will be at the position x at the time t. Thus, we must have

$$k\tau = t, \quad X_k = x.$$

We define

$$v(x,t) = P(X_k = x)$$

at time t; that is, $v(x,t)$ is the probability that at the approximate time t the particle is located (approximately) at the point x. Then, v satisfies the difference equation

$$v(x, t+\tau) = pv(x-\delta, t) + qv(x+\delta, t). \tag{6.6.38}$$

If we assume that v is differentiable, we can expand it in the Taylor series

$$\begin{aligned} v(x, t+\tau) &= v(x,t) + \tau v_t(x,t) + O(\tau^2), \\ v(x \pm \delta, t) &= v(x,t) \pm \delta v_x(x,t) + \frac{1}{2}\delta^2 v_{xx}(x,t) + O(\delta^3). \end{aligned} \tag{6.6.39}$$

Substituting into (6.6.38), we obtain

$$v_t = (q-p)\frac{\delta}{\tau}v_x + \frac{1}{2}\frac{\delta^2}{\tau}v_{xx} + \tau^{-1}O(\tau^2) + \tau^{-1}O(\delta^3)$$

Now, since $\delta^2 n = \delta^2/\tau = O(1)$, we have $\tau^{-1}O(\delta^3) = O(1)\delta^{-2}O(\delta^3)$ and we can re-write the above as

$$v_t = (q-p)\frac{\delta}{\tau}v_x + \frac{1}{2}\frac{\delta^2}{\tau}v_{xx} + O(\tau) + O(\delta)$$

and passing to the limit as $\delta \to 0, \tau \to 0$ in such a way that (6.6.36) (with $\tau = 1/\tau$) holds

$$v_t = -cv_x + \frac{1}{2}Dv_{xx}$$

where we used (6.6.37) and (6.6.36). In this equation, v must be interpreted as the probability density; that is

$$P(a \le x \le b) = \int_a^b v(x,t)ds$$

at the time t.

We note that the derivation of the diffusion limit required $\delta^2/\tau \to D$ which, in turn, implies $\delta/\tau \to \infty$ since $\delta \to 0$. In other words, for the finite diffusion coefficient (variance) the velocity of the particle must be infinite. While certainly nonphysical, it is in agreement with the properties of the diffusion equation which predicts instantaneous transmission of signals. This drawback can be removed by considering the correlated random walk.

6.2 Correlated random walk and the telegrapher's equation

It is rather unexpected but random walks can lead to wave type equations which are rather not associated with diffusion processes. Similarly to the previous discussion we consider a particle which can jump by δ to the left or to the right. Each jump is executed in time τ. However, while in the previous model the probabilities p and q described likelihoods of moving to the right or to the left, here p and q are the probabilities that the particle will, respectively, persist moving in the same direction and reverse the direction. Thus, let $\alpha(x,t)$ be the probability that a particle is at the point x and arrived there from the left, whereas $\beta(x,t)$ is the probability that a particle is at x and arrived there from the right. Thus, α and β characterize right and left moving particles, respectively. Thus, using the total probability, we have

$$\begin{aligned} \alpha(x,t+\tau) &= p\alpha(x-\delta,t)+q\beta(x-\delta,t), \\ \beta(x,t+\tau) &= q\alpha(x+\delta,t)+p\beta(x+\delta,t). \end{aligned} \tag{6.6.40}$$

Next we introduce the assumption that the shorter the time τ, the greater probability of persistence; that is, $p\to 1$ as $\tau\to 0$ which, in turn, yields $q\to 0$. Assuming that both p and q are differentiable functions of τ, we can write

$$\begin{aligned} p &= 1-\lambda\tau+o(\tau), \\ q &= \lambda\tau+o(\tau) \end{aligned} \tag{6.6.41}$$

where λ is the rate of reversal of direction as $\tau\to 0$. Indeed, in a unit time $n=1/\tau$ jumps are made and on average $qn=q/\tau$ of them result in the reversal of direction. Thus, the instantaneous rate of reversals as at $\tau=0$ is λ.

If we expand α and β, we get

$$\begin{aligned} \alpha_t(x,t)\tau+\alpha(x,t)+o(\tau) &= -p\alpha_x(x,t)\delta+p\alpha(x,t)-q\beta_x(x,t)\delta+q\beta(x,t)+o(\delta), \\ \beta_t(x,t)\tau+\beta(x,t)+o(\tau) &= p\beta_x(x,t)\delta+p\beta(x,t)+q\alpha_x(x,t)\delta+q\alpha(x,t)+o(\delta) \end{aligned}$$

or, using (6.6.41),

$$\begin{aligned} \alpha_t(x,t)\tau &= -(1-\lambda\tau)\alpha_x(x,t)\delta-\lambda\alpha(x,t)\tau-\lambda\tau\beta_x(x,t)\delta+\lambda\tau\beta(x,t)+o(\tau)+o(\delta), \\ \beta_t(x,t)\tau &= (1-\lambda\tau)\beta_x(x,t)\delta-\lambda\beta(x,t)\tau+\lambda\tau\alpha_x(x,t)\delta+\lambda\alpha(x,t)+o(\tau)+o(\delta). \end{aligned}$$

Next, assuming that $\delta/\tau\to\gamma$, the speed of motion, as $\delta,\tau\to 0$, we obtain the following coupled system of partial differential equation

$$\begin{aligned} \alpha_t(x,t) &= -\gamma\alpha_x(x,t)-\lambda\alpha(x,t)+\lambda\beta(x,t), \\ \beta_t(x,t) &= \gamma\beta_x(x,t)+\lambda\alpha(x,t)-\lambda\beta(x,t), \end{aligned} \tag{6.6.42}$$

where, as before α and β are to be interpreted as the probability densities.

Since at the point x the particle must have arrived either from the left or from the right, the function

$$v(x,t)=\alpha(x,t)+\beta(x,t) \tag{6.6.43}$$

is the probability density that a particle is at the point x at the time t hence it is the same object as in the uncorrelated random walk. If we introduce the net flux to the right

$$w(x,t)=\alpha(x,t)-\beta(x,t),$$

then (6.6.42) can be transformed, by adding and subtracting, into

$$\begin{aligned} v_t(x,t)+\gamma w_x(x,t) &= 0, \\ w_t(x,t)+\gamma v_x(x,t) &= -2\lambda w. \end{aligned} \tag{6.6.44}$$

In particular, (6.6.44) can be reduced to

$$v_{tt} - \gamma^2 v_{xx} + 2\lambda v_t = 0 \tag{6.6.45}$$

which is the damped wave equation with waves moving with the speed γ, as follows from the microscopic description.

We observe that for λ close to 0 we have strong correlations resulting practically in a pure wave motion which clearly does not display any stochasticity. On the other hand, when we divide both sides of Eq. (6.6.45) by 2λ and let $\lambda \to \infty$ (which corresponds to very weak correlations) in such a way that $\gamma^2/\lambda \to D$ (which means that the speed of jumps goes to infinity), formally the equation will become

$$-\frac{D}{2} v_{xx} + v_t = 0 \tag{6.6.46}$$

which is the diffusion equation of the uncorrelated random walk. This agrees well with intuition. However, it is important to remember that the above reasoning does not constitute a proof that a correlated random walk tends to an uncorrelated random walk if both the reversal rate and the speed tend to infinity. For instance, (6.6.45) is second order in time and requires two initial conditions whereas (6.6.46) is first order in time so specifying these initial conditions would render the problem unsolvable. Problems of this type are called *singularly perturbed* and require a delicate analysis. It can be proved, however, that of solutions of (6.6.45) to a solution of (6.6.46) holds for large t.

7 Initial and boundary conditions

As we saw earlier, a partial differential equation can have many solutions. The same situation occurs also for ordinary differential equations - then the solution typically depends on as many constants as is the order of the equation. To be able to select a unique solution of an ODE we have to impose an appropriate number of side conditions which typically are the initial conditions specified at a certain point on the real line.

However, we have seen that the solution of a second order PDE depends in general on two arbitrary function so that the specification of the side conditions will be different. Below we shall describe typical choices of such conditions which are motivated mainly by physical considerations.

Most equations which we have described above are supposed to reflect certain physical processes and due to this fact the variables were explicitly divided into time and spatial variable (of course there is rather no mathematical justification for such separation but, as we shall see later, it makes sense also from the mathematical point of view). Then it is clear that the process will take place in some spatial domain, say $\Omega \subset \mathbb{R}^3$ over some time $t \in I \in (t_0, t_1)$ $(-\infty < t_0 < t_1 \leq +\infty)$. Thus our equation is to be considered in the region $\Omega_T = \Omega \times I$. If we keep to our physical intuition, the process should have started somehow, which means that we must prescribe some initial state. In the case of the wave equation we have also the second derivative in time so, as in ODEs, we expect that we shall need also the initial speed. These conditions are called the *initial conditions*:

$$u(\boldsymbol{x}, t_0) = \psi(\boldsymbol{x}) \quad \text{in } \Omega \tag{6.7.47}$$

for equations of first order in time, and

$$u(\boldsymbol{x}, t_0) = \psi(\boldsymbol{x}), \quad u_t(\boldsymbol{x}, t_0) = \phi(\boldsymbol{x}) \quad \text{in } \Omega \tag{6.7.48}$$

for equations of second order in time, where ϕ and ψ are prescribed functions.

Initial conditions are usually not sufficient to determine a unique solution. If we think about the heat distribution in the container, mathematically represented by the region Ω, it is clear that the distribution of heat inside will be dictated by the conditions on the boundary. For equations of second order in spatial variables, such as heat and wave equations, we have three typical *boundary conditions*

(D) u is specified at the boundary $\partial\Omega$ (*Dirichlet condition*),

(N) the normal derivative $\partial u/\partial n = \nabla u \cdot \boldsymbol{n}$ is specified at the boundary $\partial\Omega$ (*Neumann condition*),

(R) the combination $\partial u/\partial n + au$ is specified at the boundary $\partial\Omega$ (*Robin conditions*),

where a is a given function of $\boldsymbol{x}, t$. Each condition is to hold for each $t \in I$ and $\boldsymbol{x} \in \partial\Omega$.

The boundary conditions are usually written in the form of equation to hold at the boundary. For instance (N) can be written as

$$\frac{\partial u}{\partial n} = g(\boldsymbol{x}, t),$$

where g is a given function (at least known on $\partial\Omega \times I$) and is called the *boundary datum*. Any of these boundary conditions is called *homogeneous* if the specified function g vanishes on the boundary.

Remark 7.1 Note that if $\Omega = (a, b) \subset \mathbb{R}$, then $\partial\Omega = \{a\} \cup \{b\}$ and at the right endpoint we have $\partial u/\partial n(b, t) = u_x(b, t)$ and at the left end point we have $\partial u/\partial n(a, t) = -u_x(a, t)$.

To make the boundary conditions more understandable, we shall put some physical meaning into them.

The heat equation

The interpretation of all three boundary conditions for the heat equation is slightly different. The Dirichlet condition is easy to explain - we keep the boundary of the body in a fixed constant temperature (e.g. by submerging it in the large container with melting ice – then we will have homogeneous boundary conditions).

To explain Neumann and Robin conditions we should remember that due to Fourier's law (or Fick's law for diffusion) the flux is proportional to the normal derivative. Therefore the homogeneous Neumann condition tells us that nothing escapes through the boundary, that is, the domain is completely insulated.

Assume now that our body is immersed in a reservoir with known temperature g, then according to Fourier's law, the flux from the domain into the boundary must be proportional to the difference of temperatures. Therefore in this case we arrive at the Robin condition

$$\frac{\partial u}{\partial n}(\boldsymbol{x}, t) = a(u(\boldsymbol{x}, t) - g(t))$$

for all $\boldsymbol{x} \in \partial\Omega$.

7.1 Conditions at infinity and other nonstandard boundary conditions

Though it is rather nonphysical, now and then people are solving problems in the whole space. The reason for this is that such problems are easier to solve in a closed form, providing on one hand benchmark solutions, and on the other hand, in many cases, approximate solutions. In such cases physics usually provide some indication as to how the solution should behave at infinity. Typically it is assumed that the solution decays sufficiently fast at infinity which sometimes takes form of the requirement that a certain integral of solution is finite. This expresses a condition of finiteness of total population, mass or energy of the system.

As we have seen in the case of McKendrick equation, sometimes boundary (or in general, side) conditions take the form of the requirement that certain expression involving the solution satisfies a prescribed relation. There is no general theory for such problems and each case may require a separate theory.

Chapter 7

First order partial differential equations

1 Linear equations

We start with the simplest transport equation. Assume that u is a concentration of some substance (pollutant) in a fluid (amount per unit length). This substance is moving to the right with a speed c. Then the differential equation for u has the form:

$$u_t + cu_x = 0. \tag{7.1.1}$$

Let us consider more general linear first order partial differential equation (PDE) of the form:

$$au_t + bu_x = 0, \quad t, x \in \mathbb{R} \tag{7.1.2}$$

where a and b are constants. This equation can be written as

$$D_{\boldsymbol{v}} u = 0, \tag{7.1.3}$$

where $\boldsymbol{v} = a\boldsymbol{j} + b\boldsymbol{i}$ ($\boldsymbol{j}$ and $\boldsymbol{i}$ are the unit vectors in, respectively, t and x directions), and $D_{\boldsymbol{v}} = \nabla u \cdot \boldsymbol{v}$ denotes the directional derivative in the direction of $\boldsymbol{v}$. This means that the solution u is a constant function along each line having direction $\boldsymbol{v}$, that is, along each line of equation $bt - ax = \xi$. Along each such a line the value of the parameter ξ remains constant. However, the solution can change from one line to another, therefore the solution is a function of ξ, that is the solution to Eq. (7.1.2) is given by

$$u(x,t) = f(bt - ax), \tag{7.1.4}$$

where f is an arbitrary differentiable function. Such lines are called the *characteristic lines* of the equation.

Example 1.1 To obtain a unique solution we must specify the initial value for u. Hence, let us consider the initial value problem for Eq. (7.1.2): find u satisfying both

$$\begin{aligned} au_t + bu_x &= 0 \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= g(x) \quad x \in \mathbb{R}, \end{aligned} \tag{7.1.5}$$

where g is an arbitrary given function. From Eq. (7.1.4) we find that

$$u(x,t) = g\left(-\frac{bt - ax}{a}\right). \tag{7.1.6}$$

We note that the initial shape propagates without any change along the characteristic lines, as seen below for the initial function $g = 1 - x^2$ for $|x| < 1$ and zero elsewhere. The speed $c = b/a$ is taken to be equal to 1.

Fig. 4.1 The graph of the solution in Example 1.1

Example 1.2 Let us consider a variation of this problem and try to solve the initial- boundary value problem

$$\begin{aligned} au_t + bu_x &= 0 \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= g(x) \quad x > 0, \qquad (7.1.7) \\ u(0,t) &= h(t) \quad t > 0, \qquad (7.1.8) \end{aligned}$$

for $a, b > 0$ From Example 1.1 we have the general solution of the equation in the form

$$u(x,t) = f(bt - ax).$$

Putting $t = 0$ we get $f(-ax) = g(x)$ for $x > 0$, hence $f(x) = g(-x/a)$ for $x < 0$. Next, for $x = 0$ we obtain $f(bt) = h(t)$ for $t > 0$, hence $f(x) = h(x/b)$ for $x > 0$. Combining these two equations we obtain

$$u(x,t) = \begin{cases} g(-\frac{bt-ax}{a}) & \text{for} \quad x > bt/a \\ h(\frac{bt-ax}{b}) & \text{for} \quad x < bt/a \end{cases}$$

Now, let us consider what happens if $a = 1 > 0, b = -1 < 0$. Then the initial condition defines $f(x) = g(-x)$ for $x < 0$ and the boundary condition gives $f(x) = h(-x)$ also for $x < 0$! Hence, we cannot specify both initial and boundary conditions in an arbitrary way as this could make the problem ill-posed.

The physical explanation of this comes from the observation that since the characteristics are given by $\xi = x + t$, the flow occurs in the negative direction and therefore the values at $x = 0$ for any t are uniquely determined by the initial condition. Therefore we see that to have a well-posed problem we must specify the boundary conditions at the point where the medium flows into the region.

The method we have just used is often called the *geometric method*. It is easy to understand, but sometimes difficult to apply, especially for non-homogeneous problems. Fortunately, this method can be easily reformulated in a more analytic language. Let us introduce the change of variables according to

$\xi = \xi(t,x),\ \eta = \eta(t,x)$; then

$$u_t = u_\xi \xi_t + u_\eta \eta_t, \quad u_x = u_\xi \xi_x + u_\eta \eta_x,$$

and the equation can be written as

$$a(u_\xi \xi_t + u_\eta \eta_t) + b(u_\xi \xi_x + u_\eta \eta_x) = u_\xi(a\xi_t + b\xi_x) + u_\eta(a\eta_t + b\eta_x) = 0$$

If we require the coefficient at u_η to be zero, the easiest way is to introduce $\eta_t = b,\ \eta_x = -a$, that is $\eta = bt - ax$. Note, that this is exactly the characteristic direction! However, this is an incomplete change of variables as originally we have had two independent variables t, x but we have only one new variable η. That means that knowing η alone we are not able to tell values of x and t. The trick is to introduce another variable, say, $\xi = \xi(x,t)$ in such a way that the system

$$\eta = bt - ax, \quad \xi = \xi(x,t)$$

is uniquely solvable. To keep things as simple as possible, we may here take $\xi(x,t) = ct + dx$ with c, d picked so as to have the determinant $bd + ac \neq 0$. If a and b are not equal to zero, then the easiest choice would be either $\xi = t$ or $\xi = x$, respectively. However, sometimes it is more convenient to use the orthogonal lines given by $\xi = at + bx$.

We illustrate this approach in the following example.

Example 1.3 Find the general solution to the following equation

$$u_t + 2u_x - (x+t)u = -(x+t).$$

Introducing new variables according to $\xi = t,\ \eta = 2t - x$ we transform the equation into

$$v_\xi - (3\xi - \eta)v = -(3\xi - \eta).$$

This equation can be regarded as a linear first order ordinary differential equation in ξ with a parameter η.

To find the integrating factor we solve the homogeneous equation

$$\mu_\xi = -(3\xi - \eta)\mu;$$

the integration gives

$$\mu(\xi,\eta) = e^{-\left(\frac{3}{2}\xi^2 - \eta\xi\right)}.$$

Multiplying both sides of the equation by μ and rearranging the terms we obtain

$$\left(e^{-\left(\frac{3}{2}\xi^2 - \eta\xi\right)} v(\xi,\eta)\right)_\xi = -(3\xi - \eta)e^{-\left(\frac{3}{2}\xi^2 - \eta\xi\right)},$$

hence the solution is given by

$$v(\xi,\eta) = 1 + C(\eta)e^{\left(\frac{3}{2}\xi^2 - \eta\xi\right)},$$

where C is an arbitrary differentiable function of one variable.

In the original variables we obtain

$$u(x,t) = 1 + C(2t - x)e^{\left(-t^2/2 + tx\right)}$$

where C is an arbitrary differentiable function of one variable.

Example 1.4 Find the solution of the equation

$$u_t + 2u_x - (x+t)u = -(x+t),$$

which satisfies the initial condition:

$$u(x,0) = f(x), \quad x > 0,$$

and

$$u(0,t) = g(t), \quad t > 0.$$

We use the general solution obtained in the previous example:

$$u(x,t) = 1 + C(2t - x) \exp\left(-t^2/2 + tx\right).$$

Thus

$$f(x) = u(x,0) = 1 + C(-x)$$

for $x > 0$. To avoid misunderstanding we introduce the variable s as the argument of C. Thus

$$C(s) = f(-s) - 1, \quad s < 0.$$

On the other hand

$$g(t) = u(0,t) = 1 + C(2t) \exp\left(-t^2/2\right),$$

thus

$$C(2t) = \exp\left(t^2/2\right)(g(t) - 1).$$

Since we need the function $s \to C(s)$, we introduce $s = 2t$; then $s > 0$ and

$$C(s) = \exp\left(s^2/8\right)(g(s/2) - 1).$$

Thus, we have defined C for all values of the argument by

$$C(s) = \begin{cases} f(-s) - 1 & \text{for} \quad s < 0, \\ \exp\left(s^2/8\right)(g(s/2) - 1) & \text{for} \quad s > 0, \end{cases}$$

In the solution to the given problem the function C is evaluated at $s = 2t - x$, so C has different definitions for $2t - x > 0$ and $2t - x < 0$. Therefore the solution to the given initial-boundary value problem is given by

$$u(x,t) = \begin{cases} 1 + (f(-2t + x) - 1) \exp\left(-t^2/2 + tx\right) & \text{for} \quad x > 2t, \\ \exp\left((x^2 + tx)/8\right)(g(t - x/2) - 1) & \text{for} \quad x < 2t. \end{cases}$$

For example, if $f(x) = x$ and $g(t) = \sin t$, then the solution is given by

$$u(x,t) = \begin{cases} 1 + (-2t + x - 1) \exp\left(-t^2/2 + tx\right) & \text{for} \quad x > 2t, \\ \exp\left((x^2 + 4tx)/8\right)(\sin(t - x/2) - 1) & \text{for} \quad x < 2t, \end{cases}$$

In the figures below note how the point where the analytic description of the solution changes across the characteristic $x - 2t = 0$.

The same principle could be used to solve equations with variable coefficients. In fact, consider the equation

$$a(x,t)u_t + b(x,t)u_x = 0. \tag{7.1.9}$$

This equation asserts that the derivative of u in the direction of the vector $(b(x,t), a(t,x))$ is equal to zero at each point (x,t) where this vector is not vanishing. We can consider a family of curves $t = t(x)$ which are tangent to these vectors at each point. In other words, these curves will have at each point the slope equal to a/b; this is equivalent to say that they satisfy the differential equation

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

Assume that this equation has solutions given by $\phi(x,t) = \eta$ where η is an arbitrary constant. Note that in principle for each η, the equation $\phi(x,t) = \eta$ determines a curve and if a point (x,t) belongs to this curve, the tangent vector at this point is given by $(b(x,t), a(x,t))$. From the general theory it follows then that the

Fig 4.2. The graph of the solution in Example 1.4.

Fig 4.3. The three-dimensional visualization of the solution in Example 1.4.

normal vector to this curve is given by the gradient of ϕ: $(\phi_x(x,t), \phi_t(x,t))$. Let us see what this means for our equation. Consider, for an arbitrary differentiable f, the function

$$u(x,t) = f(\phi(x,t)) \tag{7.1.10}$$

Inserting u into the equation (7.1.9) we obtain

$$au_t + bu_x = f' \cdot (a\phi_t + b\phi_x) = 0,$$

due to the orthogonality property of $[\phi_x, \phi_t]$ which was mentioned above. Thus, u given by Eq. (7.1.10) is the general solution to (7.1.9).

Note that the solution doesn't change along the curves $\phi(x,t) = C$ which are therefore the characteristic curves of the equation.

Remark 1.1 The extension of the above considerations to the nonhomogeneous cases can be done along the lines similarly as in the constant coefficient case. However, some difficulties can occur as $\eta = \phi(x,t)$ not always gives rise to a well-defined change of variables. Thus, we shall present an alternative way of solving problems of the form

$$\begin{aligned} u_t + c(x,t)u_x &= f(x,t), \\ u(x,0) &= u_0(x). \end{aligned} \tag{7.1.11}$$

As we explained above, the left-hand side of the equation is the derivative of u along the characteristic defined by the equation $dx/dt = c(x,t)$. Thus, we obtain the so-called *characteristic system*

$$\begin{aligned} \frac{du}{dt} &= f(x,t), \\ \frac{dx}{dt} &= c(x,t), \end{aligned} \tag{7.1.12}$$

which, in principle, can be solved. The second equation is independent of the first so that it determines the equation of characteristics $x = x(t)$. This solution can be substituted into the first equation the solution of which gives the values of u as a function of t that is a parameter along a characteristic. To determine this characteristic and express u as a function of (x,t) we must realize that the solution of (7.1.12) contains two arbitrary constants of integration. To find them, we denote by $\xi = \xi(x,t)$ the x-coordinate of the point where the characteristic passing through (x,t) crosses the x-axis. Hence, we solve (7.1.12) subject to the initial conditions

$$u(0) = u_0(\xi), \qquad x(0) = \xi. \tag{7.1.13}$$

Eliminating ξ from equations for u and x produces the solution u in terms of x and t only.

We illustrate these two methods in the example below.

Example 1.5 Find the solution to the following initial value problem

$$\begin{aligned} u_t + xu_x + u &= 0, \quad t > 0, x \in \mathbb{R}, \\ u(x,0) &= u_0(x), \end{aligned}$$

where $u_0(x) = 1 - x^2$ for $|x| < 1$, and $u_0(x) = 0$ for $|x| \geq 1$.

The differential equation for characteristic curves is

$$\frac{dt}{dx} = \frac{1}{x}$$

which gives $x = \eta e^t$, thus $\eta = xe^{-t}$. If the equation hadn't contained the zero order term u, that is, if it had been in the form

$$u_t + xu_x = 0,$$

then the general solution would have had the form

$$u(x,t) = u_0(xe^{-t}),$$

for arbitrary function f.

However, since we have the additional term, we have to perform the full change of variables. Fortunately, putting $\xi = t$ and $\eta = xe^{-t}$ produces an invertible change of variables, as $t = \xi$ and $x = e^{\xi}\eta$.

Defining $v(\xi,\eta) = u(x,t)$ we obtain that

$$\begin{aligned} 0 = u_t + xu_x + u &= v_\xi\xi_t + v_\eta\eta_t + x(v_\xi\xi_x + v_\eta\eta_x) + v \\ &= v_\xi + v \end{aligned}$$

Therefore we obtain the solution in the form

$$v(\xi,\eta) = C(\eta)e^{-\xi}$$

or

$$u(x,t) = C(xe^{-t})e^{-t}$$

where C is an arbitrary function. In fact, putting $t = 0$ we see that $u(x,0) = C(x) = u_0(x)$ so that

$$u(x,t) = u_0(xe^{-t})e^{-t}$$

The method described in Remark 1.1) we requires writing the characteristic system

$$\begin{aligned} \frac{du}{dt} &= -u, \\ \frac{dx}{dt} &= x, \end{aligned} \tag{7.1.14}$$

with the initial conditions $u(0) = u_0(\xi), \quad x(0) = \xi$ where ξ is the intercept of the characteristic and the x-axis. The solution of (7.1.14) is

$$u(t) = ae^{-t}, \qquad x(t) = be^{t},$$

where a, b are integration constants. Using the initial conditions we get

$$u(0) = a = u_0(\xi), \qquad x(0) = b = \xi$$

Thus, $x(t) = \xi e^t$ and, eliminating ξ, we get

$$u(x,t) = u_0(xe^{-t})e^{-t},$$

in accordance with the formula derived by the other method.

To find the solution of our particular initial value problem we have

$$u_0(x) = u(x,0) = C(x),$$

and according to the definition of $C(x) = 1 - x^2$ for $|x| < 1$ and $C(x) = 0$ for $|x| \geq 1$. In the solution the function C appears composed with xe^{-t}. Accordingly, $C(xe^{-t}) = 1 - x^2e^{-2t}$ for $|xe^{-t}| < 1$ and $C(xe^{-t}) = 0$ for $|xe^{-t}| \geq 1$. Thus we obtain

$$u(x,t) = \begin{cases} (1 - x^2e^{-2t})e^{-t} & \text{for} \quad |t| > \ln|x|, \\ 0 & \text{for} \quad |t| \leq \ln|x|, \end{cases}$$

Fig 4.4. Characteristics of the equation in Example 1.5.

Fig 4.5. 3-dimensional visualization of the solution in Example 1.5.

Fig 4.6. The graph of the solution in Example 1.5 for times $t = 0, 0.5, 1, 1.5, 2, 2.5$.

The described procedure is also not restricted to equations in two independent variables. In fact, let us consider the equation

$$au_t + bu_x + cu_y = 0 \tag{7.1.15}$$

This equation expresses the fact that the directional derivative in the direction of the vector $[a, b, c]$ is equal to zero, that is, that the solution does not change along any line with parametric equation $t = t_0 + as, x = x_0 + bs, y = y_0 + cs$. Such a line can be written also as the pair of equations $ax - bt = \xi, cy - bt = \eta$. For each pair ξ, η this pair describes a single line parallel to $[a, b, c]$, that is, the solution u can be a function of ξ and η only. Thus we obtain the following general solution to (7.1.15)

$$u(x, y, t) = f(ax - bt, ay - ct), \tag{7.1.16}$$

where f is an arbitrary differentiable function of two variables.

Example 1.6 Find the solution to the following initial value problem:

$$u_t + 2u_x + 3u_y + u = 0, \quad u(x, y, 0) = u_0(x, y) = e^{-x^2-y^2}.$$

We use the change of coordinates suggested by the characteristics:

$$\xi = x - 2t, \quad \eta = y - 3t$$

supplemented by

$$\alpha = t,$$

so that the change is invertible:

$$t = \alpha, \quad x = \xi + 2\alpha, \quad y = \eta - 3\alpha.$$

Then, putting $u(x,y,t) = v(\xi,\eta,\alpha)$, we find that

$$\begin{aligned} u_t &= (-2)v_\xi + (-3)v_\eta + v_\alpha, \\ u_x &= v_\xi, \\ u_y &= v_\eta \end{aligned}$$

so that

$$0 = u_t + 2u_x + 3u_y + u = v_\alpha + v.$$

The last equation has the general solution

$$v(\xi,\eta,\alpha) = C(\xi,\eta)e^{-\alpha}$$

so that

$$u(x,y,t) = C(x-2t, y-3t)e^{-t}.$$

To solve the initial value problem we put $t = 0$ to get

$$e^{-x^2-y^2} = C(x,y)$$

so that the solution is of the form

$$u(x,y,t) = e^{-(x-2t)^2-(y-3t)^2}e^{-t}.$$

Remark 1.2 The method of the characteristic system can be also easily adapted for equations in higher dimensions. In such a case, we obtain a system of $n+1$ equations, where n is the number of space variables. For instance, the characteristic system for the equation from the previous example is of the form

$$\begin{aligned} \frac{du}{dt} &= -u, \\ \frac{dx}{dt} &= 2, \\ \frac{dy}{dt} &= 3 \end{aligned}$$

with initial conditions $u(0) = u_0(\xi,\eta), x(0) = \xi, y(0) = \eta$.

2 Nonlinear equations

The linear models discussed in the previous section described propagation of signals with speed that is independent of the solution. This is the case for e.g. acoustic signals that propagate at a constant sound speed. However, signals with large amplitude propagate at a speed that is proportional to the local density of the air. Thus, it is important to consider equations in which the speed of propagation depends on the value of the solution. The qualitative picture in such cases is completely different than for linear equations.

We shall focus on the simple nonlinear value problem:

$$\begin{aligned} u_t + c(u)u_x &= 0, \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}. \end{aligned} \tag{7.2.17}$$

Here, c is a given smooth function of u. The above equation is simply the conservation law

$$u_t + \phi(u)_x = 0$$

after differentiation with respect to x so that $c(u) = \phi_u(u)$.

To analyse (7.2.17) we assume at first that a C^1 solution $u(x,t)$ to (7.2.17) exists for all $t > 0$. Motivated by the linear approach we define characteristic curves by the differential equation

$$\frac{dx}{dt} = c(u(x,t)). \tag{7.2.18}$$

Of course, contrary to the situation for linear equations, the right hand side of this equation is not known *a priori*. Thus, characteristics cannot be determined in advance. However, assuming that we know them, we can solve (7.2.18) getting

$$x = x(t,\eta)$$

where η is an integration constant. Fixing η we obtain, as in the linear case, that

$$\frac{d}{dt}u(x(t,\eta),t) = u_x x_t + u_t = u_x c(u) + u_t = 0$$

thus the solution is constant any characteristic and depends only on the single variable ξ. Now, we observe that along each characteristic at each point the slope is given by $c(u(x(t,\xi),t))$ but this depends only on ξ that is fixed along each characteristic. Thus, the slope of each characteristic is constant so that each characteristic is a straight line. Alternatively, we can prove it by differentiating $x = x(t,\xi)$ twice to get

$$\frac{d^2x}{dt^2} = \frac{d}{dt}c(u(x(t,\xi),t)) = c_u(u)\cdot(u_x x_t + u_t) = c_u(u)\cdot(u_x c(u) + u_t) = 0.$$

Thus, from each point (x,t) we draw a straight line to an unspecified (yet) point $(\xi,0)$ on the x-axis so that the equation of this characteristic is given by

$$x - \xi = c(u(\xi,0))t = c(u_0(\xi))t \tag{7.2.19}$$

because the slope is constant and therefore it must be equal to the value of $c(u)$ at the initial time $t = 0$. Note that since c and u_0 are known functions, then (7.2.19) determines ξ implicitly in terms of a given (x,t) (if it can be solved). Since the solution u is constant along characteristics, we have the implicit formula for it

$$u(x,t) = u(\xi,0) = u_0(\xi)$$

where ξ is given by (7.2.19). In some instances (7.2.19) can be solved explicitly.

Example 2.1 Find the solution to the initial value problem

$$\begin{aligned} u_t + uu_x &= 0, \quad x \in \mathbb{R}, t > 0, \\ u(x,0) &= x. \end{aligned}$$

The characteristics emanating from a point $(\xi, 0)$ on the x-axis have speed $c(u_0(\xi)) = u_0(\xi)$. Using the discussion above we see that the solution is given implicitly by

$$u(x,t) = c(u_0(\xi)) = \xi$$

where

$$x - \xi = t\xi$$

so that easily

$$\xi = \frac{x}{1+t}$$

and the solution is given by

$$u(x,t) = \frac{x}{1+t}.$$

Note that in this example the solution is defined for values of $t > 0$. This is not always the case and our prior discussion based on the assumption that for any (x,t) with $t > 0$ we have a smooth solution can be invalid. However, we can prove the following result

Theorem 2.1 *If the functions c and u_0 are $C^1(\mathbb{R})$ and if u_0 and c are either both nondecreasing or both nonincreasing on $\mathbb{R}$, then the initial value problem (7.2.17) has a unique solution defined implicitly by the parametric equations*

$$\begin{aligned} u(x,t) &= u_0(\xi), \\ x - \xi &= c(u(\xi,0))t = c(u_0(\xi))t \end{aligned} \tag{7.2.20}$$

Proof. Since we have shown that if there is a smooth solution to (7.2.17), then it must be of the form (7.2.20), all we have to do is to show that (7.2.20) is uniquely solvable and that the function $u(x,t)$ obtained from (7.2.20) is, in fact, a solution of (7.2.17). As we already mentioned, the second equation of (7.2.20) should uniquely determine $\xi = \xi(x,t)$ for any (x,t), $t > 0$. This is an implicit equation of the form

$$F(\xi, x, t) = 0$$

where $F(\xi, x, t) = \xi + c(u_0(\xi))t - x$ and, from the general theory, this equation is (locally) solvable around any point at which $F_\xi \neq 0$. In this case,

$$F_\xi(\xi, x, t) = 1 + c'_u(u_0(\xi))u'_{0,\xi}t \neq 0$$

and since from the assumption, c_0, u_0 are either both increasing, or both decreasing, the derivatives are of the same sign and the second term is always positive. Thus, under the assumptions, $F_\xi \neq 0$ everywhere and (7.2.20) is uniquely solvable for any choice of (x,t) and $\xi(x,t)$ is differentiable. Implicit differentiation of the first equation of (7.2.20) gives

$$u_t(x,t) = u'_{0,\xi}\xi_t, \qquad u_x(x,t) = u'_{0,\xi}\xi_x$$

where, by implicit differentiation of the second equation in (7.2.20),

$$-\xi_t = c'(u_0(\xi))u'_{0,\xi}\xi_t t + c(u_0(\xi)), \qquad 1 - \xi_x = c'(u_0(\xi))u'_{0,\xi}\xi_x t$$

so that

$$u_t(x,t) = -\frac{c(u_0(\xi))u'_{0,\xi}}{1 + c'_u(u_0(\xi))u'_{0,\xi}t}, \qquad u_x(x,t) = \frac{u'_{0,\xi}}{1 + c'_u(u_0(\xi))u'_{0,\xi}t},$$

where the denominator is always positive by the argument above. Hence,

$$u_t + c(u)u_x = 0$$

and we have indeed a unique solution defined everywhere for $t > 0$. ∎

The method of characteristic system, described in Remark 1.1 can be used also for nonlinear equations, as illustrated in the example below. The resulting system in the nonlinear case become coupled and, in many cases, rather impossible to solve.

Example 2.2 Consider the initial value problem

$$\begin{aligned} u_t + uu_x &= -u, \quad x \in \mathbb{R}, t > 0 \\ u(x,0) &= -\frac{x}{2}, \quad x \in \mathbb{R}. \end{aligned}$$

The characteristic system is

$$\begin{aligned} \frac{du}{dt} &= -u, \\ \frac{dx}{dt} &= u, \end{aligned}$$

with the initial data $u(0) = -\xi/2, x(0) = \xi$. The general solution of the system is

$$u(t) = ae^{-t}, \quad x(t) = b - ae^{-t}.$$

Using the initial conditions, we obtain

$$a = -\frac{\xi}{2}, \quad b = \frac{\xi}{2},$$

so that

$$u = -\frac{\xi}{2}e^{-t}, \quad x = \frac{\xi}{2}\left(1 + e^{-t}\right).$$

Dividing u by x, we eliminate ξ, getting

$$u(x,t) = -\frac{xe^{-t}}{1 + e^{-t}}.$$

This is a smooth solution defined for all $x \in \mathbb{R}$ and $t > 0$.

It should be stressed that the situation described in the previous two examples are far from being typical. Firstly, in most cases the implicit equation for ξ in (7.2.20) cannot be solved explicitly. Secondly, a solution may cease to exist after finite time, as shown in the example below.

Example 2.3 Consider the equation

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, t > 0,$$

with the initial condition

$$u_0(x) = \begin{cases} 2 & \text{for} \quad x < 0, \\ 2 - x & \text{for} \quad 0 \le x \le 1, \\ 1 & \text{for} \quad x > 1 \end{cases}$$

The characteristic system is

$$\begin{aligned} \frac{dv}{dt} &= 0, \\ \frac{dx}{dt} &= u, \end{aligned} \tag{7.2.21}$$

with the initial data $u(0) = u(\xi)$ and $x(0) = \xi$. From the first equation we obtain $v(t,\xi) = u(\xi)$; that is, u is constant along each characteristic, $u(x(t,\xi),t) = v(t,\xi) = u(\xi)$. But then, for a given ξ, the characteristic emanating from ξ satisfies

$$\frac{dx}{dt} = u(\xi), \qquad x(0) = \xi;$$

that is,

$$x = tu(\xi) + \xi.$$

In other words, the characteristics are straight lines, with the slope (x against t) equal to the initial value of the solution at the characteristic's intercept with the x axis, ξ and the solution is constant along each such line. To find the solution at a particular point (x,t), we see that the characteristic system (9.5.45) is equivalent to the system of algebraic equations

$$u(x,t) = u(\xi), \qquad x = tu(\xi) + \xi \tag{7.2.22}$$

and the solution u is obtained eliminating, if possible, the parameter ξ from (9.5.46).

To find the solution of our particular problem, we see that for $\xi < 0$ the characteristics lines are $t = (x-\xi)/2$ and for $\xi > 1$ the lines are $t = x - \xi$. For $0 \le \xi \le 1$ the equation of the characteristic is

$$t = \frac{1}{2-\xi}(x-\xi)$$

and we see that all these lines pass through the point $(x,t) = (2,1)$. This means that the solution $(2,1)$ cannot exist as it should be equal to the value carried by each characteristic, and each characteristic carries different value of the solution. Thus, the smooth solution cannot continue beyond $t = 1$. We call $t = 1$ the *breaking time* of the wave.

To find the solution for $t < 1$, we first note that $u(x,t) = 2$ for $x < 2t$ and $u(x,t) = 1$ for $x > t+1$. For $2t < x < t+1$, the second equation in (7.2.20) becomes

$$x - \xi = (2-\xi)t$$

that gives

$$\xi = \frac{x-2t}{1-t}$$

and the first equation gives then

$$u(x,t) = u_0(\xi) = 2 - \frac{x-2t}{1-t} = \frac{2-x}{1-t},$$

valid for $2t < x < t+1, t < 1$. The explicit form of the solution also indicates the difficulty at the breaking time $t = 1$.

The phenomenon observed in the above example is typical when the speed of propagation c is increasing and u_0 is increasing, or conversely, as in the above example. To explain this, let us concentrate on the initial value problem

$$\begin{aligned} u_t + c(u)u_x &= 0, \\ u(x,0) &= u_0(x) \end{aligned} \tag{7.2.23}$$

where $c(u) > 0$, $c'(u) > 0$, and u_0 is a differentiable function on $\mathbb{R}$. We have already seen that if $u_0' \ge 0$, then a smooth solution $u(x,t)$ exists for all $t > 0$ and is given implicitly by

$$u(x,t) = u_0(\xi), \qquad x - \xi = c(u_0(\xi))t.$$

Let, on the contrary, u_0 be such that for some $\xi_1 < \xi_2$ we have $u_0(\xi_1) > u_0(\xi_2)$. Then clearly also $c(u_0(\xi_1)) > c(u_0(\xi_2))$, that is the characteristic emanating from ξ_1 is faster (has greater speed) that that emanating from ξ_2. Therefore the characteristics cross at some point (x,t) which is a contradiction as the value $u(x,t)$ should be uniquely determined as $u_0(\xi_1)$ (or $u_0(\xi_2)$?).

As we observed in the previous example, at this point the gradient u_x becomes infinite. That is why we say that at this point a *gradient catastrophe* occurs. Certainly, along different characteristics the gradient catastrophe can occur in different times; along some it is possible that it will never occur. It is possible to determine the breaking time, when a gradient catastrophe occurs, even if we do not know the explicit form of the solution. To do this, let us calculate the gradient u_x along a characteristic that has the equation

$$x - \xi = c(u_0(\xi))t$$

. We assume that $c' > 0$ and $u_0'(\xi) < 0$ for this ξ. Let $g(t) = u_x(x(t), t)$ denotes the gradient of the solution along this characteristic. Then

$$\frac{dg}{dt} = u_{tx} + c(u)u_{xx}$$

and, on the other hand, differentiating the partial differential equation in (7.2.23) with respect to x, we find

$$u_{tx} + c'(u)(u_x)^2 + c(u)u_{xx} = 0,$$

thus we obtain

$$\frac{dg}{dt} = -c'(u)g^2,$$

along the characteristic. Along the characteristic $c'(u)$ is constant (as u is constant) and equal to $c'(u_0(\xi))$ and therefore this equation can be solved, giving

$$g(t) = \frac{g(0)}{1 + g(0)c'(u_0(\xi))t},$$

where $g(0)$ is the initial gradient at $t = 0$. But along the characteristic the initial gradient is $u_0'(\xi)$, thus

$$u_x = \frac{u_0'(\xi)}{1 + u_0'(\xi)c'(u_0(\xi))t}.$$

This is the same formula for that gradient that we obtained in the proof of Theorem 2.1 and clearly the finitness of the gradient is equivalent to the solvability of the implicit equation for ξ. In any case, as u_0' and c' have opposite sign, the product is negative and there always exist time t for which

$$tu_0'(\xi)c'(u_0(\xi)) = -1.$$

If the initial condition is decreasing on $\mathbb{R}$, then the gradient catastrophe will occur along any characteristic. To find when the wave brakes, we must find the characteristic along which the catastrophe occurs first. Since, for a given ξ the catastrophe time along this characteristic is given by

$$t(\xi) = -\frac{1}{u_0'(\xi)c'(u_0(\xi)}$$

and since the denominator is the derivative of

$$F(\xi) = c(u_0(\xi))$$

we conclude that the wave first breaks along the characteristic $\xi = \xi_b$ for which $F'(\xi) < 0$ attains minimum. Thus, the time of (the first) breaking is

$$t_b = -\frac{1}{F'(\xi_b)}.$$

The positive time t_b is called the *breaking time* of the wave.

We remark that is the initial function u_0 is not monotone, breaking will first occur on the characteristic $\xi = \xi_b$, for which $F'(\xi_b) < 0$ and $F'(\xi_b)$ is a minimum.

Example 2.4 Consider the problem

$$\begin{aligned} u_t + uu_x &= 0, \\ u(x, 0) &= e^{-x^2}. \end{aligned}$$

To determine the breaking time, we find

$$F(\xi) = e^{-\xi^2}$$

so that

$$F'(\xi) = -2\xi e^{-\xi^2}, \qquad F''(\xi) = -(4\xi^2 - 2)e^{-\xi^2}.$$

Thus, $F'(\xi)$ attains minimum at $\xi_b = \frac{1}{\sqrt{2}}$ and the breaking time is

$$t_b = -\frac{1}{F'(\xi_b)} = \frac{e^{1/2}}{\sqrt{2}} \approx 1.16.$$

Thus, the breaking will occur first along the characteristic emanating from $\xi_b = \frac{1}{\sqrt{2}}$ at $t_b \approx 1.16$.

Is there life after the break time?

We have observed that at the break time the solution should become multivalued as it is defined by many characteristics each carrying a different value. However, in most physical problems described by this theory, the solution is just the density of some medium and is inherently single-valued. Therefore when breaking occurs, then

$$u_t + c(u)u_x = 0 \tag{7.2.24}$$

must cease to be valid as a description of the physical problem. Even in cases such as water waves, where a multivalued solution for the height of the surface could at least be interpreted, it is still found that (7.2.24) is inadequate to describe the process. Thus the situation is that some assumption or approximate relation leading to (7.2.24) is no longer valid. In principle one must return to the physics of the problem, see what went wrong, and formulate an improved theory. However, it turns out, that fortunately the foregoing solution can be saved by allowing discontinuities into the solution: a single-valued solution with a simple jump discontinuity replaces the multivalued continuous solution. This requires some mathematical extension of what we mean by a "solution" to (7.2.24) since, strictly speaking, the derivatives of u do not exist at a discontinuity.

Let us recall the modelling process that led to (7.2.24). We started with the conservation law

$$\frac{d}{dt}\int_a^b u(x,t)dx = \phi(a,t) - \phi(b,t), \tag{7.2.25}$$

where ϕ was the flux, and under assumption that both u and ϕ are differentiable, we arrived at (7.2.24). Now, clearly u is not known beforehand and, as we have observed in several examples, it can be not differentiable. Thus, while for functions u and ϕ with jump discontinuities the conservation law (7.2.25) may have sense, one cannot derive (7.2.24) from it. Since in the modelling process (7.2.25) is more basic than (7.2.24), we will insist on the validity of the former without necessarily having (7.2.24).

Let us find out what type of discontinuities are allowed by (7.2.25). Assume that $x = s(t)$ is a smooth curve in space-time along which u suffers a simple discontinuity, that is, $u(x,t)$ is continuously differentiable for $x > s(t)$ and $x < s(t)$, and that u and its derivatives have one sided limits as $x \to s(t)^-$ and $x \to s(t)^+$, for any $t > 0$. If we chose a and b such that $a < s(t) < b$ then we can write (7.2.25) in the form

$$\frac{d}{dt}\int_a^{s(t)} u(x,t)dx + \frac{d}{dt}\int_{s(t)}^b u(x,t)dx = \phi(a,t) - \phi(b,t), \tag{7.2.26}$$

The Leibniz rule for differentiating an integral whose integrand and limits depend on a parameter can be applied as the integrands are differentiable. In this way we obtain

$$\int_a^{s(t)} u_t(x,t)dx + \int_{s(t)}^b u_t(x,t)dx + u(s(t)^-,t)\frac{ds}{dt} - u(s(t)^+,t)\frac{ds}{dt} = \phi(a,t) - \phi(b,t), \tag{7.2.27}$$

where

$$u(s^{\pm}(t),t) = \lim_{x \to s^{\pm}(t)} u(x,t).$$

and ds/dt is the speed of discontinuity $x = s(t)$. Passing with $a \to s(t)^-$ and $b \to s(t)^+$, we see that the integral terms on the left-hand side become zero as the integrands are bounded and the length of integration shrinks to zero. Thus, we obtain

$$-\frac{ds}{dt}[u] = [\phi(u)], \tag{7.2.28}$$

where the brackets denote the jump of the quantity inside across the discontinuity. Equation (7.2.28) is called the *jump condition* or *Rankine-Hugoniot condition*, and it relates the conditions ahead of the discontinuity and behind the discontinuity to the speed of the discontinuity itself. The discontinuity in u that propagates along the curve $x = s(t)$ is called a *shock wave*, the curve itself is called the *shock path*, ds/dt is the shock speed, and the magnitude of the jump $[u]$ is called the *shock strength*. We illustrate this discussion by continuing the solution of Example 2.3.

Example 2.5 Consider the conservation law

$$\frac{d}{dt}\int_a^b u(x,t)dx = \frac{1}{2}u^2(a,t) - \frac{1}{2}u^2(b,t)$$

that, for smooth solutions, reduces to

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, t > 0,$$

with the initial condition

$$u_0(x) = \begin{cases} 2 & \text{for} \quad x < 0, \\ 2 - x & \text{for} \quad 0 \le x \le 1, \\ 1 & \text{for} \quad x > 1 \end{cases}$$

Since characteristic cannot carry values of the solution across the shock, the only possible values behind the shock are $u^- = 2$ and in front of the shock are $u^+ = 1$, where we denoted $u^\pm = u(s^\pm(t), t)$. As $\phi(u) = \frac{1}{2}u^2$, we have

$$\frac{ds}{dt} = \frac{3}{2}$$

and the shock propagates as $x(t) = s(t) = \frac{3t}{2} + \frac{1}{2}$.

It must be remembered, however, that in most cases fitting the shock cannot be done explicitly as the values of u^+ and u^- are not known.

Chapter 8

Travelling waves

1 Introduction

One of the cornerstones in the study of both linear and nonlinear PDEs is the wave propagation. A *wave* is a recognizable signal which is transferred from one part of the medium to another part with a recognizable speed of propagation. Energy is often transferred as the wave propagates, but matter may not be. We mention here a few areas where wave propagation is of fundamental importance.

Fluid mechanics (water waves, aerodynamics)

Acoustics (sound waves in air and liquids)

Elasticity (stress waves, earthquakes)

Electromagnetic theory (optics, electromagnetic waves)

Biology (epizootic waves)

Chemistry (combustion and detonation waves)

The simplest form of a mathematical wave is a function of the form

$$u(x,t) = f(x - ct). \tag{8.1.1}$$

We adopt the convention that $c > 0$. We have already seen such a wave as a solution of the constant coefficient transport equation

$$u_t + cu_x = 0,$$

it can, however, appear in many other contexts.

At $t = 0$ the wave has the form $f(x)$ which is the initial wave profile. Then $f(x - ct)$ represents the profile at time t, that is just the initial profile translated to the right by ct spatial units. Thus the constant c represents the speed of the wave and thus, evidently, (8.1.1) represents a wave travelling to the right with speed $c > 0$. Similarly,

$$u(x,t) = f(x + ct)$$

represents a wave travelling to the right with speed c. These waves propagate undistorted along the lines $x \pm ct = const.$

One of the fundamental questions in the theory of nonlinear PDEs is whether a given PDE admit such a travelling wave as a solution. This question is generally asked without regard to initial conditions so that the wave is assumed to have existed for all times. However, boundary conditions of the form

$$u(-\infty, t) = constant, \qquad u(+\infty, t) = constant \tag{8.1.2}$$

are usually imposed. A *wavefront type-solution* to a PDE is a solution of the form $u(x,t) = f(x \pm ct)$ subject to the condition (8.1.2) of being constant at $\pm\infty$ (this constant are not necessarily the same); the function f is assumed to have the requisite degree of smoothness defined by the PDE. If u approaches the same constant at both $\pm\infty$, then the wavefront is called a *pulse*.

2 Examples

We have already seen that the transport equation with constant c admits the travelling wave solution and it is the only solution this equation can have. To illustrate the technique of looking for travelling wave solutions on a simple example first, let us consider the wave equation.

Example 2.1 Find travelling wave solutions to the wave equation

$$u_{tt} - a^2 u_{xx} = 0.$$

According to definition, travelling wave solution is of the form $u(x,t) = f(x - ct)$. Inserting this into the equation, we find

$$u_{tt} - a^2 u_{xx} = c^2 f'' - a^2 f'' = f'' \cdot (c^2 - a^2) = 0$$

so that either $f(s) = A + Bs$ for some constants A, B or $c = \pm a$ and f arbitrary. In the first case we would have

$$u(x,t) = A + B(x \pm ct)$$

but the boundary conditions (8.1.2) cannot be satisfied unless $B = 0$. Thus, the only travelling wave solution in this case is constant. For the other case, we see that clearly for any twice differentiable function f such that

$$\lim_{s \to \pm\infty} = d_{\pm\infty}$$

the solution

$$u(x,t) = f(x \pm at)$$

is a travelling wave solution (a pulse if $d_{+\infty} = d_{-\infty}$).

In general, it follows that any solution to the wave equation can be obtained as a superposition of two travelling waves: one to the right and one to the left

$$u(x,t) = f(x - at) + g(x + at).$$

Not all equations admit travelling wave solutions, as demonstrated below.

Example 2.2 Consider the diffusion equation

$$u_t = D u_{xx}.$$

Substituting the travelling wave formula $u(x,t) = f(x - ct)$, we obtain

$$-cf' - Df'' = 0$$

that has the general solution

$$f(s) = a + b \exp\left(-\frac{cs}{D}\right).$$

It is clear that for f to be constant at both plus and minus infinity it is necessary that $b = 0$. Thus, there are no nonconstant travelling wave solutions to the diffusion equation.

We have already seen that the non-viscid Burger's equation

$$u_t + uu_x = 0$$

does not admit a travelling wave solution: any profile will either smooth out or form a shock wave (which can be considered as a generalized travelling wave - it is not continuous!). However, some amount of dissipation represented by a diffusion term allows to avoid shocks.

Example 2.3 Consider Burgers' equation with viscosity

$$u_t + uu_x - \nu u_{xx} = 0, \qquad \nu > 0. \tag{8.2.3}$$

The term uu_x will have a shocking up effect that will cause waves to break and the term νu_{xx} is a diffusion like term. We attempt to find a travelling wave solution of (8.2.3) of the form

$$u(x,t) = f(x - ct).$$

Substituting this to (8.2.3) we obtain

$$-cf'(s) + f(s)f'(s) - \nu f''(s) = 0,$$

where $s = x - ct$. Noting that $ff' = \frac{1}{2}(f^2)'$ we re-write the above as

$$-cf' + \frac{1}{2}(f^2)' - \nu f'' = 0,$$

hence we can integrate getting

$$-cf + \frac{1}{2}f^2 - \nu f' = B$$

where B is a constant of integration. Hence

$$\frac{df}{ds} = \frac{1}{2\nu}\left(f^2 - 2cf - 2B\right). \tag{8.2.4}$$

Let us consider the case when the quadratic polynomial above factorizes into real linear factors, that is

$$\left(f^2 - 2cf - 2B\right) = (f - f_1)(f - f_2)$$

where

$$f_1 = c - \sqrt{c^2 + 2B}, \qquad f_2 = c + \sqrt{c^2 + 2B}. \tag{8.2.5}$$

This requires $c^2 > 2B$ and yields, in particular, $f_2 > f_1$. Eq. (8.2.4) can be easily solved by separating variables. First note that f_1 and f_2 are the only equilibrium points and $(f - f_1)(f - f_2) < 0$ for $f_1 < f < f_2$ so that any solution starting between f_1 and f_2 will stay there tending to f_1 as $s \to +\infty$ and to f_2 as $s \to -\infty$. Any solution starting above f_2 will tend to ∞ as $s \to +\infty$ and any one starting below f_1 will tend to $-\infty$ as $s \to -\infty$. Thus, the only non constant travelling wave solutions are possible for $f_1 < f < f_2$. For such f integration if (9.5.26) yields

$$\begin{aligned} \frac{s - s_0}{2\nu} &= \int \frac{df}{(f - f_1)(f - f_2)} = -\frac{1}{f_2 - f_1}\int \left(\frac{1}{f - f_1} + \frac{1}{f_2 - f}\right) df \\ &= \frac{1}{f_2 - f_1} \ln \frac{f_2 - f}{f - f_1}. \end{aligned}$$

Solving for f yields

$$f(s) = \frac{f_2 + f_1 e^{K(s - s_0)}}{1 + e^{K(s - s_0)}}, \tag{8.2.6}$$

where $K = \frac{1}{2\nu}(f_2 - f_1) > 0$. We see that for large positive s $f(s) \sim f_1$ whereas for negative values of s we obtain asymptotically $f(s) \sim f_2$. It is clear that the initial value s_0 is not essential so that we shall suppress it in the sequel. The derivative of f is

$$f'(s) = K\frac{e^{Ks}(f_1 - f_2)}{(1 + e^{Ks})^2} < 0.$$

It is easy to see that for large $|s|$ the derivative $f(s)$ is close to zero so that f is almost flat. Moreover, the larger ν (so that the smaller K), the more flat is f as the derivative is closer to zero. Hence, for small ν we obtain a very steep wave front that is consistent with the fact for $\nu = 0$ we obtain inviscid Burger's equation that admits only discontinuous travelling waves.

The formula for travelling wave solution to (8.2.3) is then

$$u(x,t) = \frac{f_2 + f_1 e^{K(x-ct)}}{1 + e^{K(x-ct)}}, \tag{8.2.7}$$

where the speed of the wave is determined from (8.2.5) as

$$c = \frac{1}{2}(f_1 + f_2).$$

Graphically the travelling wave solution is the profile f moving to the right at speed c. This solution, because it resembles the actual profile of a shock wave, is called the *shock structure* solution; it joints the asymptotic states f_1 and f_2. Without the term νu_{xx} the solutions of (8.2.3) would shock up and tend to break. The presence of the diffusion term prevents this breaking effect by countering the nonlinearity. The result is competition and balance between the nonlinear term uu_x and the diffusion term $-\nu u_{xx}$, much the same as occurs in a real shock wave in the narrow region where the gradient is steep. In this context the $-\nu u_{xx}$ term could be interpreted as a viscosity term.

The last example related to travelling waves is concerned with the so called solitons that appear in solutions of numerous important partial differential equations. The simples equation producing solitons is the Korteweg-deVries equation that governs long waves in shallow water.

Example 2.4 Find travelling wave solutions of the KdV equation

$$u_t + uu_x + ku_{xxx} = 0, \tag{8.2.8}$$

where $k > 0$ is a constant. As before, we are looking for a solution of the form

$$u(x,t) = f(s), \qquad s = x - st,$$

where the waveform f and the wave speed c are to be determined. Substituting f into (8.2.8) we get

$$-cf' + \frac{1}{2}(f^2)' + kf''' = 0,$$

integration of which gives

$$-cf + \frac{1}{2}f^2 + kf'' = a$$

for some constant a. This is a second order equation that does not contain the independent variable, hence we can use substitution $f' = F(f)$, discussed in Subsection 3.4. Hence, $f''_s = F'_f f'_s = F'_f F$ and we can write

$$-cf + \frac{1}{2}f^2 + kF'_f F = -cf + \frac{1}{2}f^2 + \frac{k}{2}(F_f^2)' = a.$$

Integrating, we get

$$F^2 = \frac{1}{k}\left(cf^2 - \frac{1}{3}f^3 + 2af + 2b\right)$$

where b is a constant. Thus,

$$f' = \pm\sqrt{\frac{1}{3k}}\left(-f^3 + 3cf^2 + 6af + 6b\right)^{1/2}. \tag{8.2.9}$$

To fix attention we shall take the "+" sign. Denote by $\phi(f)$ the cubic polynomial on the right-hand side. We have the following possibilities:

(i) ϕ has one real root α;

(ii) ϕ has three distinct real roots $\gamma < \beta < \alpha$;

(iii) ϕ has three real roots satisfying $\gamma = \beta < \alpha$;

(iv) ϕ has three real roots satisfying $\gamma < \beta = \alpha$;

(v) ϕ has a triple root γ.

Since we are looking for travelling wave solutions that should be bounded at $\pm\infty$ and nonnegative, we can rule out most of the cases by qualitative analysis. Note first that the right-hand side of (8.2.9) is defined only where $\phi(f) > 0$. Then, in the case (i), α is a unique equilibrium point, $\phi > 0$ only for $f < \alpha$ and hence any solution converges to α as $t \to \infty$ and diverges to $+\infty$ as $t \to -\infty$. Similar argument rules out case (v).

If we have three distinct roots, then by the same argument, the only bounded solutions can exist in the cases (iii) and (ii) (in the case (iv) the bounded solutions could only appear where $\phi < 0$.) The case (ii) leads to the so-called *cnoidal* waves expressible through special functions called cn-functions and hence the name. We shall concentrate on case (iii) so that

$$\phi(f) = -f^3 + 3cf^2 + 6af + 6b = (\gamma - f)^2(\beta - f)$$

and, since $f > \gamma$

$$\sqrt{\phi(f)} = (f - \gamma)\sqrt{\alpha - f}.$$

Thus, the differential equation (8.2.9) can be written

$$\frac{s}{\sqrt{3k}} = \int \frac{df}{(f-\gamma)\sqrt{\alpha - f}}. \tag{8.2.10}$$

To integrate, we first denote $v = f - \gamma$ and $B = \alpha - \gamma$ (with $0 < v < b$), getting

$$\frac{s}{\sqrt{3k}} = \int \frac{dv}{v\sqrt{B - v}}.$$

Next, we substitute $w = \sqrt{B - v}$, hence $v = B - w^2$ and $dw = -ds/2\sqrt{B - v}$ so that the above will be transformed to

$$\frac{s}{\sqrt{3k}} = 2\int \frac{dw}{w^2 - B}.$$

This integral can be evaluated by partial fractions, so that, using $0 < w < \sqrt{B}$, we get

$$\frac{s}{\sqrt{3k}} = \frac{1}{\sqrt{B}}\int\left(\frac{dw}{w - \sqrt{B}} - \frac{dw}{w + \sqrt{B}}\right)dw = \frac{1}{\sqrt{B}}\ln\frac{\sqrt{B} - w}{\sqrt{B} + w}.$$

Solving with respect to w we obtain

$$\frac{\sqrt{B} - w}{\sqrt{B} + w} = \exp s\sqrt{\frac{B}{3k}}$$

and thus

$$w = \sqrt{B}\,\frac{1 - \exp s\sqrt{\frac{B}{3k}}}{1 + \exp s\sqrt{\frac{B}{3k}}} = -\sqrt{B}\,\frac{\sinh s\sqrt{\frac{B}{12k}}}{\cosh s\sqrt{\frac{B}{12k}}}.$$

Returning to the original variables $w^2 = B - v = \alpha - \gamma - f + \gamma = \alpha - f$ and using the hyperbolic identity $\cosh^2\theta - \sinh^2\theta = 1$, we get

$$\begin{aligned} f(s) &= \alpha - w^2(s) = \alpha - (\alpha-\gamma)\frac{\sinh^2 s\sqrt{\frac{\alpha-\gamma}{12k}}}{\cosh^2 s\sqrt{\frac{\alpha-\gamma}{12k}}} \\ &= \gamma + (\alpha-\gamma)\mathrm{sech}^2\left(s\sqrt{\frac{\alpha-\gamma}{12k}}\right) \end{aligned}$$

Clearly, $f(s) \to \gamma$ as $s \to \pm\infty$ so that the travelling wave here is a pulse. It is instructive to write the roots α and γ in terms of the original parameters. To identify them we observe

$$\begin{aligned} \phi(f) &= -f^3 + 3cf^2 + 6af + 6b = (\gamma - f)^2(\beta - f) \\ &= -f^3 + f^2(\alpha + 2\gamma) + f(-\gamma^2 - 2\alpha\gamma) + \gamma^2\alpha. \end{aligned}$$

Thus, the wave speed is given by

$$c = \frac{\alpha + 2\gamma}{3} = \frac{\alpha - \gamma}{3} + \gamma.$$

Since γ is just the level of the wave at $\pm\infty$, by moving the coordinate system we can make it equal to zero. In this case we can write the travelling wave solution to the KdV equation as

$$u(x,t) = 3c\,\mathrm{sech}^2\left(\sqrt{\frac{\sqrt{c}}{4k}}(x - ct)\right) \tag{8.2.11}$$

It is important to note that the velocity of this wave is proportional to its amplitude which makes it different from linear waves governed by, say the wave equation, where the wave velocity is the property of the medium rather than of the wave itself.

3 The Fisher equation

Many natural processes involve mechanisms of both diffusion and reaction, and such problems are often modelled by so-called *reaction-diffusion equations* of the form

$$u_t - Du_{xx} = f(u), \tag{8.3.12}$$

where f is a given, usually nonlinear function of u. We introduced earlier the Fisher equation

$$u_t - Du_{xx} = ru\left(1 - \frac{u}{K}\right) \tag{8.3.13}$$

to model the diffusion of a species (e.g. insect population) when the reaction (or, at this instance) growth term is given by the logistic law. Here D is the diffusion constant, and r and K are the growth rate and carrying capacity, respectively.

We shall examine the Fisher equation and, in particular, we shall address the question of existence of a travelling wave solution.

Let us consider the Fisher equation in dimensionless form

$$u_t - u_{xx} = u(1 - u) \tag{8.3.14}$$

and, as before, we shall look for solutions of the form

$$u(x,t) = U(s), \qquad s = x - ct, \tag{8.3.15}$$

where c is a positive constant and U has the property that it approaches constant values at $s \to \pm\infty$. The function U to be determined should be twice differentiable. The wave speed c is a priori unknown and must be determined as a part of the solution of the problem. Substituting (8.3.15) into (8.3.14) yields a second order ordinary differential equation for U:

$$-cU' - U'' = U(1-U). \tag{8.3.16}$$

Contrary to the previous cases, this equation cannot be solved in a closed form and the best approach to analyze it is to perform the phase plane analysis. In a standard way we write (8.3.16) as a simultaneous system of first order equations by defining $V = U'$. In this way we obtain

$$\begin{aligned} U' &= V, \\ V' &= -cV - U(1-U). \end{aligned} \tag{8.3.17}$$

We find equilibrium points of this system solving

$$\begin{aligned} 0 &= V, \\ 0 &= -cV - U(1-U), \end{aligned}$$

which gives two: $(0,0)$ and $(0,1)$. The Jacobi matrix of the system is

$$J(U,V) = \begin{pmatrix} 0 & 1 \\ 2U-1 & -c \end{pmatrix}$$

so that

$$J(0,0) = \begin{pmatrix} 0 & 1 \\ -1 & -c \end{pmatrix}$$

with eigenvalues

$$\lambda_{\pm}^{0,0} = \frac{-c \pm \sqrt{c^2-4}}{2}$$

and

$$J(1,0) = \begin{pmatrix} 0 & 1 \\ 1 & -c \end{pmatrix}$$

with eigenvalues

$$\lambda_{\pm}^{1,0} = \frac{-c \pm \sqrt{c^2+4}}{2}.$$

It is easily seen that for any c, $\lambda_{\pm}^{1,0}$ are real and of opposite sign and therefore $(1,0)$ is a saddle. On the other hand, $\lambda_{\pm}^{0,0}$ are both real and negative if $c \geq 2$ and in this case $(0,0)$ is a stable node (for the linearized system), and are complex with negative real part if $0 < c < 2$ in which case $(0,0)$ is a stable focus.

Since the wave profile $U(s)$ must have finite limits as $s \to \pm\infty$ and since we know that the only limit points of solutions of autonomous systems are equilibrium points, search for travelling wave solutions of (8.3.16) is equivalent to looking for orbits of (8.3.17) joining equilibria, that is approaching them as $s \to \pm\infty$. Such orbits are called *heteroclinic* if they join different equilibrium points and *homoclinic* if the orbit returns to the same equilibrium point from which it started.

We shall use the Stable Manifold Theorem. According to it, there are two orbits giving rise, together with the equilibrium point $(1,0)$, to the unstable manifold defined at least in some neighbourhood of the saddle point $(1,0)$, such that each orbit $\phi(s) = (U(s), V(s))$ satisfies: $\phi(s)$ $to (1,0)$ as $s \to -\infty$. Our aim is then to show that at least one of these orbits can be continued till $s \to \infty$ and reaches then $(0,0)$ in a monotonic way. Let us consider the orbit that moves into the fourth quadrant $U > 0, V < 0$. This quadrant is divided into to regions by the isocline $0 = V' = -cV - U(1-U)$: region I with $\frac{1}{c}(U^2 - U) < V < 0$ and region II where $V < \frac{1}{c}(U^2 - U)$, see Fig. 5. 1. In region I we have $V' < 0, U' < 0$ and in region II we have $V' > 0, U' < 0$. From the earlier theory we know that if a solution is not defined for all values of the argument s then it must blow up as $s \to s'$ where $(-\infty, s')$ is the maximal interval of existence of the solution. We note first that the

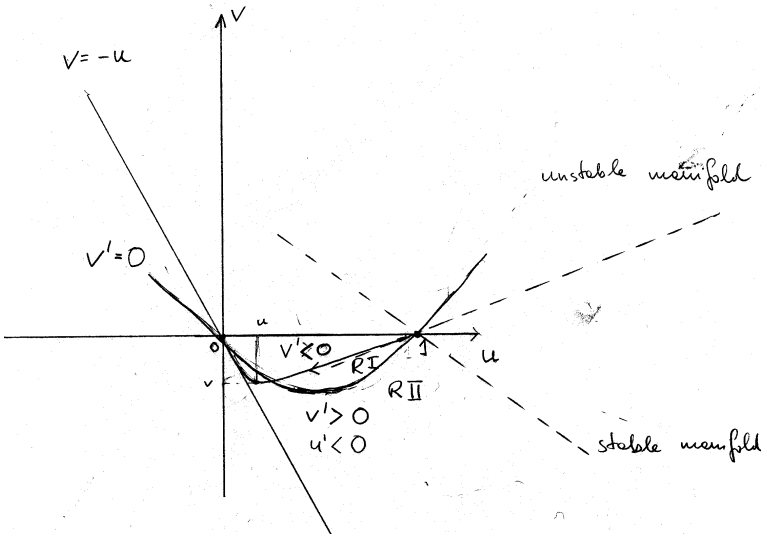

Fig. 5.1 Phase portrait of system (8.3.17).

selected orbit on our unstable manifold enters Region I so that $(U(s), V(s)) \in$ Region I for $-\infty < s \leq s_0$ for some s_0. In fact, the tangent of the isocline at $(1, 0)$ is $1/c$ and the tangent of the unstable manifold is $\lambda_+^{1,0} = \frac{\sqrt{c^2+4}-c}{2}$. Denoting

$$\psi(c) = \frac{c(\sqrt{c^2+4}-c)}{2}$$

we see that $\psi(0) = 0$, $\lim_{c\to+\infty} \psi(s) = 1$ and

$$\psi'(c) = \frac{4}{\sqrt{c^2+4}(\sqrt{c^2+4}+c)^2} > 0,$$

thus $0 \leq \psi(c) \leq 1$ for all $c \geq 0$ and hence

$$\frac{1}{c} \geq \frac{\sqrt{c^2+4}-c}{2},$$

so that the slope of the isocline is larger than that of the orbit and the orbit must enter Region 1. Then, since $V' < 0$ there, $V(s) < V(s_0)$ as long as $V(s)$ stays in Region I. Hence, the orbit must leave Region I in finite time as there is no equilibrium point with strictly negative V coordinate. Of course, there cannot be any blow up as long as the orbit stays in Region I. However, as at the crossing point the sign of V' changes, this point is a local minimum for V so that the orbit starts moving up and continues to move left.

The slope of $V = \frac{1}{c}U(1-U)$ at the origin is $-1/c$ so for $c \geq 1$ (and in particular for $c \geq 2$) the parabola $V = \frac{1}{c}U(1-U)$ stays above the line $V = -U$ so that any orbit entering Region II from Region I must stay for some time above $V = -U$, that is, $V/U > -1$. Consider any point $(U, -U)$, $U > 0$, and estimate the slope of the vector field on the line, see Fig. 5.2, We have

$$\tan\phi = \frac{-cV - U(1-U)}{-V} = c + \frac{U}{V}(1-U) = c - 1 + U > c - 1$$

Considering the direction, we see that if $c \geq 2$, then the vector field points to the left from the line $U = -V$. Hence, the whole trajectory must stay between $U = -V$ and $V = \frac{1}{c}U(1-U)$. Hence, the orbit is bounded and therefore exists for all s and enters $(0, 0)$ in a monotonic way, that is U is decreasing monotonically from 1 at $s = -\infty$ to 0 at $s = +\infty$ while $U' = V$ is non-positive and goes from zero at $s = -\infty$ through minimum back to 0 at $s = +\infty$.

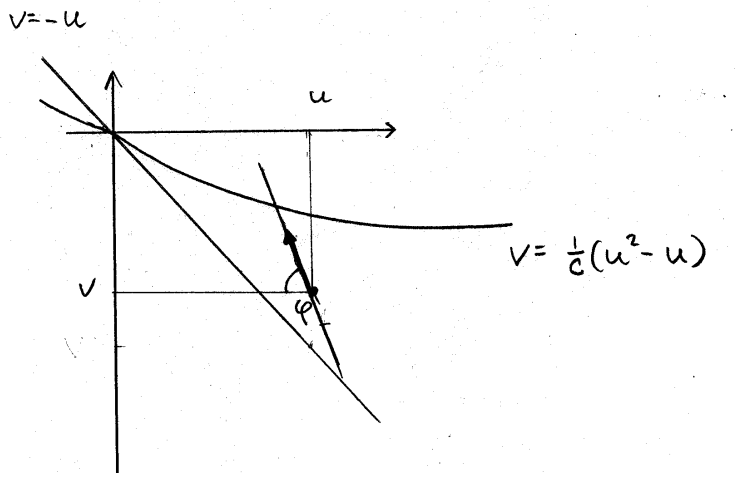

Fig. 5.2. The orbit in Region II.

Thus, summarizing, the orbit $(U(s), V(s))$ is globally defined for $-\infty < s < +\infty$ joining the equilibrium points $(1,0)$ and $(0,0)$. Thus, $U(s) \to 1$ as $s \to -\infty$ and $U(s) \to 0$ as $s \to \infty$. Moreover, as $0 > U'(s) = V(s) \to 0$ as $s \to \pm\infty$, U is monotonically decreasing and becomes flat at both "infinities" giving a travelling wavefront solution.

We note that for $c < 2$ the orbit no longer enters $(0,0)$ monotonically but, as suggested by linearization, spirals into the equilibrium point with U passing through positive and negative values. Thus, if we are interested only in positive values of U in order to have a physically realistic solution (e.g. if U is a population density), we should reject this case.

3.1 An explicit travelling wave solution to the Fisher equation

In many cases by postulating a special form of the travelling wave, one can obtain explicit solutions having this particular form. An important example in the Fisher equation case are solutions of the form

$$U_d(z) = \frac{1}{(1+ae^{bz})^d}, \qquad z = x - ct, \tag{8.3.18}$$

where constants a, b, d are to be determined. To determine these constants, we substitute (8.3.18) to (8.3.16)

$$-cU' - U'' = U(1-U).$$

First let us evaluate the necessary derivatives.

$$\begin{aligned} U_d' &= -dbae^{bz}(1+ae^{bz})^{-d-1} = -db(1+ae^{bz}-1)(1+ae^{bz})^{-d-1} \\ &= -db\left((1+ae^{bz})^{-d} - (1+ae^{bz})^{-d-1}\right) = -db(U_d - U_{d+1}) \end{aligned}$$

and similarly

$$U_d'' = db^2(dU_d - (2d+1)U_{d+1} + (d+1)U_{d+2}).$$

On the other hand

$$U(1-U) = U_d - U_{2d}$$

so that

$$cdb(U_d - U_{d+1}) - db^2(dU_d - (2d+1)U_{d+1} + (d+1)U_{d+2}) = U_d - U_{2d}$$

Since $U_d = (U_1)^d$ and $y = U_1(z)$ is a one-to-one mapping of $(-\infty, \infty)$ onto $(0,1)$, the above equation is equivalent to the polynomial equation

$$y^{2d} - d(d+1)b^2y^{d+2} + (d(2d+1)b^2 - cdb)y^{d+1} + (cdb - (db)^2 - 1)y^d = 0, \qquad y \in (0,1).$$

Since a polynomial is identically equal zero on an open interval if and only if all coefficients are 0, we obtain three possible cases: $2d = d+2$, $2d = d+1$ or $2d = d$. The last one gives $d = 0$ and thus a constant solution. The second results in $d = 1$ and thus $d(d+1)b = 0$, yielding $b = 0$ and again we obtain a constant solution. Let us consider the first case. Then $d = 2$ and thus, equating to zero coefficients of like powers

$$\begin{aligned} 1 - 6b^2 &= 0, \\ 2b(5b - c) &= 0, \\ 2cb - 4b^2 - 1 &= 0, \end{aligned}$$

we immediately obtain $b = \pm 1/\sqrt{6}$, $c = \pm 5/\sqrt{6}$ and it is easy to see that the last equation is automatically satisfied. Hence, we obtained a family of travelling wave of the Fisher equation

$$u(x,t) = \frac{1}{(1 + ae^{\pm \frac{\sqrt{6}x \mp 5t}{6}})^2}.$$

The arbitrary parameter $a > 0$ determines how steep is the wave.

4 The Nagumo equation

The Fisher equation describes spreading of a population which locally reproduces according to the logistic law. More sophisticated reproduction laws include the so-called Allee effect describing the situation in which small populations die out. Mathematically the Allee effect is represented by a term with three equilibria $0 < L < K$, where K is the carrying capacity of the environment and L is the threshold below which the population perishes. In the normalized case we consider the equation

$$u_t = u_{xx} + u(u - a)(1 - u) \tag{8.4.19}$$

where $0 < a < 1$. u is normalized density of the population and we assume that $0 < u < 1$. This equation is called the Nagumo equation. We look for some explicit travelling wave solutions to this equation. As before, we introduce $u(x,t) = U(x - ct)$, $0 < U < 1$, which converts (8.4.19) to

$$-cU'_z = U''_{zz} + U(U - a)(1 - U), \qquad z = x - ct.$$

As long as $U'_z \neq 0$, we can reduce this equation to the first order equation

$$(\Psi^2)'_U = -2c\Psi - 2U^3 - 2(1 + a)U^2 + 2aU$$

where $\Psi(U) = U'_z$ and $U''_{zz} = \Psi'_U \Psi$. If we try to find polynomial (in U) solutions to the above equation, then we see that the lowest order polynomial which could provide a solution is quadratic. Consider thus

$$\Psi(U) = \gamma + \beta U + \alpha U^2.$$

Then

$$(\Psi^2(U))' = 2\gamma\beta + 2(2\gamma\alpha + \beta^2)U + 6\alpha\beta U^3 + 4\alpha^2 U^3$$

and, comparing like powers of U, we obtain

$$\begin{aligned} 2\gamma\beta &= -2c\gamma, \\ 2(2\gamma\alpha + \beta^2) &= -2c\beta + 2a, \\ 6\alpha\beta &= -2c\alpha - 2(1 + a), \\ 4\alpha^2 &= 2. \end{aligned}$$

From the last equation $\alpha = \pm 1/\sqrt{2}$. We shall work with the plus sign to keep the notation simpler. Thus

$$\gamma(\beta + c) = 0, \tag{8.4.20}$$

$$\sqrt{2}\gamma + \beta^2 = -c\beta + a, \tag{8.4.21}$$

$$3\beta = -c - \sqrt{2}(1 + a). \tag{8.4.22}$$

From the last equation $c = -3\beta - \sqrt{2}(1+a)$ and the first equation becomes

$$\gamma(2\beta + \sqrt{2}(1+a)) = 0.$$

Now, we have to distinguish two cases.

Case 1. $\gamma = 0$.
In this case the second equation gives

$$4\beta^2 + 2\sqrt{2}(1+a) + 2a = 0$$

and we obtain solutions

$$\beta_{1,2} = -\frac{a}{\sqrt{a}}, -\frac{1}{\sqrt{2}}.$$

In the first case

$$\Psi(U) = \frac{1}{\sqrt{2}}U(U-a)$$

so that $\Psi(U) = U'_z$ becomes 0 for $U = a$. It means that there is a possibility that the transformation reducing the second order equation to the first order maybe not invertible at some point and thus we discard this solution. The second case gives

$$\Psi(U) = \frac{1}{\sqrt{2}}U(U-1)$$

and $\Psi(U) \neq 0$ in the range where U is allowed to change. So, we shall use this solution with

$$c = \sqrt{2}\left(\frac{1}{2} - a\right). \tag{8.4.23}$$

Case 2. $\gamma \neq 0$.
In this case

$$-c = \beta = -\frac{1}{\sqrt{2}}(1+a)$$

and, from the second equation in (8.4.22), $\gamma = a/\sqrt{2}$. Thus

$$\Psi(U) = \frac{a}{\sqrt{2}} - \frac{1}{\sqrt{2}}(1+a)U + \frac{1}{\sqrt{2}}U^2 = \frac{1}{\sqrt{2}}(a-U)(1-U)$$

and, again, we see that U' can vanish for some $U \in (0,1)$. Hence, again we rule this case out.

Thus, we are left with the equation

$$U' = \frac{1}{\sqrt{2}}U(U-1).$$

This is a separable equation which we solve by partial fractions. For $U \in (0,1)$

$$K + \frac{z}{\sqrt{2}} = \int \frac{dU}{U} + \int \frac{dU}{1-U} = \ln\frac{U}{1-U}$$

for some constant K, that is,

$$U(z) = \frac{\bar{K}e^{\frac{z}{\sqrt{2}}}}{1 + \bar{K}e^{\frac{z}{\sqrt{2}}}}$$

and we obtain a travelling wave in the form

$$u(x,t) = \frac{\bar{K}e^{\frac{x-ct}{\sqrt{2}}}}{1 + \bar{K}e^{\frac{x-ct}{\sqrt{2}}}}.$$

Chapter 9

Similarity methods for linear and non-linear diffusion

1 Similarity method

The method described in this section can be applied to equations of arbitrary order. However, having in mind concrete application we shall focus on the general second order partial differential equation in two independent variables in the form

$$G(t, x, u, u_x, u_t, u_{xx}, u_{xt}, u_{xx}) = 0. \tag{9.1.1}$$

To shorten notation we introduce notation

$$p = u_x, \quad q = u_t, \quad r = u_{xx}, \quad s = u_{xt}, \quad v = u_{tt}.$$

We introduce the *one-parameter family of stretching transformations*, denoted by T_ϵ, by

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^b t, \quad \bar{u} = \epsilon^c u, \tag{9.1.2}$$

where a, b, c are real constants and ϵ is a real parameter restricted to some open interval I containing $\epsilon = 1$. We note that (9.1.2) induces a transformation of the derivatives in the following way:

$$\bar{p} = \frac{\partial \bar{u}}{\partial \bar{x}} = \epsilon^c \frac{\partial u}{\partial x} \frac{dx}{d\bar{x}} = \epsilon^{c-a} p \tag{9.1.3}$$

and similarly for other derivatives

$$\bar{q} = \epsilon^{c-b} q, \quad , \bar{r} = \epsilon^{c-2a} r, \quad , \bar{s} = \epsilon^{c-a-b} s, \quad \bar{v} = \epsilon^{c-2b} \bar{v}. \tag{9.1.4}$$

Further, we say that PDE (9.1.1) is invariant under the one parameter family T_ϵ of stretching transformations if there exists a smooth function $f(\epsilon)$ such that

$$G(\bar{t}, \bar{x}, \bar{u}, \bar{p}, \bar{q}, \bar{r}, \bar{s}, \bar{v}) = f(\epsilon) G(t, x, u, p, q, r, s, v) \tag{9.1.5}$$

for all $\epsilon \in I$, with $f(1) = 1$. If $f(\epsilon) \equiv 1$ for all $\epsilon \in I$, then the PDE (9.1.1) is said to be *absolutely invariant.*

We shall formulate and prove the fundamental theorem.

Theorem 1.1 *If the equation (9.1.1) is invariant under the family T_ϵ defined by (9.1.2), then the transformation*

$$u = t^{c/b} y(z), \qquad z = \frac{x}{t^{a/b}} \tag{9.1.6}$$

reduces (9.1.1) to a second order ordinary differential equation in $y(z)$.

Proof. By invariance, we know that (9.1.5) holds for all ϵ in some open interval containing 1, thus we can differentiate (9.1.5) and set $\epsilon = 1$ after differentiation, getting

$$btG_t + axG_x + cuG_u + (c-a)pG_p + (c-b)qG_q + (c-2a)rG_r + (c-a-b)sG_s + (c-2b)vG_v = f'(1)G$$

where we used formulae like

$$\left.\frac{d\bar{x}}{d\epsilon}\right|_{\epsilon=1} = a\epsilon^{a-1}x\big|_{\epsilon=1} = ax,$$

etc. The above equation is a first order equation so that we can integrate it using t as the parameter along characteristics. The characteristic system will be then

$$\begin{aligned}
\frac{dF}{dt} &= \frac{f'(1)F}{bt},\\
\frac{dx}{dt} &= \frac{ax}{bt},\\
\frac{du}{dt} &= \frac{cu}{bt},\\
\frac{dp}{dt} &= \frac{(c-a)p}{bt},\\
\frac{dq}{dt} &= \frac{(c-b)q}{bt},\\
\frac{dr}{dt} &= \frac{(c-2a)r}{bt},\\
\frac{ds}{dt} &= \frac{(c-a-b)s}{bt},\\
\frac{dv}{dt} &= \frac{(c-2b)v}{bt},
\end{aligned}$$

where F is the function G written in the characteristic coordinates. Thus, we obtain characteristics defined by

$$\begin{aligned}
xt^{-a/b} &= z,\\
ut^{-c/b} &= \xi_1, \quad pt^{-(c-a)/b} = \xi_2, \quad qt^{-(c-b)/b} = \xi_3,\\
rt^{-(c-2a)/b} &= \xi_4, \quad st^{-(c-a-b)/b} = \xi_5, \quad vt^{-(c-2b)/b} = \xi_6
\end{aligned}$$

and

$$F = t^{\frac{f'(1)}{b}}\Psi(z, \xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6) \tag{9.1.7}$$

where Ψ is an arbitrary function. Now, we have $y = ut^{-c/b} = \xi_1$, $p = u_x = t^{c/b}y'_z z'_x = y'_z t^{(c-a)/b}$, hence $\xi_2 = y'_z$. Further,

$$q = u_t = \frac{c}{b}t^{-1+c/b}y - \frac{a}{b}t^{c/b-a/b-1}xy'_z,$$

thus

$$\xi_3 = qt^{1-c/b} = \frac{c}{b}y - \frac{a}{b}zy'_z.$$

Further,

$$r = u_{xx} = p_x = y''_{zz}t^{(c-a)/b}z'_x = y''_{zz}t^{(c-2a)/b},$$

hence $\xi_4 = y''_{zz}$. Similarly,

$$s = u_{tx} = q_x = \frac{c}{b}t^{-1+c/b}y'_z z'_x - \frac{a}{b}t^{c/b-a/b-1}y'_z - \frac{a}{b}t^{c/b-a/b-1}y''_{zz}z'_x = t^{c/b-a/b-1}\left(\frac{c-a}{b}y'_z - \frac{a}{b}y''_{zz}z\right)$$

giving $\xi_5 = \frac{c-a}{b}y'_z - \frac{a}{b}y''_{zz}z$ and finally

$$v = q_t = \frac{c}{b}\left(\frac{c}{b} - 1\right)t^{-2+c/b}y + \frac{c}{b}t^{-1+c/b}y'_z z'_t$$

$$
\begin{aligned}
& -\frac{a}{b}\left(\left(\frac{c}{b}-1\right)t^{-2+c/b}zy'_z + t^{-1+c/b}z'_t y'_z + t^{-1+c/b}zy''_{zz}z'_t\right) \\
= \;& t^{-2+c/b}\left(\frac{c}{b}\left(\left(\frac{c}{b}-1\right)y - \frac{a}{b}zy'_z\right) - \frac{a}{b}\left(\left(\frac{c}{b}-1\right)zy'_z - \frac{a}{b}zy'_z - \frac{a}{b}z^2y''_{zz}\right)\right) \\
= \;& t^{-2+c/b}\left(\frac{c}{b}\left(\frac{c}{b}-1\right)y - 2\frac{ac}{b^2}zy'_z + \frac{a}{b}zy'_z + \frac{a^2}{b^2}zy'_z + \frac{a^2}{b^2}z^2y''_{zz}\right)
\end{aligned}
$$

so that

$$\xi_6 = \frac{c}{b}\left(\frac{c}{b}-1\right)y - \frac{a}{b}\left(2\frac{c}{b}-1-\frac{a}{b}\right)zy'_z + \frac{a^2}{b^2}z^2y''_{zz}$$

Hence, combining (9.1.7) with the original equation (9.1.1) we obtain

$$\Psi\left(z, y, y', \frac{c}{b}y - \frac{a}{b}zy'_z, y''_{zz}, \frac{c-a}{b}y'_z - \frac{a}{b}y''_{zz}z, \frac{c}{b}\left(\frac{c}{b}-1\right)y - \frac{a}{b}\left(2\frac{c}{b}-1-\frac{a}{b}\right)zy'_z + \frac{a^2}{b^2}z^2y''_{zz}\right) = 0$$

which is a second order ordinary differential equation in z. ■

2 Linear diffusion equation

Consider the diffusion equation

$$u_t - Du_{xx}. \tag{9.2.1}$$

We shall try to find a stretching transformation under which this equation is invariant. Using our simplified notation for derivatives we have

$$\bar{q} - D\bar{r} = \epsilon^{c-b}q - D\epsilon^{c-2a}r.$$

We achieve invariance if ϵ is risen to the same power. Thus, we must have

$$b = 2a$$

with c and a at this moment arbitrary. Thus, (9.2.1) is invariant under the stretching transformation

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^{2a}t, \quad \bar{u} = \epsilon^c u \tag{9.2.2}$$

and the similarity transformation is given by

$$u = t^{c/2a}y(z), \quad z = \frac{x}{\sqrt{t}}. \tag{9.2.3}$$

We have $z'_x = \frac{1}{\sqrt{t}}$, $z'_t = -\frac{1}{2}xt^{-3/2} = -\frac{1}{2}zt^{-1}$, hence

$$u_t = \frac{c}{2a}t^{-1+c/2a}y + t^{c/2a}y'_z z'_t = t^{-1+c/2a}\left(\frac{c}{2a}y - \frac{z}{2}y'_z\right)$$

and

$$u_x = t^{c/2a-1/2}y'_z, \qquad u_{xx} = t^{c/2a-1}y''_{zz}.$$

Substituting the above relations into the diffusion equation yields

$$Dy'' + \frac{z}{2}y' - \frac{c}{2a}y = 0. \tag{9.2.4}$$

Constants c and a are, in general, arbitrary.

An important comment is in place here. Though the diffusion equation has been reduced to an ordinary differential equation, one should not think that any problem related to the diffusion equation is reducible to an ODE problem so the similarity approach by no means solves all PDE problems. In fact, an inherent part

of a diffusion problem are initial and boundary conditions and these, in general, cannot be translated into side conditions for (9.2.4). For instance, consider the initial value problem for the diffusion equation

$$\begin{aligned} u_t &= Du_{xx}, \quad t > 0, -\infty < x < \infty, \\ u(x,0) &= u_0(x). \end{aligned} \tag{9.2.5}$$

Using the similarity transformation, we can convert the equation into an ODE for f defined as $u(x,t) = t^{c/2a} y(x/\sqrt{t})$ but then putting $t = 0$ in the preceding formula in general does not make any sense as, at best, we would have something like

$$\begin{aligned} y(\infty) &= \lim_{t\to 0^+} t^{-c/2a} u(x,t), \quad x > 0, \\ y(-\infty) &= \lim_{t\to 0^+} t^{-c/2a} u(x,t), \quad x < 0, \end{aligned} \tag{9.2.6}$$

with the right hand side equal to 0 if $c/2a < 0$, ∞ if $c/2a > 0$ or $u_0(x)$ if $c = 0$. Now, in the first two cases all the information coming from the initial condition is lost and the last one imposes a strict condition on u_0: u_0 must be constant on each semi-axis. Note that such a condition is also invariant under the transformation $z = x/\sqrt{t}$: $z \lessgtr 0$ if and only if $x \lessgtr 0$. In general, the similarity method provides a full solution to the initial-boundary value problems only if the side conditions are also invariant under the same transformation or, in other words, can be expressed in terms of the similarity variable. Otherwise, this method can serve as a first step in building more general solution, as we shall see below.

Consider the initial value problem

$$\begin{aligned} u_t &= Du_{xx}, \quad t > 0, -\infty < x < \infty, \\ u(x,0) &= H(x) \end{aligned} \tag{9.2.7}$$

where H is the Heaviside function: $H(x) = 1$ for $x \geq 0$ and $H(x) = 0$ for $x < 0$. According to the discussion above, this initial condition yields to the similarity method provided $c = 0$; in this case a is irrelevant and we put it equal to 1. Thus, the initial value problem (9.2.7) is transformed into

$$\begin{aligned} y'' + \frac{z}{2D} y' &= 0, \\ y(-\infty) = 0, \qquad & y(\infty) = 1. \end{aligned} \tag{9.2.8}$$

Denoting $y' = h$, we reduce the equation to the first order equation

$$h' + \frac{z}{2D} h = 0$$

which can be integrated to $y' = h = c_1 \exp(-z^2/4D)$. Integrating this once again, we obtain

$$y(z) = c_1 \int_0^z e^{-\frac{\eta^2}{4D}} d\eta + c_2 \tag{9.2.9}$$

and the constants c_1 and c_2 can be obtained from the initial conditions

$$\begin{aligned} 1 &= \lim_{z\to+\infty} y(z) = c_1 \int_0^\infty e^{-\frac{\eta^2}{4D}} d\eta + c_2 = \sqrt{4D} c_1 \int_0^\infty e^{-s^2} ds + c_2 = c_1 \frac{\sqrt{4D\pi}}{2} + c_2, \\ 0 &= \lim_{z\to-\infty} y(z) = c_1 \int_0^{-\infty} e^{-\frac{\eta^2}{4D}} d\eta + c_2 = \sqrt{4D} c_1 \int_0^{-\infty} e^{-s^2} ds + c_2 = -c_1 \frac{\sqrt{4D\pi}}{2} + c_2, \end{aligned}$$

where we used $\int_0^\infty e^{-s^2} ds = \sqrt{\pi}/2$, so that

$$c_2 = \frac{1}{2}, \quad c_1 = \frac{1}{\sqrt{4\pi D}}.$$

Hence

$$u(x,t) = y\left(\frac{x}{\sqrt{t}}\right) = \frac{1}{2} + \frac{1}{\sqrt{4\pi D}} \int_0^{\frac{x}{\sqrt{t}}} e^{-\frac{\eta^2}{4D}} d\eta \tag{9.2.10}$$

The fundamental rôle in the theory of the diffusion equation is played by the derivative of u with respect to x:

$$S(x,t) = \frac{\partial u}{\partial x}(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}} \tag{9.2.11}$$

that is called *source function* or *fundamental solution* of the diffusion equation.

The reason for the importance of the fundamental solution is that it describes diffusion of a unit quantity of the medium concentrated at the origin and thus provides the solution to the initial value problem for the diffusion equation of the form

$$\begin{aligned} u_t &= Du_{xx}, \\ u(x,0) &= \delta(x) \end{aligned}$$

where $\delta(x)$ is Dirac's delta "function". One can check that $S(x,t)$ has the following properties:

$$\lim_{t\to 0^+} S(x,t) = 0, \quad x \neq 0, \tag{9.2.12}$$

$$\lim_{t\to 0^+} S(x,t) = \infty, \quad x = 0, \tag{9.2.13}$$

$$\lim_{x\to \pm\infty} S(x,t) = 0, \quad x > 0, \tag{9.2.14}$$

$$\int_{-\infty}^{\infty} S(x,t)dx = 1, \quad t > 0, \tag{9.2.15}$$

In general, it can be proved that the initial value problem (9.2.5) has a unique solution given by

$$u(x,t) = \int_{-\infty}^{\infty} S(x-\xi)u_0(\xi)d\xi \tag{9.2.16}$$

provided u_0 is sufficiently regular (e.g. bounded and continuous).

3 Miscellaneous examples for solving the linear diffusion equation

As we have shown, the solution to the initial value problem for the diffusion equation is given by the integral

$$u(x,t) = \frac{1}{\sqrt{4D\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4Dt} \phi(y)dy,$$

where ϕ is the stipulated initial value of the solution. Unfortunately, this integral can be evaluated explicitly only in very few cases. In this section we shall present some such cases and discuss several other situations when one can obtain the solution to the problem, not necessarily resorting the integral formula.

Example 3.1 Matching the known solution
Find the solution of the problem

$$u_t = 2u_{xx}, \quad u(x,0) = e^{-x^2+x}.$$

To solve this equation we shall use the fact that if $v(x,t)$ is a solution to the diffusion equation, then for any constants a, x_0 and t_0 the function $u(x,t) = av(x+x_0, t+t_0)$ is also a solution. We know that

$$S(x,t) = \frac{1}{\sqrt{8\pi t}} e^{-x^2/8t}$$

is a solution to the equation above, and since the initial value has a form similar to S, we shall try and find constants such that the function

$$u(x,t) = aS(x+x_0, t+t_0) = \frac{a}{\sqrt{8\pi(t+t_0)}} e^{-(x+x_0)^2/8(t+t_0)}$$

satisfies the initial condition. We must have

$$e^{-x^2+x} = \frac{a}{\sqrt{8\pi t_0}} e^{-(x+x_0)^2/8t_0} = \frac{a}{\sqrt{8\pi t_0}} e^{-(x^2+2xx_0+x_0^2)/8t_0}$$

Comparing both sides we see that

$$\begin{aligned} 1 &= 1/8t_0, \\ 1 &= -x_0/4t_0, \\ 1 &= \frac{ae^{-x_0^2/8t_0}}{\sqrt{8\pi t_0}}. \end{aligned}$$

Therefore $t_0 = 1/8$, $x_0 = -1/2$ and $a = e^{1/4}\sqrt{\pi}$ and the solution is given by

$$u(x,t) = \frac{e^{1/4}}{\sqrt{8t+1}} e^{-\frac{(x-1/2)^2}{8t+1}}.$$

Note that this method can be used only for the initial data which have the particular form

$$\phi(x) = ae^{-bx^2+cx}$$

where a, b, c are constants with $b > 0$.

A variation of this method can be used for a little more general initial conditions. For example, to solve

$$u_t = u_{xx}, \quad u(x,0) = x^2 e^{-x^2} \tag{9.3.1}$$

we may use the fact that

$$x^2 e^{-x^2} = \frac{1}{4}\frac{d^2}{dx^2} e^{-x^2} + \frac{1}{2} e^{-x^2}$$

Using the approach described above, we obtain that the solution to the problem

$$u_t = u_{xx}, \quad u(x,0) = e^{-x^2}$$

is given by

$$u_1(x,t) = \frac{1}{\sqrt{4t+1}} e^{-x^2/(4t+1)}.$$

Since we know that the derivative of a solution is a solution and that sum of two solutions is a solution, we obtain that the solution of the problem (9.3.1) is given by

$$\begin{aligned} u(x,t) &= \frac{1}{4}\frac{d^2 u_1(x,t)}{dx^2} + \frac{1}{2} u_1(x,t) \\ &= -\frac{1}{2(4t+1)^{3/2}} e^{-x^2/(4t+1)} + \frac{x^2}{(4t+1)^{5/2}} e^{-x^2/(4t+1)} + \frac{1}{2(4t+1)^{1/2}} e^{-x^2/(4t+1)} \\ &= \frac{e^{-x^2/(4t+1)}}{(4t+1)^{3/2}} \left(\frac{x^2}{4t+1} + 2t \right). \end{aligned}$$

This method can be used for initial values of the form

$$\phi(x) = P_n(x) e^{-bx^2+cx}$$

where P_n is an arbitrary polynomial and a, b, c are as above.

Example 3.2 Using the integral formula
If the initial datum is a polynomial, an exponential function, sine or cosine, or a linear combination of them, then the integral in the representation formula can be evaluated explicitly and the solution is given as a combination of elementary functions. As an example we shall solve the problem

$$u_t = u_{xx}, \qquad u(x,0) = \cos^2 x.$$

First we observe that

$$\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x,$$

thus we can split the problem into two simpler ones:

$$v_t = v_{xx}, \qquad v(x,0) = \cos 2x,$$

and

$$w_t = w_{xx}, \qquad w(x,0) = 1.$$

If we find v and w, then from the linearity (superposition principle) we obtain that

$$u(x,t) = \frac{1}{2}v + \frac{1}{2}w.$$

Let us start with finding v. Using the representation formula we obtain

$$v(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4t} \cos 2y dy = Re \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4t} e^{2iy} dy,$$

where Re denotes the real part of the complex number. For the exponents, completing the square, we obtain,

$$\begin{aligned} -(x-y)^2/4t + 2iy &= -\frac{x^2 - 2xy + y^2 - 8iyt}{4t} = -\frac{(x-y+4it)^2 - 8ixt + 16t^2}{4t} \\ &= -\frac{(x-y+4it)^2}{4t} + 2ix - 4t, \end{aligned}$$

therefore

$$v(x,t) = Re \frac{e^{2ix-4t}}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y+4it)^2/4t} dy.$$

The integral

$$\frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y+4it)^2/4t} dy = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(p+2i\sqrt{t})^2} dp$$

resembles the integral $\int_{-\infty}^{\infty} e^{-p^2} dp$, however, it is evaluated along the line in the complex plane. Using the methods of complex integration we obtain that both integrals are equal with common value $\sqrt{\pi}$. Therefore

$$v(x,t) = Re\, e^{2ix-4t} = e^{-4t}\cos 2x.$$

To find w we evaluate the integral

$$w(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4t} \cdot 1 dy = 1.$$

Therefore the solution to the original problem is given by

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

Example 3.3 Polynomial initial values
To solve the initial value problem

$$u_t = Du_{xx}, \qquad u(x,0) = P_n(x),$$

where $P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ is a polynomial of n-th degree we can use the method of the previous example but for higher order polynomials the evaluation of integrals becomes quite labourious. Here we present an alternative way of solving the problem.

Let us assume that $u(x,t)$ is the solution; then its $n+1$ derivative

$$v(x,t) = \frac{d^{n+1}u(x,t)}{dx^{n+1}}$$

also is a solution to the diffusion equation satisfying the initial condition

$$v(x,0) = \frac{d^{n+1}u(x,0)}{dx^{n+1}} = \frac{d^{n+1}P_n(x)}{dx^{n+1}} = 0.$$

However, the only solution to the problem

$$v_t = Dv_{xx}, \qquad v(x,0) = 0,$$

is $v(x,t) \equiv 0$. Thus, the solution u of the original problem can be obtained by integrating the zero function $n+1$ times with respect to x, that is

$$u(x,t) = A_n(t)x^n + A_{n-1}(t)x^{n-1} + \ldots + A_0,$$

where $A_i(t)$ are as yet undetermined functions of t only, which must be found by inserting u to the equation, and comparing it to the initial value.

As an example we shall solve the initial value problem

$$u_t = Du_{xx}, \qquad u(x,0) = x^3 + x.$$

Following the approach outlined above we are looking for u in the form

$$u(x,t) = A_3(t)x^3 + A_2(t)x^2 + A_1(t)x + A_0(t).$$

Inserting such defined u into the equation we obtain

$$A_3'(t)x^3 + A_2'(t)x^2 + A_1'(t)x + A_0'(t) = 6DA_3(t)x + 2DA_2(t)$$

and comparing coefficients at the same powers of x we obtain

$$\begin{aligned} A_3' &= 0, \\ A_2' &= 0, \\ A_1' &= 6DA_3, \\ A_0' &= 2DA_2. \end{aligned}$$

Next, from the initial condition we obtain

$$A_3(0)x^3 + A_2(0)x^2 + A_1(0)x + A_0(0) = x^3 + x$$

which yields

$$\begin{aligned} A_3(0) &= 1, \\ A_2(0) &= 0, \\ A_1(0) &= 1, \\ A_0(0) &= 0. \end{aligned}$$

Solving the above system of ordinary differential equations we obtain

$$\begin{aligned} A_3(t) &= 1, \\ A_2(t) &= 0, \\ A_1(t) &= 6Dt + 1, \\ A_0' &= 0. \end{aligned}$$

Thus, the solution is given by

$$u(x,t) = x^3 + (6Dt+1)x.$$

Example 3.4 Drift-diffusion equation
Let us consider the general parabolic equation in two variables (drift-diffusion equation)

$$u_t = Au + Bu_x + Cu_{xx}.$$

In Example **??** we showed how to reduce this equation to the diffusion equation by introducing a new independent variable. Here we show that it is also possible to reduce it to the diffusion equation by changing the unknown function. We introduce the new unknown function v according to the formula

$$u(x,t) = e^{ax+bt}v(x,t),$$

where a and b are coefficients to be determined. Differentiating, we obtain

$$\begin{aligned} u_t &= be^{ax+bt}v(x,t) + e^{ax+bt}v_t(x,t), \\ u_x &= ae^{ax+bt}v(x,t) + e^{ax+bt}v_x(x,t), \\ u_{xx} &= a^2e^{ax+bt}v(x,t) + 2ae^{ax+bt}v_x(x,t) + e^{ax+bt}v_{xx}(x,t) \end{aligned}$$

Inserting the above expressions into the equation, collecting terms and dividing by e^{ax+bt} we arrive at

$$v_t = (A + Ba + Ca^2 - b)v + (B + 2Ca)v_x + Cv_{xx}.$$

From the above equation we see that taking

$$a = -\frac{B}{2C}, \qquad b = A - \frac{B^2}{4C}$$

will make the coefficients multiplying v and v_x equal to zero, so that v will be the solution to

$$v_t = Cv_{xx}.$$

To illustrate this technique let us consider the initial-value problem

$$u_t = 2u_x + u_{xx}, \qquad u(x,0) = x^2.$$

Following the above considerations we find $a = b = -1$ so that

$$u(x,t) = e^{-x-t}v(x,t).$$

Consequently, v solves the following initial-value problem

$$v_t = v_{xx}, \qquad v(x,0) = x^2e^x.$$

Using the integral formula we obtain

$$v(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4t} y^2 e^y dy$$

For the exponents, completing the square, we obtain,

$$\begin{aligned} -(x-y)^2/4t + y &= -\frac{x^2 - 2xy + y^2 + 4yt}{4t} = -\frac{(x-y+2t)^2 - 4xt - 4t^2}{4t} \\ &= -\frac{(x-y+2t)^2}{4t} + x + t, \end{aligned}$$

therefore

$$v(x,t) = \frac{e^{x+t}}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-(x-y+2t)^2/4t} y^2 dy.$$

Changing the variable according to $p = -(x-y+2t)/2\sqrt{t}$ so that $y = x + 2t + 2p\sqrt{t}$ we obtain

$$\begin{aligned} v(x,t) &= \frac{e^{x+t}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} (x + 2t + 2p\sqrt{t})^2 dp \\ &= \frac{(x+2t)^2 e^{x+t}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} dp + \frac{4(x+2t)\sqrt{t} e^{x+t}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} p dp \\ &\quad + \frac{4te^{x+t}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} p^2 dp \end{aligned}$$

The first integral in the above is equal to $\sqrt{\pi}$. The second is zero, as pe^{-p^2} is an odd function. The last one can be evaluated by parts as follows

$$\int_{-\infty}^{\infty} e^{-p^2} p^2 dp = -\frac{1}{2} e^{-p^2} p \Big|_{-\infty}^{+\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-p^2} dp = \frac{\sqrt{\pi}}{2}.$$

Putting these together we obtain

$$v(x,t) = \frac{e^{x+t}}{\sqrt{\pi}} \left(\sqrt{\pi}(x+2t)^2 + 4t\frac{\sqrt{\pi}}{2} \right) = e^{x+t}(2t + (x+2t)^2),$$

thus we finally obtain

$$u(x,t) = (2t + (x+2t)^2).$$

let us now compere this approach with that introduced in Example **??**. Following it, we introduce the function $w(y,t) = u(y-2t,t)$ so that $w_t = w_{xx}$. The initial condition becomes $w(y,0) = u(x,0) = y^2$ as $x|_{t=0} = y$. For this problem we can apply the approach of Example 3.3 and look for w in the form

$$w(y,t) = A_2(t)y^2 + A_1(t)y + A_0.$$

Inserting this formula into the diffusion equation, we get

$$A_2' = 0, \quad A_1' = 0, \quad A_0' = 2A_0$$

with initial conditions $A_2(0) = 1, A_1(0) = A_0(0) = 0$. This yields $A_1(t) = 0, A_2(t) = 1$ and $A_0(t) = 2t$. Consequently, $w(y,t) = y^2 + 2t$ and $u(x,t) = (x+t)^2 + 2t$, in accordance with the previous method. Here, the latter method appears simpler things, however, change, when one discusses boundary conditions.

4 Black-Scholes formulae

In this section we shall derive the Black-Scholes formulas for option pricing. Let us recall that we are looking for the solution of the following initial-boundary value problem

$$\begin{aligned} \frac{\partial V}{\partial t} &= -\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r\left(V - S\frac{\partial V}{\partial S}\right), \\ V(S,T) &= \max\{S-E,0\}, \quad \text{for all } S>0, \\ V(0,t) &= 0, \quad \text{for all } t \le T, \\ \lim_{S\to+\infty} V(S,t)/S &= 1, \quad \text{for all } t \le T. \end{aligned} \tag{9.4.2}$$

Using the change of variables (**??**)

$$\begin{aligned} x(S,t) &= \ln S + \left(r - \frac{1}{2}\sigma^2\right)(T-t), \\ y(t) &= \frac{1}{2}\sigma^2(T-t), \end{aligned}$$

with the inverse

$$\begin{aligned} T-t &= \frac{2}{\sigma^2}y, \\ S &= e^{x-\left(\frac{2r}{\sigma^2}-1\right)y}, \end{aligned}$$

we reduced the problem (9.4.2) to the following initial-boundary value problem for the diffusion equation:

$$\begin{aligned} G_y &= G_{xx}, \quad \text{for } -\infty < x < \infty, y > 0 \\ G(x,0) &= \max\{e^x - E, 0\}, \quad \text{for } -\infty < x < \infty, \\ \lim_{x\to-\infty} G(x,y) &= 0, \qquad \text{for any } y \ge 0, \\ \lim_{x\to-\infty} e^{-(x+y)} G(x,y) &= 1, \qquad \text{for any } y \ge 0. \end{aligned} \tag{9.4.3}$$

So, if we have the solution to the diffusion problem, $G(x,y)$, then the solution to the Black-Scholes problem is given by

$$V(S,t) = e^{-r(T-t)} G\left(\ln S + \left(r - \frac{1}{2}\sigma^2\right)(T-t), \frac{1}{2}\sigma^2(T-t)\right) \tag{9.4.4}$$

From the previous sections we know that the solution of the pure initial value problem with the initial condition ϕ is given by

$$u(x,t) = \frac{1}{2\sqrt{\pi D t}} \int_{-\infty}^{\infty} e^{-(x-y)^2/4Dt} \phi(y) dy.$$

In our case $D = 1$ and since e^x is an increasing function, the initial value takes the simpler form

$$\phi(x) = \begin{cases} e^x - E & \text{for} \quad x > \ln E \\ 0 & \text{for} \quad x \le \ln E. \end{cases}$$

Thus, we obtain

$$G(x,y) = \frac{1}{2\sqrt{\pi y}} \int_{\ln E}^{\infty} e^{-\frac{(x-s)^2}{4y}} (e^s - E) ds. \tag{9.4.5}$$

This solution is closely related to the cumulative distribution function for a normal random variable and it is sensible to express it in terms of the distribution function of standardised normal random variable defined as

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{p^2}{2}} dp.$$

To write G (and subsequently V) in terms of N we split the integral (9.4.5) and evaluate them using to subsequent changes of variables.

$$\begin{aligned}
G(x,y) &= \frac{1}{2\sqrt{\pi y}}\left(\int_{\ln E}^{\infty} e^{-\frac{(x-s)^2}{4y}} e^s ds - E\int_{\ln E}^{\infty} e^{-\frac{(x-s)^2}{4y}} ds\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\frac{x-\ln E}{\sqrt{2y}}} e^{-\frac{p^2}{2}} e^{x-\sqrt{2y}} dp - E\int_{-\infty}^{\frac{x-\ln E}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(e^{x+y}\int_{-\infty}^{\frac{x-\ln E}{\sqrt{2y}}} e^{-\frac{(p+\sqrt{2y})^2}{2}} dp - E\int_{-\infty}^{\frac{x-\ln E}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp\right) \\
&= \frac{1}{\sqrt{2\pi}}\left(e^{x+y}\int_{-\infty}^{\frac{x-\ln E+2y}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp - E\int_{-\infty}^{\frac{x-\ln E}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp\right) \\
&= e^{x+y} N\left(\frac{x-\ln E+2y}{\sqrt{2y}}\right) - EN\left(\frac{x-\ln E}{\sqrt{2y}}\right).
\end{aligned}$$

Before we write down the Black-Scholes formula, let us note that despite the fact that we have derived G as the solution of the initial value problem, luckily it also satisfies the boundary conditions. To prove it, let us first observe that

$$\begin{aligned}
\lim_{x\to-\infty} N(x) &= 0, \\
\lim_{x\to+\infty} N(x) &= 1.
\end{aligned}$$

Since $\lim_{x\to-\infty} e^{x+y} = 0$ for any fixed y, we immediately see that

$$\lim_{x\to-\infty} G(x,y) = \lim_{x\to-\infty}\left(e^{x+y} N\left(\frac{x-\ln E+2y}{\sqrt{2y}}\right) - EN\left(\frac{x-\ln E}{\sqrt{2y}}\right)\right) = 0$$

and

$$\lim_{x\to\infty} e^{-(x+y)} G(x,y) = \lim_{x\to\infty}\left(N\left(\frac{x-\ln E+2y}{\sqrt{2y}}\right) - e^{-(x+y)} EN\left(\frac{x-\ln E}{\sqrt{2y}}\right)\right) = 1,$$

so that G is the function which we have been looking for.

To derive explicit formulas for option pricing we must change back the variables x and y into S and t. We get

$$\frac{x-\ln E}{\sqrt{2y}} = \frac{\ln S/E + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and

$$\frac{x-\ln E+2y}{\sqrt{2y}} = \frac{\ln S/E + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}.$$

Since

$$e^{x+y} = e^{\ln S + r(T-t)} = Se^{r(T-t)},$$

we obtain explicitly from (9.4.4)

$$\begin{aligned}
V(S,t) &= SN\left(\frac{\ln S/E + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right) \\
&\quad - Ee^{-r(T-t)} N\left(\frac{\ln S/E + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right).
\end{aligned} \tag{9.4.6}$$

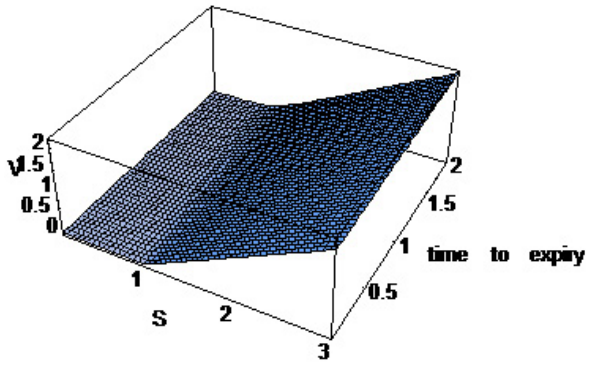

Figure 9.1: 3-dimensional visualization of the solution to the Black-Scholes equation giving the pricing of European call options.

The figures show the graphs of the solution.

We complete the section by giving the formula for the European put options.

Example 4.1 Black-Scholes formula for put options

In the main body of the course we have focused on the European call options, that is, contracts allowing the holder to buy a share at the prescribed expiry time T at a prescribed exercise price E. A somewhat complementary contract which allows the holder to sell a share at a prescribed price at a prescribed time is called the put option. The equation governing the evolution of the put option price is the same Black-Scholes equation (recall that in the process of deriving this equation we have never used the fact that we have call options in the portfolio), that is, the put option price satisfies

$$\frac{\partial V}{\partial t} = -\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r\left(V - S\frac{\partial V}{\partial S}\right).$$

The difference appears in side conditions. Let us consider the pay-off at the expiry time. If the value of the share S is smaller than the exercise price E, then clearly it is worthwhile to sell the share at E and get the profit of $E - S$. On the other hand, if $S > E$, then it is better to sell the share elsewhere, so the option is worthless. Thus we arrived at the final condition

$$V(S,T) = \max\{E - S, 0\}.$$

As far as the boundary conditions are concerned, we note that if at some stage S becomes zero, then it will stay at this level and then we shall exercise the option getting certain profit of E at time T. However, our basic assumption is that we operate in an arbitrage-free market and therefore this profit must be equal to that obtained by depositing the amount of $V(0,t)$ in a bank at time t. Assuming constant interest rate r we obtain that

$$E = V(0,t)e^{r(T-t)},$$

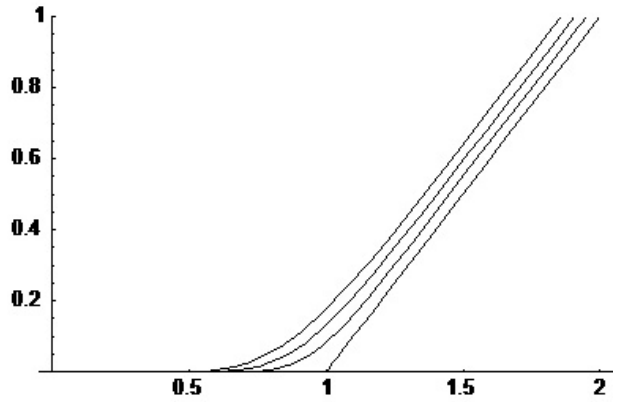

Figure 9.2: The price of a European call option as a function of the price of the underlying share as time approaches the expiry time. The bottom line gives the pay-off

so that we obtain the boundary condition at $S = 0$ to be

$$V(0,t) = Ee^{-r(T-t)}.$$

The other natural boundary is as $S \to \infty$. If S becomes very large, then it is unlikely that the option will be exercised, thus the value of the option is zero. Hence, the full initial-boundary value problem for the Black-Scholes equation for pricing put options is

$$\begin{aligned} \frac{\partial V}{\partial t} &= -\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r\left(V - S\frac{\partial V}{\partial S}\right), \\ V(S,T) &= \max\{E - S, 0\}, \quad \text{for all } S > 0, \\ V(0,t) &= Ee^{-r(T-t)}, \quad \text{for all } t \leq T, \\ \lim_{S\to+\infty} V(S,t) &= 0, \quad \text{for all } t \leq T. \end{aligned} \tag{9.4.7}$$

Using the same change of variables as for call options (the equation is the same) we obtain the relation between V and the solution G to the diffusion equation

$$V(S,t) = e^{-r(T-t)} G\left(\ln S + \left(r - \frac{1}{2}\sigma^2\right)(T-t), \frac{1}{2}\sigma^2(T-t)\right) \tag{9.4.8}$$

Clearly, if $t = T$, then $V(S,T) = G(\ln S, 0)$, that is the initial value for G is

$$G(x,0) = V(e^x, 0) = \max\{E - e^x, 0\}.$$

As S approaches 0, then $x = \ln S + \left(r - \frac{1}{2}\sigma^2\right)(T-t)$ approaches $-\infty$ and we obtain

$$Ee^{-r(T-t)} = \lim_{S\to 0} V(S,t) = \lim_{x\to-\infty} e^{-r(T-t)} G(x,y)$$

so that one boundary condition for G reads:

$$\lim_{x\to-\infty} G(x,y) = E.$$

On the other hand, as $S \to \infty$, so does x, and the last boundary condition for G takes the form

$$\lim_{x\to+\infty} G(x,y) = 0.$$

Summarizing, to obtain the solution for the Black-Scholes equation via formula (9.4.8) we must solve the initial-boundary value problem

$$\begin{aligned} G_y &= G_{xx}, \quad \text{for } -\infty < x < \infty, y > 0 \\ G(x,0) &= \max\{E - e^x, 0\}, \quad \text{for } -\infty < x < \infty, \\ \lim_{x\to-\infty} G(x,y) &= E, \qquad \text{for any } y \geq 0, \\ \lim_{x\to\infty} G(x,y) &= 0, \qquad \text{for any } y \geq 0. \end{aligned} \tag{9.4.9}$$

The following calculation are almost the same as in Section 4. The initial condition can be written as

$$\phi(x) = \begin{cases} E - e^x & \text{for} \quad x < \ln E \\ 0 & \text{for} \quad x \geq \ln E, \end{cases}$$

hence we obtain

$$G(x,y) = \frac{1}{2\sqrt{\pi y}} \int_{-\infty}^{\ln E} e^{-\frac{(x-s)^2}{4y}} (E - e^s) ds. \tag{9.4.10}$$

As before we express this solution in terms of the distribution function of standardised normal random variable defined as

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{p^2}{2}} dp.$$

We split the integral (9.4.10) and performing subsequently two changes of variables we get

$$\begin{aligned} G(x,y) &= \frac{1}{2\sqrt{\pi y}} \left(E \int_{-\infty}^{\ln E} e^{-\frac{(x-s)^2}{4y}} ds - \int_{-\infty}^{\ln E} e^{-\frac{(x-s)^2}{4y}} e^s ds \right) \\ &= \frac{1}{\sqrt{2\pi}} \left(E \int_{-\infty}^{\frac{\ln E - x}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp - \int_{-\infty}^{\frac{\ln E - x}{\sqrt{2y}}} e^{-\frac{p^2}{2}} e^{x - \sqrt{2y}} dp \right) \\ &= \frac{1}{\sqrt{2\pi}} \left(E \int_{-\infty}^{\frac{\ln E - x}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp - e^{x+y} \int_{-\infty}^{\frac{\ln E - x}{\sqrt{2y}}} e^{-\frac{(p-\sqrt{2y})^2}{2}} dp \right) \\ &= \frac{1}{\sqrt{2\pi}} \left(E \int_{-\infty}^{\frac{\ln E - x}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp - e^{x+y} \int_{-\infty}^{\frac{\ln E - x - 2y}{\sqrt{2y}}} e^{-\frac{p^2}{2}} dp \right) \\ &= EN\left(\frac{\ln E - x}{\sqrt{2y}} \right) - e^{x+y} N\left(\frac{\ln E - x - 2y}{\sqrt{2y}} \right). \end{aligned}$$

As before let us note that despite the fact that we derived G as the solution of the initial value problem, luckily it also satisfies the boundary conditions. To prove it, let us recall that

$$\begin{aligned} \lim_{x\to-\infty} N(x) &= 0, \\ \lim_{x\to+\infty} N(x) &= 1. \end{aligned}$$

Since $\lim_{x\to-\infty} e^{x+y} = 0$ for any fixed y, we immediately see that

$$\lim_{x\to-\infty} G(x,y) = \lim_{x\to-\infty}\left(EN\left(\frac{\ln E - x}{\sqrt{2y}}\right) - e^{x+y}N\left(\frac{\ln E - x - 2y}{\sqrt{2y}}\right)\right) = E,$$

and

$$\lim_{x\to\infty} e^{-(x+y)}G(x,y) = \lim_{x\to\infty}\left(e^{-(x+y)}EN\left(\frac{\ln E - x}{\sqrt{2y}}\right) - N\left(\frac{\ln E - x - 2y}{\sqrt{2y}}\right)\right) = 0,$$

so that G is the function which we have been looking for.

To derive explicit formulas for option pricing we must change back the variables x and y into S and t. As for call options we get

$$\frac{\ln E - x}{\sqrt{2y}} = -\frac{\ln S/E + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and

$$\frac{\ln E - x - 2y}{\sqrt{2y}} = -\frac{\ln S/E + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}.$$

Since

$$e^{x+y} = e^{\ln S + r(T-t)} = Se^{r(T-t)},$$

we obtain explicitly from (9.4.8)

$$\begin{aligned} V(S,t) &= Ee^{-r(T-t)}N\left(-\frac{\ln S/E + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right) \\ &\quad -SN\left(-\frac{\ln S/E + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right). \end{aligned} \tag{9.4.11}$$

Next we derive the formula relating call and put options. To distinguish them we denote by V_c and V_p the price of call and put options, respectively.

Let us first observe that since the function $e^{-p^2/2}$ is even and

$$\lim_{x\to\infty} N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{p^2}{2}}dp = 1,$$

we have

$$N(-x) = 1 - N(x). \tag{9.4.12}$$

Denoting

$$\alpha = \frac{\ln S/E + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}$$

and

$$\beta = \frac{\ln S/E + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}},$$

we see that the price formula for call options (9.4.6) can be written as

$$V_c(S,t) = SN(\beta) - Ee^{-r(T-t)}N(\alpha),$$

whereas for put options we have from (9.4.11)

$$V_p(S,t) = Ee^{-r(T-t)}N(-\alpha) - SN(-\beta).$$

Subtracting we obtain using (9.4.12)

$$V_c(S,t) - V_p(S,t) = SN(\beta) - Ee^{-r(T-t)}N(\alpha) - Ee^{-r(T-t)}(1-N(\alpha)) + S(1-N(\beta)) = S - Ee^{-r(T-t)}.$$

The derived formula

$$S - V_c(S,t) + V_p(S,t) = Ee^{-r(T-t)} \tag{9.4.13}$$

is called *put-call parity* and expresses the relationship between the underlying asset and its options.

5 Nonlinear diffusion models

The linear diffusion equation, though extremely useful and important in applications, has at least one major drawback - it is physically incorrect as it admits infinite speed of signal transmission. To alleviate this problem a number of nonlinear versions of this equation has been presented and we shall discuss some of them.

5.1 Models

Let us recall that the starting point for the derivation of the diffusion equation was the conservation law

$$u_t + \phi_x = 0 \tag{9.5.1}$$

and the linear equation was obtained by postulating the flux in the form of Fick's law

$$\phi = -Du_x. \tag{9.5.2}$$

In population models usually there is the increase in diffusion due to the population pressure and it is then reasonable to assume that the coefficient D itself should depend on the density u, that is,

$$\phi = -D(u)u_x \tag{9.5.3}$$

and substituting this to the conservation law (9.5.1) yields the nonlinear equation

$$u_t = (D(u)u_x)_x. \tag{9.5.4}$$

Another place where the nonlinearity can occur is at the time derivative. For instance, deriving the heat equation we start with the energy conservation law

$$(\rho CT)_t + \phi_x = 0 \tag{9.5.5}$$

where T is the temperature, ρ is the density and C is the specific heat of the medium (so that the first terms describes the rate of change of the amount of energy contained in a unit volume). The flux ϕ is given by the Fourier law

$$\phi = -KT_x,$$

where K is thermal conductivity of the medium. In many applications, where the temperature range is limited, the specific heat and the density may be regarded as constants. However, over a wide temperature ranges, they are not constants but rather depend on the temperature and in this case (9.5.5) combined with Fourier's law give the equation

$$(\rho(T)C(T)T)_t = KT_{xx}, \tag{9.5.6}$$

which is a nonlinear diffusion equation for the temperature T.

These two types of nonlinearities can, of course, appear in a single equation. Consider, for instance, the porous medium equation, where we wish to describe a fluid (e.g. water) seeping downward through the soil. Let $\rho(x,t)$ be the density of the fluid, with positive x measured downward. In a given volume of soil only a fraction of the space is available to the fluid, the remaining being reserved for the soil itself. If the fraction of the volume that is available to the fluid is denoted by κ, then mass balance for the fluid can be written as

$$\kappa\rho_t + (\rho v)_x = 0, \tag{9.5.7}$$

where we used the formula $\phi = v\rho$ with v being the volumetric flow rate (velocity of the flow). The conservation law (9.5.7) contains two unknowns: the density and the velocity of the flow so we need another constitutive equation relating them. Without going through too much of physics we state that in many cases the constitutive relation

$$v = a\rho_x^\gamma$$

for some constants $\gamma > 1, a > 0$. Thus, the porous media equation can be written as

$$\kappa\rho_t = a(\rho\rho_x^\gamma)_x = a\gamma(\rho^\gamma\rho_x)_x = \alpha\rho_{xx}^m \tag{9.5.8}$$

where $\alpha = a\gamma/(\gamma+1)$ and $m = \gamma + 1 > 2$. Since, in general, κ can be ρ-dependent, the porous media equation combines nonlinearities of (9.5.6) and (9.5.4) (with $D(\rho) = a\gamma\rho^\gamma$.)

5.2 Some solutions

Let us consider a special case of Eq. (9.5.4) with $D(u) = u$, that is,

$$u_t - (uu_x)_x = 0. \tag{9.5.9}$$

To compare the nonlinear equation with the linear diffusion, we shall try to solve it on $\mathbb{R}$ subject to the initial condition of a unit point source applied at $x = 0$, that is,

$$u(x, 0) = \delta(x).$$

To simplify considerations we shall change this condition into two, more amenable to the similarity method, so that we shall look for solutions satisfying

$$\int_{-\infty}^{\infty} u(x,t)dx = 1, \qquad t > 0 \tag{9.5.10}$$

and

$$\lim_{x\to\pm\infty} = 0. \tag{9.5.11}$$

As follows from Eqs. (9.2.12)–(9.2.15), in the linear case the solution satisfying these conditions is the fundamental solution $S(x, t)$ to the diffusion equation.

Let us introduce the stretching transformation

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^b t, \bar{u} = \epsilon^c u \tag{9.5.12}$$

and substitute it into our nonlinear diffusion equation

$$u_t - (uu_x)_x = u_t - u_x^2 - uu_{xx} = p - q^2 - ur = \epsilon^{b-c}\bar{p} - \epsilon^{2(a-c)}\bar{q}^2 - \epsilon^{-2c+2a}\bar{u}\bar{r}$$

from where we see that $b = 2a - c$ so that

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^b t, \bar{u} = \epsilon^{2b-a} u \tag{9.5.13}$$

and the similarity transformation is given by

$$u(x,t) = t^{(2a-b)/b} y(z), \qquad z = \frac{x}{t^{a/b}}. \tag{9.5.14}$$

Let us first specialize the parameters a and b by imposing the condition (9.5.10). We have

$$1 = \int_{-\infty}^{\infty} u(x,t)dx = t^{(2a-b)/b} \int_{-\infty}^{\infty} y\left(\frac{x}{t^{a/b}}\right) dx = t^{\frac{3a-b}{b}} \int_{-\infty}^{\infty} y(z)dz$$

so that $3a - b = 0$. Hence

$$u(x,t) = t^{-1/3} y(z), \qquad z = xt^{-1/3}. \tag{9.5.15}$$

and substituting these equations we obtain

$$\begin{aligned} u_t &= -\frac{1}{3} t^{-4/3}(y + y'z), \\ u_x &= t^{-2/3} y', \\ (uu_x)_x &= t^{-1}(yy')_x = t^{-1}(yy')' z_x = t^{-4/3}(yy')' \end{aligned}$$

and consequently (9.5.11) turns into

$$3(yy')' + y + zy' = 0. \tag{9.5.16}$$

As $y + zy' = (zy)'$, this equation can be integrated at once giving

$$3yy' + zy = constant. \tag{9.5.17}$$

Now, the equation is invariant under the change of variable $x \to -x$ and the side conditions also are not altered when we make this change so that it is reasonable to expect the solution to have the same property, that is $u(x) = u(-x)$. In other ways, we expect the solution to be even and in such a case we derive another condition, namely $u_x(0,t) = 0$ for any $t > 0$. Since $y'(z) = t^{2/3}u_x(x,t)$ we see that $y'(0) = 0$. Putting $z = 0$ in (9.5.17) we see that the constant must be zero and hence we obtain a first order separable equation

$$3yy' + zy = 0$$

that can be immediately integrated giving either $y = 0$ or $y = \frac{-z^2+A^2}{6}$, where A is a constant of integration. Separately, neither solution makes sense as the former does not satisfy the integral condition and the second becomes negative for $|z| > A$. Let us then patch these two solutions together and consider

$$y(z) = \begin{cases} \frac{-z^2+A^2}{6} & \text{if } |z| < A, \\ 0 & \text{if } |z| \geq A. \end{cases} \tag{9.5.18}$$

At first glance this seems to be a ridiculous idea as such a function is merely continuous and we are looking for a solution to the second order differential equation. However, a closer look at (9.5.16) shows that the equation does not involve the second derivative of y by itself but rather requires differentiability of yy' that is less stringent as y is zero at the discontinuity of y'. We note thus that

$$y'(z) = \begin{cases} \frac{-z}{3} & \text{if } |z| < A, \\ 0 & \text{if } |z| > A, \end{cases} \tag{9.5.19}$$

with one-sided derivatives at $z = \pm A$: $y'_-(-A) = y'_+(A) = 0$, $y'_+(-A) = A/3$ and $y'_-(A) = -A/3$. Thus

$$yy'(z) = \begin{cases} \frac{-z(A^2-z^2)}{18} & \text{if } |z| < A, \\ 0 & \text{if } |z| \geq A, \end{cases} \tag{9.5.20}$$

so that

$$(yy')'(z) = \begin{cases} \frac{-A^2+3z^2}{18} & \text{if } |z| < A, \\ 0 & \text{if } |z| > A, \end{cases} \tag{9.5.21}$$

with one-sided derivatives at $\pm A$: $(yy')'_-(-A) = (yy')'_+(A) = 0$, $(yy')'_+(-A) = (yy')'_-(A) = A^2/9$. Clearly, the equation is satisfied on open intervals $|z| < A$ and $|z| > A$. Taking one-sided values from the left at $-A$ and from the right at A we have zeros; from the right at $-A$ we obtain $3A^2/9 + 0 + (-A)A/3 = 0$ and similarly from the left at A: $3A^2/9 + 0 + A(-A/3) = 0$. Hence, the equation is satisfied everywhere if we allow interpretation of derivatives as one-sided ones. Such solutions are called piecewise smooth.

Having accepted (9.5.18) as a solution to our problem, we can specify A by requiring (9.5.10) to be satisfied. Thus

$$1 = \int_{-\infty}^{\infty} y(z)dz = \int_{-A}^{A} y(z)dz = \frac{2}{9}A^3$$

that gives $A = \sqrt[3]{2/9}$.

Therefore we have constructed a piecewise smooth solution

$$u(x,t) = \begin{cases} \frac{1}{6}t^{-2/3}\left((9t/2)^{2/3} - x^2\right) & \text{if } |x| < (9t/2)^{1/3}, \\ 0 & \text{if } |x| \geq (9t/2)^{1/3}, \end{cases} \tag{9.5.22}$$

It is easy to see that as $t \to 0^+$, $u(x,t)$ converges to $\delta(x)$ in the sense of (9.2.12)–(9.2.13). Snapshots of the solution are shown on Fig. 6.1.

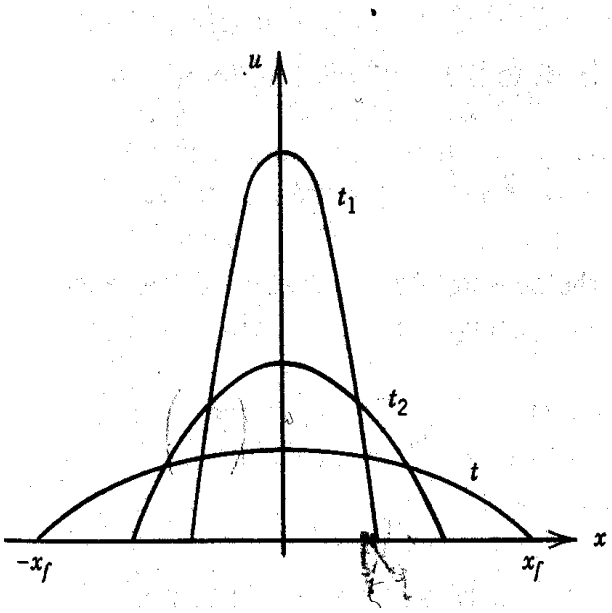

Fig. 6.1 Snapshots of (9.5.22).

Thus, we see that solution (9.5.22) is fundamentally different from the smooth and everywhere positive solution $u(x,t) = (4\pi)^{-1/2}\exp\left(-x^2/4t\right)$ of the corresponding linear problem. In fact, (9.5.22) represents a type of a sharp wavefront $x_f = (9t/2)^{1/3}$ propagating into the medium with speed

$$\frac{dx_f}{dt} = \sqrt[3]{t/6}.$$

The last example we are going to discuss is the nonlinear heat equation

$$uu_t - u_{xx} = 0, \quad x > 0, t > 0 \tag{9.5.23}$$

subject to side conditions

$$\begin{aligned} u(x,0) &= 0, \quad x > 0, && (9.5.24)\\ u(\infty,t) &= 0, \quad t > 0, && (9.5.25)\\ u_x(0,t) &= -1, \quad t > 0. && (9.5.26) \end{aligned}$$

This problem is a simplified version of the nonlinear heat equation (9.5.6) with temperature zero initially and as $x \to \infty$. The flux condition describes the heat flowing into the medium at the constant rate -1 (gradient of the temperature is negative and heat flows from regions of higher temperature to regions of lower temperature).

As before, we introduce the stretching transformation

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^b t, \bar{u} = \epsilon^c u \tag{9.5.27}$$

and substitute it to our nonlinear diffusion equation

$$uu_t - u_{xx} = up - r = \epsilon^{b-2c}\bar{p} - \epsilon^{2a-c}\bar{r}$$

from where we see that we have $c = b - 2a$ so that

$$\bar{x} = \epsilon^a x, \quad \bar{t} = \epsilon^b t, \bar{u} = \epsilon^{b-2a} u \tag{9.5.28}$$

and the similarity transformation is given by

$$u(x,t) = t^{(b-2a)/b} y(z), \qquad z = \frac{x}{t^{a/b}}. \tag{9.5.29}$$

Restrictions on the constants a and b can be determined by the initial and boundary conditions. Evaluating $u_x(x,t)$ gives

$$u_x(x,t) = t^{(b-3a)/b} y'(z) \tag{9.5.30}$$

and therefore

$$u_x(0,t) = t^{(b-3a)/b} y'(0) = -1 \tag{9.5.31}$$

which could be possible only if $b = 3a$. Consequently, the similarity transformation is given by

$$u(x,t) = t^{1/3} y(z), \qquad z = \frac{x}{\sqrt{3} t^{1/3}}, \tag{9.5.32}$$

where the factor $\sqrt{3}$ was introduced to simplify further calculations. Since in the initial condition (9.5.26) $t > 0$ is arbitrary, the flux condition will become

$$y'(0) = -\sqrt{3}, \tag{9.5.33}$$

where the factor $\sqrt{3}$ was introduced due to the modified expression for z. Condition (9.5.25) implies

$$y(\infty) = 0 \tag{9.5.34}$$

and the initial condition can be written as

$$\lim_{t \to 0^+} u(x,t) = \lim_{z \to 0} t^{1/3} y(z) = \lim_{z \to 0} \frac{x}{\sqrt{3} z} y(z) = 0, \qquad x > 0$$

that is even stronger condition than (9.5.34). Thus, conditions (9.5.24) and (9.5.26) have coalesced into the single condition (9.5.34). Once again we emphasize here that the fact that the original boundary conditions can be written as conditions in similarity variables follows from their special form - in general such a transformation is impossible.

Let us now transform the equation. We have

$$\begin{aligned}
z_t &= -\frac{z}{3\sqrt{3} t}, \\
u_t &= \frac{1}{3} t^{-2/3} \left(y - y' z \right), \\
u_x &= \frac{y'}{\sqrt{3}}, \\
u_{xx} &= t^{-1/3} \frac{y''}{3}, \\
u u_t &= \frac{1}{3} t^{-1/3} \left(y^2 - \frac{y y' z}{\sqrt{3}} \right)
\end{aligned}$$

and consequently (9.5.23) turns into

$$f'' - f(f - z f') = 0. \tag{9.5.35}$$

Unfortunately, contrary to the previous cases, this is second order non-autonomous equation and cannot be easily solved. However, we can still simplify it using once again the similarity variables. We shall formulate and prove the following lemma.

Lemma 5.1 *Assume that the equation*

$$f'' - G(z, f, f') = 0 \tag{9.5.36}$$

is invariant under the transformation

$$s = \epsilon z, \qquad g = \epsilon^b f, \qquad \epsilon \in I \tag{9.5.37}$$

where I is an open interval containing 1 and b is a constant, that is,

$$\frac{d^2g}{ds^2} - G\left(s, g, \frac{dg}{ds}\right) = A(\epsilon)\left(\frac{d^2f}{dz^2} - G\left(z, f, \frac{df}{dz}\right)\right).$$

Then (9.5.36) can be reduced to a first order ordinary differential equation of the form

$$\frac{d\eta}{d\xi} = \frac{H(\xi,\eta) - (b-1)\eta}{\eta - b\xi}, \tag{9.5.38}$$

where $\xi = \phi(z,f)$ and $\eta = \psi(z,f,p)$ are solutions of the characteristic system

$$\begin{aligned} \frac{df}{dz} &= \frac{bf}{z}, \\ \frac{dp}{dz} &= \frac{(b-1)p}{z}, \end{aligned} \tag{9.5.39}$$

and H is some function depending on G.

Proof. The proof is similar to the proof of Theorem 1.1. Denoting $p = f'$, we obtain from invariance

$$\frac{d^2g}{ds^2} - G\left(s, g, \frac{dg}{ds}\right) = \epsilon^{b-2}f'' - G(\epsilon z, \epsilon^b f, \epsilon^{b-1}p) = A(\epsilon)\left(\frac{d^2f}{dz^2} - G\left(z, f, p\right)\right),$$

so that

$$\epsilon^{b-2}G(\epsilon z, \epsilon^b f, \epsilon^{b-1}p) = G\left(z, f, p\right)$$

for all $\epsilon \in I$. Differentiating with respect to ϵ and putting $\epsilon = 1$ we obtain

$$zG_z + bfG_f + (b-1)pG_p = (b-2)G.$$

The characteristic system is

$$\begin{aligned} \frac{dG}{dz} &= (b-2)\frac{G}{z}, \\ \frac{df}{dz} &= \frac{bf}{z}, \\ \frac{dp}{dz} &= (b-1)\frac{p}{z}. \end{aligned}$$

The solutions of the characteristic system are

$$\xi = fz^{-b}, \qquad \eta = pz^{1-b}$$

so that $G = z^{b-2}H(\xi,\eta)$ for some function H. To arrive at (9.5.38) we observe that, since

$$\frac{d\xi}{dz} = f'_z z^{-b} - bfz^{-b-1} = \eta z^{-1} - b\xi z^{-1},$$

we obtain

$$\begin{aligned} f'' &= \frac{dp}{dz} = z^{1-b}\eta'_z + \eta(b-1)z^{b-2} = z^{1-b}\frac{d\eta}{d\xi}\frac{d\xi}{dz} + \eta(b-1)z^{b-2} \\ &= z^{1-b}\frac{d\eta}{d\xi}(\eta z^{-1} - b\xi z^{-1}) + \eta(b-1)z^{b-2} = z^{b-2}\left(\frac{d\eta}{d\xi}(\eta - b\xi) + (b-1)\eta\right) \end{aligned}$$

so that

$$\frac{d\eta}{d\xi} = \frac{H(\xi,\eta) - (b-1)\eta}{\eta - b\xi},$$

that is exactly (9.5.38). ■

Returning to our problem, we take the stretching transformation $s = \epsilon x, \quad g = \epsilon^b y$ and substitute it to (9.5.35). We get

$$g''_{ss} - g(g - sg') = \epsilon^{b-2} y''_{zz} - \epsilon^{2b} y(y - zy'_z)$$

that gives invariance if $b = -2$. Hence, introducing new variables

$$\xi = z^2 y, \qquad \eta = z^3 y' \tag{9.5.40}$$

we obtain directly

$$\frac{d^2y}{dz^2} = \eta_z z^{-3} - 3\eta z^{-4} = \frac{d\eta}{d\xi}\frac{d\xi}{dz} - 3\eta z^{-4} = z^{-4}\left(\frac{d\eta}{d\xi}(\eta + 2\xi) - 3\eta\right)$$

so that

$$\frac{d\eta}{d\xi} = \frac{\xi^2 - \xi\eta + 3\eta}{\eta + 2\xi}. \tag{9.5.41}$$

This is first order non-autonomous equation that still cannot be solved. However, we can obtain some information on the behaviour of the solution by re-writing it as an autonomous system and performing phase plane analysis. To identify the proper parameter, we write

$$\frac{d\xi}{dz} = z^{-1}(\eta + 2\xi)$$

and, using (9.5.41)

$$\frac{d\eta}{dz} = 3z^2 y' + z^3 y'' = 3z^{-1}\eta + z^3(y^2 - zyy') = z^{-1}(3\eta + \xi^2 - \xi\eta)$$

Thus, introducing a new variable via $d\tau/dz = 1/z$, that is, $\tau = \ln z$, we have

$$\begin{aligned} \xi'_\tau &= \eta + 2\xi, \\ \eta'_\tau &= \xi^2 - \xi\eta + 3\eta, \end{aligned} \tag{9.5.42}$$

Moreover, we see that as $z \to 0^+$, $\tau \to -\infty$ and as $z \to +\infty$, $\tau \to +\infty$, thus the boundary conditions (9.5.33) and (9.5.34) can be translated into conditions at $\pm\infty$. Firstly, we observe that (9.5.33) implicitly imposes the condition of boundedness on f at 0 so that $\xi = \eta = 0$ at $s = 0$ that is at $\tau = -\infty$.

If we look at the equilibrium points of (9.5.42), then we obtain $(0, 0)$ and $(2, -4)$. The Jacobi matrix of the right-hand side is

$$J(\xi, \eta) = \begin{pmatrix} 2 & 1 \\ 2\xi - \eta & 3 - \xi \end{pmatrix}.$$

Eigenvalues at $(0, 0)$ are calculated from

$$\begin{vmatrix} 2 - \lambda & 1 \\ 0 & 3 - \lambda \end{vmatrix} = (2 - \lambda)(3 - \lambda) = 0$$

hence $\lambda_\pm = 2, 3$ and $(0, 0)$ is an unstable node (source) which is consistent with the requirement $(\xi(\tau), \eta(\tau)) \to (0, 0)$ as $\tau \to \infty$.

Eigenvalues at $(2, -4)$ are calculated from

$$\begin{vmatrix} 2 - \lambda & 1 \\ 8 & 1 - \lambda \end{vmatrix} = \lambda^2 - 3\lambda - 6 = 0$$

hence $\lambda_\pm = \frac{2 \pm \sqrt{33}}{2}$ and since $\lambda_\pm$ are real with opposite sign, $(2, -4)$ is a saddle. It is important to note that the tangent of the stable manifold at $(2, -4)$ is $\lambda_+ - 2$.

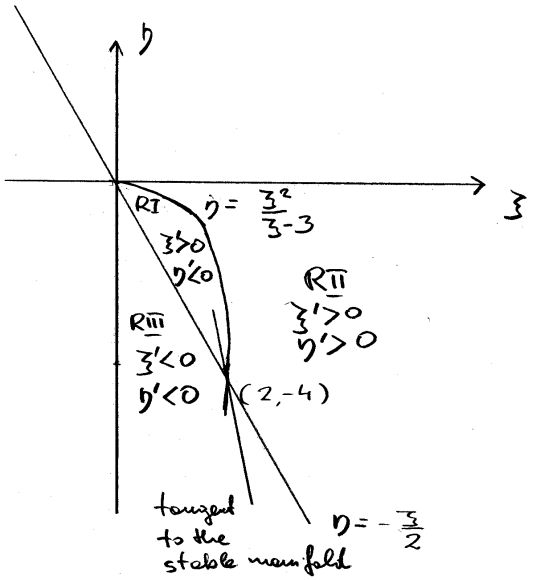

Fig. 6.2 Phase plane for the system (9.5.42).

The isoclines are $\eta = -2\xi$ and $\eta = \xi^2/\xi - 3$ and since tangents of these isoclines at $(2,-4)$ are -2 and -8 respectively ($\eta' = (\xi^2 - 6\xi)/(\xi - 3)^2$ at $\xi = 3$), we see that the stable manifold is between these two tangents, as seen in Fig. 6.2 Now, we observe that in the bounded region RI between isoclines we have $\xi' > 0, \eta' < 0$, in the region RII immediately above $\xi' > 0, \eta' > 0$ and in $RIII$ immediately below $\xi' < 0, \eta' < 0$.

For the phase-plane analysis, it may be easier to change the direction of τ defining $\bar{\tau} = -\tau$ and consequently $\bar{\xi}(\bar{\tau}) = \xi(\tau)$ and $\bar{\eta}(\bar{\tau}) = \eta(\tau)$ so that the stable manifold at $(2,-4)$ becomes the unstable manifold and the source at $(0,0)$ becomes a sink. Then in RI we have $\bar{\xi}' < 0$, $\bar{\eta}' > 0$, in RII $\bar{\xi}' < 0$ and $\bar{\eta}' < 0$, and in $RIII$ $\bar{\xi}' > 0$ and $\bar{\eta}' < 0$. In these new variables, there is a unique trajectory starting at $(2,-4)$ at $\bar{\tau} = -\infty$ and entering RI. If this trajectory was to leave RI through the isocline $\bar{\eta}' = 0$, that is, into RII, then the intercept would be a local maximum of $\bar{\eta}$ with $\bar{\xi}$ still moving to the left. This is, however, impossible, as the isocline is a decreasing function. Similarly, if the trajectory was to leave through the isocline $\bar{\xi}' = 0$, then at the intercept $\bar{\xi}$ would have local minimum, with $\bar{\eta}$ still moving up. This is again impossible as the isocline is a decreasing function. Thus, the trajectory must stay in RI and since $\bar{\xi}$ and $\bar{\eta}$ are monotonic functions, it must converge to $(0,0)$.

Returning to the old variable, we have proved the existence of a solution (ξ, η) such that $(\xi(z), \eta(z)) \to (0,0)$ as $z \to 0$ and $(\xi(z), \eta(z)) \to (2,-4)$ as $z \to \infty$. In particular,

$$\xi(z) \sim \frac{2}{z^2}, \qquad z \to \infty,$$

and going back to the original problem (9.5.23) we have determined the asymptotic behaviour of the solution $u(x,t)$

$$u(x,t) = t^{1/3} y\left(\frac{x}{\sqrt{3}t^{1/3}}\right) \sim \frac{6t}{x^2} \tag{9.5.43}$$

for large $x/t^{1/3}$.

5.3 The Burgers equation

The Burgers equation is a fundamental partial differential equation of mathematical physics. It occurs in various areas of applied sciences, such as the traffic flow or fluid dynamics. In the latter it can be considered

a prototype version of the Euler and Navier-Stokes systems of equations. There are two typical forms of the Burgers equation: the inviscid form

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, t > 0$$

and the viscous form

$$u_t + uu_x = \nu u_{xx}, \tag{9.5.44}$$

where the coefficient ν is called the viscosity. Both equations are examples of the conservation law with the flux density $\Phi(x,u) = \frac{1}{2}u^2$ in the first case and $\Phi_\nu(x,u) = \frac{1}{2}u^2 - \nu u_x$.

If we remember that the coefficient at the u_x derivative of the transport equation, here u, is the speed of the transport of the described substance of the density u, we see that in the Burgers equation the speed increases with the density. This has a build-up effect leading to creation of the so-called shock waves. We will see how they arise in the subsection below. On the other hand, the diffusive term νu_{xx} describes a dissipation of the substance, counteracting the effect of transport term. This will be demonstrated in the next subsection as well as in Example 2.3.

Inviscid Burgers equation and the formation of a shock wave

Consider the equation

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, t > 0,$$

with the initial condition

$$u(x) = \begin{cases} 2 & \text{for} \quad x < 0, \\ 2 - x & \text{for} \quad 0 \le x \le 1, \\ 1 & \text{for} \quad x > 1 \end{cases}$$

The characteristic system is

$$\begin{aligned} \frac{dv}{dt} &= 0, \\ \frac{dx}{dt} &= v, \end{aligned} \tag{9.5.45}$$

with the initial data $v(0) = u(\xi)$ and $x(0) = \xi$. From the first equation we obtain $v(t,\xi) = u(\xi)$; that is, u is constant along each characteristic, $u(x(t,\xi),t) = v(t,\xi) = u(\xi)$. But then, for a given ξ, the characteristic emanating from ξ satisfies

$$\frac{dx}{dt} = u(\xi), \qquad x(0) = \xi;$$

that is,

$$x = tu(\xi) + \xi.$$

In other words, the characteristics are straight lines, with the slope (x against t) equal to the initial value of the solution at the characteristic's intercept with the x axis, ξ and the solution is constant along each such line. To find the solution at a particular point (x,t), we see that the characteristic system (9.5.45) is equivalent to the system of algebraic equations

$$u(x,t) = u(\xi), \qquad x = tu(\xi) + \xi \tag{9.5.46}$$

and the solution u is obtained eliminating, if possible, the parameter ξ from (9.5.46).

To find the solution of our particular problem, we see that for $\xi < 0$ the characteristics lines are $t = (x-\xi)/2$ and for $\xi > 1$ the lines are $t = x - \xi$. For $0 \le \xi \le 1$ the equation of the characteristic is

$$t = \frac{1}{2-\xi}(x - \xi)$$

and we see that all these lines pass through the point $(x,t) = (2,1)$. This means that the solution $(2,1)$ cannot exist as it should be equal to the value carried by each characteristic, and each characteristic carries

different value of the solution. Thus, the smooth solution cannot continue beyond $t = 1$. We call $t = 1$ the *breaking time* of the wave.

To find the solution for $t < 1$, we first note that $u(x,t) = 2$ for $x < 2t$ and $u(x,t) = 1$ for $x > t + 1$. For $2t < x < t + 1$, the second equation in (9.5.46) becomes

$$x - \xi = (2 - \xi)t$$

that gives

$$\xi = \frac{x - 2t}{1 - t}$$

and the first equation gives then

$$u(x,t) = u_0(\xi) = 2 - \frac{x - 2t}{1 - t} = \frac{2 - x}{1 - t},$$

valid for $2t < x < t + 1, t < 1$. The explicit form of the solution also indicates the difficulty at the breaking time $t = 1$.

The Cole-Hopf transformation and analytic solution to the viscous Burgers equation

Let us consider the simplified version of the Burgers equation

$$u_t + uu_x - u_{xx} = 0. \tag{9.5.47}$$

It can be proved that the Burgers equation is equivalent to the linear diffusion equation. First we introduce a new function by

$$u = w_x.$$

Then (9.5.47) takes the form

$$w_{tx} + w_x w_{xx} - w_{xxx} = w_{tx} + \frac{1}{2}(w_x)_x^2 - w_{xxx} = 0$$

which can be integrated with respect to x to give

$$w_t + \frac{1}{2}w_x^2 - w_{xx} = c(t)$$

where $x(t)$ is a function of time t. Next we introduce a new function v by the formula

$$w = -2 \ln v.$$

Differentiation gives

$$w_t = -2\frac{v_t}{v}, \quad w_x = -2\frac{v_x}{v}, \quad w_{xx} = -2\frac{v_{xx}}{v} + 2\frac{v_x^2}{v^2}.$$

Hence

$$w_t + \frac{1}{2}w_x^2 - w_{xx} = -2\frac{v_t}{v} + 2\frac{v_x^2}{v^2} + 2\frac{v_{xx}}{v} - 2\frac{v_x^2}{v^2} = -2\frac{v_t}{v} + 2\frac{v_{xx}}{v} = c(t)$$

which is the diffusion equation (with a growth/decay term).

$$v_t - v_{xx} = -2c(t)v.$$

If we denote $C(t) = \int c(t)dt$ then, using the integrating factor, we can write the above as the standard diffusion equation

$$(ve^{C(t)})_t - (ve^{C(t)})_{xx} = 0.$$

Summarizing, if we can find a solution $V(t,x) = ve^{C(t)}$ of the diffusion equation, then

$$u(x,t) = -2\frac{\partial}{\partial x} \ln \left(v(x,t)e^{C(t)} \right) = -2\frac{v_x(x,t)}{v(x,t)} \tag{9.5.48}$$

is a solution to the Burgers equation. We see, that the arbitrary function $c(t)$ does not enter into the solution u and thus we can focus on solving just

$$v_t - v_{xx} = 0.$$

So, any nonzero solution of the diffusion equation generates a unique solution of the Burgers equation. Conversely, if u is a solution to the Burgers equation, then the formula (9.5.48) determines v according to

$$v(x,t) = \phi(t) e^{-\frac{1}{2}\int_0^x u(s,t)ds}$$

where ϕ is an arbitrary function of t. Differentiating, we obtain

$$v_t(x,t) = \phi'(t) e^{-\frac{1}{2}\int_0^x u(s,t)ds} - \frac{1}{2}\phi(t)\int_0^x u_t(s,t)ds e^{-\int_0^x u(s,t)ds} = \frac{\phi'(t)}{\phi(t)} v(x,t) - \frac{1}{2} v(x,t)\int_0^x u_t(s,t)ds$$

and, similarly,

$$v_x = -\frac{1}{2}uv, \qquad v_{xx} = -\frac{1}{2}u_x v + \frac{1}{4}u^2 v.$$

Consequently,

$$v_t - v_{xx} = \frac{1}{2}\frac{\phi'}{\phi} v - \frac{1}{2} v\left(\int_0^x u_t(s,t)ds - u_x + \frac{1}{2}u^2\right)$$

where the expression in brackets is the integral with respect to x of the left hand side of (9.5.47), which is an arbitrary function of t. Hence, any solution to the Burgers equation generates a solution to the diffusion equation

$$v_t - v_{xx} = \phi(t) v$$

for some function $\phi(t)$.

As our main interest is finding the solution to the Burgers equation, we write down the analytical formula for the solution to the initial value problem

$$u_t + uu_x - u_{xx} = 0, \qquad u(x,0) = u(x). \tag{9.5.49}$$

Using (9.5.48), the initial condition (9.5.49) on u can be obtained from the initial condition on v given by

$$v(x) = e^{-\frac{1}{2}\int_0^x u(s)ds}$$

The solution of the initial value problem for the diffusion equation is given by (9.2.16)

$$v(x,t) = \frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} e^{-\frac{(x-\xi)^2}{4t}} v(\xi)d\xi.$$

Therefore

$$v_x(x,t) = -\frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} \frac{(x-\xi)}{2t} e^{-\frac{(x-\xi)^2}{4t}} v(\xi)d\xi$$

and

$$u(x,t) = -2\frac{v_x(x,t)}{v(x,t)} = -\frac{\int_{-\infty}^{\infty} \frac{(x-\xi)}{t} e^{-\frac{(x-\xi)^2}{4t}} v(\xi)d\xi}{\int_{-\infty}^{\infty} e^{-\frac{(x-\xi)^2}{4t}} v(\xi)d\xi}$$

or, introducing the kernel

$$G(x,\xi,t) = -\frac{1}{2}\int_0^{\xi} u(s)ds - \frac{(x-\xi)^2}{4t}$$

we obtain the formula explicitly involving u as

$$u(x,t) = -\frac{\int_{-\infty}^{\infty} \frac{(x-\xi)}{t} e^{G(x,\xi,t)} d\xi}{\int_{-\infty}^{\infty} e^{G(x,\xi,t)} d\xi}.$$

Remark 5.1 The considerations above have been done for a spacial case of the Burgers equation with $\nu = 1$. If we have the general form of the equation

$$u_t + auu_x = \nu u_{xx}$$

then the substitution $u(x,t) = \nu a^{-1} U(x, \nu t)$ gives the simplified equation

$$U_t + UU_x = U_{xx}.$$

Example 5.1 Find the solution to the following initial value problem

$$u_t + uu_x = 2u_{xx}, \qquad u(x,0) = 2x.$$

Introducing a new unknown function $u(x,t) = 2U(x,2t)$ we see that $u_t = 4U_\tau, u_x = 2U_x, u_{xx} = 2U_{xx}$ and thus have

$$0 = u_t + uu_x = 2u_{xx} = 4U_\tau + 4UU_x - 4U_{xx}$$

so that U satisfies the simplified Burgers equation with the initial condition $U(x,0) = x$. We can then apply the transformation

$$U = -2\frac{V_x}{V}$$

to get the diffusion equation

$$V_t - V_{xx} = 0$$

subject to the initial condition

$$\frac{x}{2} = -2\frac{V_x(x,0)}{V(x,0)}$$

that is

$$V(x,0) = e^{-\frac{1}{4}x^2}.$$

To solve the diffusion equation with this initial condition we can use several methods. We present the solution by matching and by using the integral formula (9.2.16).

Method 1.

We recognize that $V(x,0)$ has the form similar to the fundamental solution of the diffusion equation (9.2.11)

$$S(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}}.$$

Since a solution of the diffusion equation multiplied by a constant is again a solution, we look for time t and constant c such that

$$cS(x,t) = e^{-\frac{1}{4}x^2},$$

which gives $t = 1$ and $c = 2\sqrt{\pi}$. Thus $2\sqrt{\pi}S(x,t)$ takes the value $V(x,0)$ at $t = 1$. Since the diffusion equation is invariant with respect to the shift of time, we obtain that

$$V(x,t) = 2\sqrt{\pi}S(x,t+1) = \frac{e^{-\frac{x^2}{4(t+1)}}}{\sqrt{t+1}}$$

is the sought solution. Consequently

$$U(x,t) = -2\frac{V_x}{V} = \frac{x}{t+1}$$

and

$$u(x,t) = 2U(x,2t) = \frac{2x}{2t+1}.$$

Method 2.
We have

$$V(x,t) = \frac{1}{2\sqrt{\pi}t} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4t}} e^{-\frac{y^2}{4}} dy.$$

Completing the quare in the exponent we get

$$\frac{x^2 - 2xy + y^2}{4t} + \frac{y^2}{4} = \frac{x^2 - 2xy + y^2 + ty^2}{4t} = \frac{(x - y(1+t))^2}{4t(1+t)} + \frac{x^2}{4(1+t)}$$

hence

$$V(x,t) = \frac{e^{-\frac{x^2}{4(t+1)}}}{2\sqrt{\pi}t} \int_{-\infty}^{\infty} e^{-\frac{(x-y(1+t))^2}{4t(1+t)}} dy.$$

Introducing the change of variable $z = (x - y(1+t))/2\sqrt{t(1+t)}$, so that $dy = -2\sqrt{t}dz/\sqrt{1+t}$, we obtain

$$V(x,t) = \frac{e^{-\frac{x^2}{4(t+1)}}}{\sqrt{\pi}(1+t)} \int_{-\infty}^{\infty} e^{-z^2} dz = \frac{e^{-\frac{x^2}{4(t+1)}}}{\sqrt{(1+t)}}.$$

The last step follows as before.