Ordinary Differential Equation

Alexander Grigorian
University of Bielefeld

Lecture Notes, April - July 2007

Contents

1 Introduction: the notion of ODEs and examples

A differential equation (*eine Differenzialgleichung*) is an equation for an unknown function that contains not only the function but also its derivatives. In general, the unknown function may depend on several variables and the equation may include various partial derivatives. However, in this course we consider only the differential equations for a function of a *single* real variable. Such equations are called *ordinary differential equations* - shortly ODE (*die gewöhnliche Differenzialgleichungen*). The theory of *partial* differential equations, that is, the equations containing partial derivatives, is a topic for a different lecture course.

A most general ODE has the form

$$F\left(x, y, y', ..., y^{(n)}\right) = 0 \tag{1.1}$$

where F is a given function of $n+2$ variables and $y = y(x)$ is an unknown function. The problem is usually to find a solution $y(x)$, possibly with some additional conditions, or to investigate various properties of a solution.

The *order* of an ODE is the maximal value of n such that the derivative $y^{(n)}$ is presented in the equation.

In Introduction we consider various examples and specific classes of ODEs of the first and second but then develop a general theory, which includes the existence and uniqueness results in rather general setting for an arbitrary order.

Consider the differential equation of the first order

$$y' = f(x, y), \tag{1.2}$$

where $y = y(x)$ is the unknown real-valued function of x and $f(x, y)$ is a given function of x, y.

The difference with (1.1) is that (1.2) is *resolved* with respect to y. Consider a couple (x, y) as a point in $\mathbb{R}^2$ and assume that function f is defined on a set $D \subset \mathbb{R}^2$, which is called the *domain* of the equation (1.2).

Definition. A real valued function $y(x)$ defined on an interval $I \subset \mathbb{R}$, is called a (*particular*) solution of (1.2) if $y(x)$ is differentiable at any $x \in I$, the point $(x, y(x))$ belongs to D for any $x \in I$ and the identity $y'(x) = f(x, y(x))$ holds for all $x \in I$.

The family of all solutions of (1.2) is called the *general* solution. The graph of a particular solution is called an *integral curve* of the equation. Note that any integral curve is contained in the domain D.

Typically, a given ODE cannot be solved explicitly. We'll consider below some classes of $f(x, y)$ when one find the general solution to (1.2) in terms of indefinite integration. Start with a simplest example.

Example. Assume that the function f does not depend on y so that (1.2) becomes $y' = f(x)$. Hence, y must be a primitive function of f. Assuming that f is a continuous function on an interval I, we obtain the general solution in the form

$$y = \int f(x)\,dx = F(x) + C,$$

where F is a primitive of $f(x)$ and C is an arbitrary constant.

1.1 Separable ODE

Consider a *separable ODE*, that is, an ODE of the form

$$y' = f(x)\, g(y). \tag{1.3}$$

It is called separable because the right hand side splits into the product of a function of x and a function of y.

Theorem 1.1 (The method of separation of variables) *Let $f(x)$ and $g(y)$ be continuous functions on intervals I and J, respectively, and assume that $g(y) \neq 0$ on J. Let $F(x)$ be a primitive function of $f(x)$ on I and $G(y)$ be a primitive function of $\frac{1}{g(y)}$ on J. Then a function $y : I \to J$ solves the differential equation* (1.3) *if and only if it satisfies the identity*

$$G(y(x)) = F(x) + C, \tag{1.4}$$

where C is any real constant.

Proof. Let $y(x)$ solve (1.3). Dividing (1.3) by $g(y)$ and integrating in x, we obtain

$$\int \frac{y'dx}{g(y)} = \int f(x)\, dx, \tag{1.5}$$

where we regard y as a function of x. Since $F(x)$ is a primitive of $f(x)$, we have

$$\int f(x)\, dx = F(x) + C'.$$

In the left hand side of (1.5), we have $y'dx = dy$. By the change of a variable in the indefinite integral (Theorem 1.4 from Analysis II), we obtain

$$\int \frac{y'dx}{g(y)} = \int \frac{dy}{g(y)} = G(y) + C'',$$

where in the middle integral y is considered as an independent variable. Combining the above lines, we obtain the identity (1.4) with $C = C' - C''$.

Conversely, let $y(x)$ be a function from I to J that satisfies (1.4). Since the function $g(y)$ does not change the sign, by the inverse function theorem (from Analysis I) function G has the inverse G^{-1}, whence we obtain from (1.4)

$$y(x) = G^{-1}(F(x) + C). \tag{1.6}$$

Since G^{-1} and F are differentiable functions, by the chain rule also y is differentiable. It follows from (1.4) by differentiation in x that

$$G'(y)\, y' = F'(x) = f(x).$$

Substituting here $G'(y) = \frac{1}{g(y)}$, we obtain (1.3). ■

Corollary. *Under the conditions of* Theorem 1.1, *for any $x_0 \in I$ and $y_0 \in J$ there is exactly one solution $y(x)$ of the equation* (1.3) *defined on I and such that $y(x_0) = y_0$.*

In other words, for every point $(x_0, y_0) \in I \times J$ there is exactly one integral curve of the ODE that goes through this point.

Proof. The identity (1.6) determines for any real C a particular solution $y(x)$ defined on the interval I. We only need to make sure that C can be chosen to satisfy the condition $y(x_0) = x_0$. Indeed, by (1.4), the latter condition is equivalent to $G(y_0) = F(x_0) + C$, which is true exactly for one value of the constant C, that is, for $C = G(y_0) - F(x_0)$, whence the claim follows. ■

Let us show some examples using this method.

Example. *(Heat conduction)* Let $x(t)$ denote the temperature of a body at time t and assume that the body is immersed into a media with a constant temperature T. Without sources and sinks of heat, the temperature of the body will over time tend to T. The exact temperature at time t can be determined by using the Fourier law of heat conductance: the rate of decrease of $x(t)$ is proportional to the difference $x(t) - T$, that is,

$$x'(t) = -k(x(t) - T),$$

where $k > 0$ is the coefficient of thermoconductance between the body and the media.

This equation is separable, and solving it in each of the domains $x > T$ or $x < T$, we obtain the identity

$$\int \frac{dx}{x-T} = -k \int dt,$$

$$\ln|x - T| = -kt + C,$$

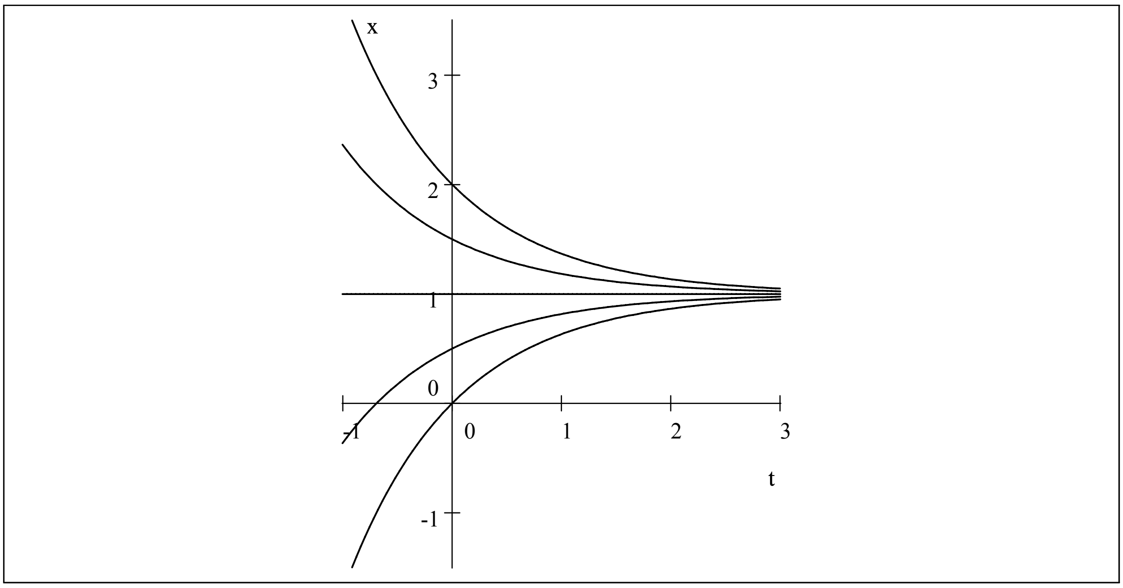
whence

$$|x - T| = e^C e^{-kt}.$$

Renaming $\pm e^C$ by C, we obtain the solution

$$x = T + Ce^{-kt}. \tag{1.7}$$

Note that by the definition of C, we have $C \neq 0$. More precisely, $C > 0$ correspond to a solution $x(t) > T$ and $C < 0$ corresponds to a solution $x(t) < T$. However, $C = 0$ gives also a solution $x \equiv T$, which was not accounted for by the above method. Hence, the identity (1.7) determines a solution to the given equation for all real C. Here are some integrals curves of this equation with $T = 1$ and $k = 1$:



The value of C can be found, for example, if one knows the initial value of the temperature that is $x(0) = x_0$. Setting $t = 0$ in (1.7), we obtain $x_0 = T + C$ whence $C = x_0 - T$. Hence, (1.7) becomes

$$x = T + (x_0 - T)\, e^{-kt}.$$

The value of k can be determined if one has one more measurement of $x(t)$ at some time $t > 0$.

Remark. If in the equation $y' = f(x)\, g(y)$ the function $g(y)$ vanishes at a sequence of points, say $y_1, y_2, ...,$ enumerated in the increasing order, then we have a family of constant solutions $y(x) = y_k$. The method of separation of variables provides solutions in any domain $y_k < y < y_{k+1}$. The integral curves in the domains $y_k < y < y_{k+1}$ can in general touch the constant solution, as will be shown in the next example.

Example. Consider the equation

$$y' = \sqrt{|y|},$$

which is defined for all $y \in \mathbb{R}$. Since the right hand side vanish for $y = 0$, the constant function $y \equiv 0$ is a solution. In the domains $y > 0$ and $y < 0$, the equation can be solved using separation of variables. For example, in the domain $y > 0$, we obtain

$$\int \frac{dy}{\sqrt{y}} = \int dx$$

whence

$$2\sqrt{y} = x + C$$

and

$$y = \frac{1}{4}\,(x + C)^2\,,\;\; x > -C.$$

Similarly, in the domain $y < 0$, we obtain

$$\int \frac{dy}{\sqrt{-y}} = \int dx$$

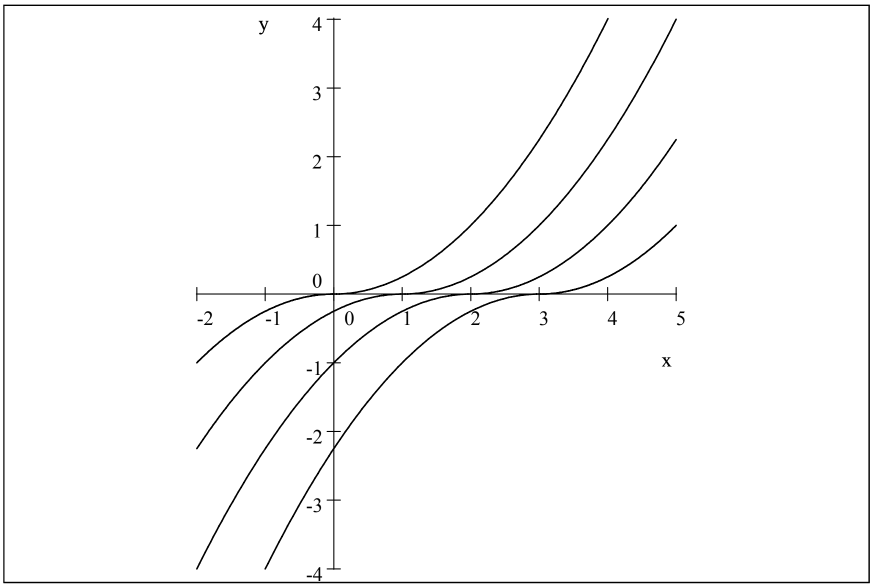
whence

$$-2\sqrt{-y} = x + C$$

and

$$y = -\frac{1}{4}(x + C)^2, \ x < -C.$$

We obtain the following integrals curves:



We see that the integral curves in the domain $y > 0$ touch the curve $y = 0$ and so do the integral curves in the domain $y < 0$. This allows us to construct more solution as follows: take a solution $y_1(x) < 0$ that vanishes at $x = a$ and a solution $y_2(x) > 0$ that vanishes at $x = b > a$. Then define a new solution:

$$y(x) = \begin{cases} y_1(x), & x < a \\ 0, & a \le x \le b, \\ y_2(x), & x > b. \end{cases}$$

Such solutions are not obtained automatically by the method of separation of variables.

1.2 Linear ODE of 1st order

Consider the ODE of the form

$$y' + a(x)\,y = b(x) \tag{1.8}$$

where a and b are given functions of x, defined on a certain interval I. This equation is called *linear* because it depends linearly on y and y'.

A linear ODE can be solved as follows.

Theorem 1.2 (The method of variation of parameter) *Let functions $a(x)$ and $b(x)$ be continuous in an interval I. Then the general solution of the linear ODE* (1.8) *has the form*

$$y(x) = e^{-A(x)} \int b(x)\, e^{A(x)} dx, \tag{1.9}$$

where $A(x)$ is a primitive of $a(x)$ on I.

Note that the function $y(x)$ given by (1.9) is defined on the full interval I.

Proof. Let us make the change of the unknown function $u(x) = y(x)\, e^{A(x)}$, that is,

$$y(x) = u(x)\, e^{-A(x)}. \tag{1.10}$$

Substituting this to the equation (1.8) we obtain

$$\left(ue^{-A}\right)' + aue^{-A} = b,$$

$$u'e^{-A} - ue^{-A}A' + aue^{-A} = b.$$

Since $A' = a$, we see that the two terms in the left hand side cancel out, and we end up with a very simple equation for $u(x)$:

$$u'e^{-A} = b$$

whence $u' = be^{A}$ and

$$u = \int be^{A} dx.$$

Substituting into (1.10), we finish the proof. ∎

One may wonder how one can guess to make the change (1.10). Here is the motivation. Consider first the case when $b(x) \equiv 0$. In this case, the equation (1.8) becomes

$$y' + a(x)\,y = 0$$

and it is called *homogeneous*. Clearly, the homogeneous linear equation is separable. In the domains $y > 0$ and $y < 0$ we have

$$\frac{y'}{y} = -a(x)$$

and

$$\int \frac{dy}{y} = -\int a(x)\, dx = -A(x) + C.$$

Then $\ln|y| = -A(x) + C$ and

$$y(x) = Ce^{-A(x)}$$

where C can be any real (including $C = 0$ that corresponds to the solution $y \equiv 0$).

For a general equation (1.8) take the above solution to the homogeneous equation and replace a constant C by a function $u(x)$, which will result in the above change. Since we have replaced a constant parameter by a function, this method is called the method of variation of parameter. It applies to the linear equations of higher order as well.

Corollary. *Under the conditions of* Theorem 1.2, *for any $x_0 \in I$ and any $y_0 \in \mathbb{R}$ there is exists exactly one solution $y(x)$ defined on I and such that $y(x_0) = y_0$.*

That is, though any point $(x_0, y_0) \in I \times \mathbb{R}$ there goes exactly one integral curve of the equation.

Proof. Let $B(x)$ be a primitive of be^{-A} so that the general solution can be written in the form

$$y = e^{-A(x)}(B(x) + C)$$

with an arbitrary constant C. Obviously, any such solution is defined on I. The condition $y(x_0) = y_0$ allows to uniquely determine C from the equation:

$$C = y_0 e^{A(x_0)} - B(x_0),$$

whence the claim follows. ■

Example. Consider the equation

$$y' + \frac{1}{x}y = e^{x^2}$$

in the domain $x > 0$. Then

$$A(x) = \int a(x)\,dx = \int \frac{dx}{x} = \ln x$$

(we do not add a constant C since $A(x)$ is *one* of the primitives of $a(x)$),

$$y(x) = \frac{1}{x}\int e^{x^2} x dx = \frac{1}{x}\left(\frac{1}{2}e^{x^2} + C\right) = \frac{1}{2x}e^{x^2} + \frac{C}{x},$$

where C is an arbitrary constant.

1.3 Exact differential forms

Let $F(x,y)$ be a real valued function defined in an open set $\Omega \subset \mathbb{R}^2$. Recall that F is differentiable at a point $(x,y) \in \Omega$ if there exists a matrix A of dimension 1×2 (called the full derivative of F) full such that

$$F(x + dx, y + dy) - F(x,y) = A\begin{pmatrix} dx \\ dy \end{pmatrix} + o(|dx| + |dy|)$$

as $|dx| + |dy| \to 0$. Here we denote by dx and dy the increments of x and y, respectively, which are considered as new independent variables. The linear function $\binom{dx}{dy} \mapsto A\binom{dx}{dy}$ is called the differential of F and is denoted by dF. Let $A = (a \;\; b)$ so that

$$dF = adx + bdy.$$

If F is differentiable at any point $(x, y) \in \Omega$ then a and b are functions of (x, y). Recall also that $a = F_x$ and $b = F_y$.

Definition. Given two functions $a(x, y)$ and $b(x, y)$ in Ω, consider the expression

$$a(x, y)\, dx + b(x, y)\, dy,$$

which is called a *differential form*. The differential form is called *exact* in Ω if there is a differentiable function F in Ω such that

$$dF = adx + bdy, \tag{1.11}$$

and *inexact* otherwise. If the form is exact then the function F from (1.11) is called an *integral* of the form.

Observe that not every differential form is exact as one can see from the following claim.

Claim. *If functions a, b belong to $C^1(\Omega)$ then the necessary condition for the form $adx + bdy$ to be exact is $a_y = b_x$.*

Proof. Indeed, if there is F is an integral of the form $adx + bdy$ then $F_x = a$ and $F_y = b$, whence it follows that $F \in C^2(\Omega)$. Then $F_{xy} = F_{yx}$, which implies $a_y = b_x$. ■

Example. The form $ydx - xdy$ is not exact because $a_y = 1$ while $b_x = -1$.

The form $ydx + xdy$ is exact because it has an integral $F(x, y) = xy$.

The form $2xydx + (x^2 + y^2)\, dy$ is exact because it has an integral $F(x, y) = x^2y + \frac{y^3}{3}$ (it will be explained later how one can obtain an integral).

If the differential form $adx + bdy$ is exact then this allows to solve easily the following differential equation:

$$a(x, y) + b(x, y)\, y' = 0. \tag{1.12}$$

This ODE is called *quasi-linear* because it is linear with respect to y' but not necessarily linear with respect to y. One can write (1.12) in the form

$$a(x, y)\, dx + b(x, y)\, dy = 0,$$

which explains why the equation (1.12) is related to the differential form $adx + bdy$. We say that the equation (1.12) is exact if the form $adx + bdy$ is exact.

Theorem 1.3 *Let Ω be an open subset of $\mathbb{R}^2$, a, b be continuous functions on Ω, such that the form $adx + bdy$ is exact. Let F be an integral of this form. Consider a differentiable function $y(x)$ defined on an interval $I \subset \mathbb{R}$ such that the graph of y is contained in Ω. Then y solves the equation* (1.12) *if and only if*

$$F(x, y(x)) = \text{const} \ \ \textit{on } I.$$

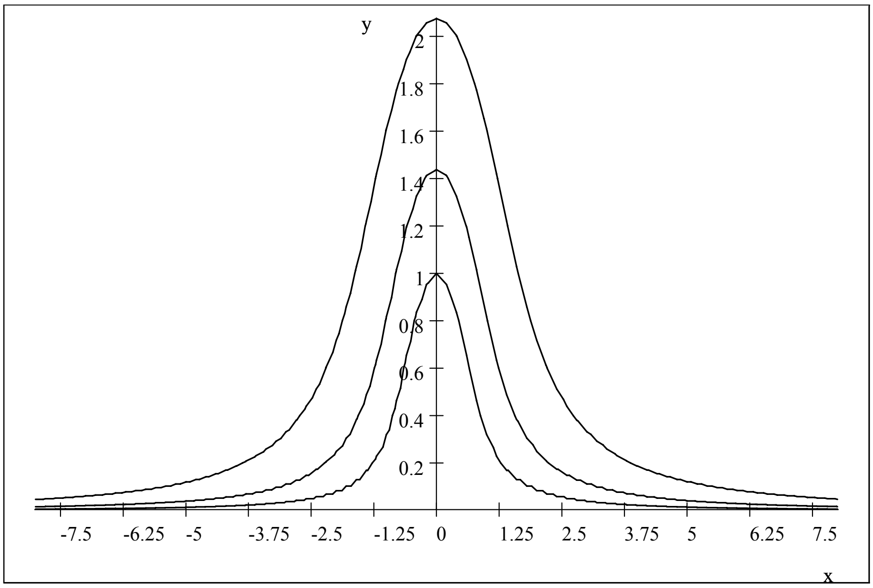
Proof. The hypothesis that the graph of $y(x)$ is contained in Ω implies that the composite function $F(x, y(x))$ is defined on I. By the chain rule, we have

$$\frac{d}{dx} F(x, y(x)) = F_x + F_y y' = a + by'.$$

Hence, the equation $a + by' = 0$ is equivalent to $\frac{d}{dx}F(x, y(x)) = 0$, and the latter is equivalent to $F(x, y(x)) = \text{const}$. ■

Example. The equation $y + xy' = 0$ is exact and is equivalent to $xy = C$ because $ydx + xdy = d(xy)$. The same can be obtained using the method of separation of variables.

The equation $2xy + (x^2 + y^2)\,y' = 0$ is exact and is equivalent to $x^2y + \frac{y^3}{3} = C$. Below are some integral curves of this equation:



How to decide whether a given differential form is exact or not? A partial answer is given by the following theorem.

We say that a set $\Omega \subset \mathbb{R}^2$ is a *rectangle* (box) if it has the form $I \times J$ where I and J are intervals in $\mathbb{R}$.

Theorem 1.4 (The Poincaré lemma) *Let Ω be an open rectangle in $\mathbb{R}^2$. Let a, b be functions from $C^1(\Omega)$ such that $a_y \equiv b_x$. Then the differential form $adx + bdy$ is exact in Ω.*

Let us first prove the following lemma, which is of independent interest.

Lemma 1.5 *Let $g(x, t)$ be a continuous function on $I \times J$ where I and J are bounded closed intervals in $\mathbb{R}$. Consider the function*

$$f(x) = \int_\alpha^\beta g(x, t)\,dt,$$

where $[\alpha, \beta] = J$, which is defined for all $x \in I$. If the partial derivative g_x exists and is continuous on $I \times J$ then f is continuously differentiable on I and, for any $x \in I$,

$$f'(x) = \int_\alpha^\beta g_x(x, t)\,dt.$$

In other words, the operations of differentiation in x and integration in t, when applied to $g\left(x,t\right)$, are interchangeable.

Proof. We need to show that, for all $x \in I$,

$$\frac{f\left(x',t\right)-f\left(x,t\right)}{x'-x} \rightarrow \int_{\alpha}^{\beta} g_x\left(x,t\right) dt \text{ as } x' \rightarrow x,$$

which amounts to

$$\int_{\alpha}^{\beta} \frac{g\left(x',t\right)-g\left(x,t\right)}{x'-x} dt \rightarrow \int_{\alpha}^{\beta} g_x\left(x,t\right) dt \text{ as } x' \rightarrow x.$$

Note that by the definition of a partial derivative, for any $t \in [\alpha, \beta]$,

$$\frac{g\left(x',t\right)-g\left(x,t\right)}{x'-x} \rightarrow g_x\left(x,t\right) \text{ as } x' \rightarrow x. \tag{1.13}$$

Consider all parts of (1.13) as functions of t, with fixed x and with x' as a parameter. Then we have a convergence of a sequence of functions, and we would like to deduce that their integrals converge as well. By a result from Analysis II, this is the case, if the convergence is *uniform* in the whole interval $[\alpha, \beta]$, that is, if

$$\sup_{t\in[\alpha,\beta]} \left| \frac{g\left(x',t\right)-g\left(x,t\right)}{x'-x} - g_x\left(x,t\right) \right| \rightarrow 0 \quad \text{as } x' \rightarrow x. \tag{1.14}$$

By the mean value theorem, for any $t \in [\alpha, \beta]$, there is $\xi \in [x, x']$ such that

$$\frac{g\left(x',t\right)-g\left(x,t\right)}{x'-x} = g_x\left(\xi,t\right).$$

Hence, the difference quotient in (1.14) can be replaced by $g_x\left(\xi,t\right)$. To proceed further, recall that a continuous function on a compact set is uniformly continuous. In particular, the function $g_x\left(x,t\right)$ is uniformly continuous on $I \times J$, that is, for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$x, x' \in I, |x-x'| < \delta \text{ and } t, t' \in J, |t-t'| < \delta \ \Rightarrow\ |g_x\left(x,t\right) - g_x\left(x',t'\right)| < \varepsilon. \tag{1.15}$$

If $|x-x'| < \delta$ then also $|x-\xi| < \delta$ and by (1.15)

$$|g_x\left(\xi,t\right) - g_x\left(x,t\right)| < \varepsilon \text{ for all } t \in J.$$

In other words, $|x-x'| < \delta$ implies that

$$\sup_{t\in J} \left| \frac{g\left(x',t\right)-g\left(x,t\right)}{x'-x} - g_x\left(x,t\right) \right| \leq \varepsilon,$$

whence (1.14) follows. ■

$\boxed{\text{Lecture 3}}$ 10.04.2007

Proof of Theorem 1.4. Assume first that the integral F exists and $F(x_0, y_0) = 0$ for some point $(x_0, y_0) \in \Omega$ (the latter can always be achieved by adding a constant to F). For any point $(x, y) \in \Omega$, also the point $(x, y_0) \in \Omega$; moreover, the intervals $[(x_0, y_0), (x, y_0)]$ and $[(x, y_0), (x, y)]$ are contained in Ω because Ω is a rectangle. Since $F_x = a$ and $F_y = b$, we obtain by the fundamental theorem of calculus that

$$F(x, y_0) = F(x, y_0) - F(x_0, y_0) = \int_{x_0}^{x} F_x(s, y_0)\, ds = \int_{x_0}^{x} a(s, y_0)\, ds$$

and

$$F(x, y) - F(x, y_0) = \int_{y_0}^{y} F_y(x, t)\, dt = \int_{y_0}^{y} b(x, t)\, dt,$$

whence

$$\boxed{F(x, y) = \int_{x_0}^{x} a(s, y_0)\, ds + \int_{y_0}^{y} b(x, t)\, dt.} \tag{1.16}$$

Now forget about this argument and just define function $F(x, y)$ by (1.16). Let us show that F is indeed the integral of the form $adx + bdy$. It suffices to verify that $F_x = a$ and $F_y = b$ because then we can conclude that $F \in C^1(\Omega)$ (and even $F \in C^2(\Omega)$) and, hence,

$$dF = F_x dx + F_y dy = adx + bdy.$$

It is easy to see from (1.16) that $F_y = b(x, y)$. Let us show that $F_x = a(x, y)$. Indeed, using Lemma 1.5 and the hypothesis $a_y = b_x$, we obtain

$$\begin{aligned} F_x &= \frac{d}{dx} \int_{x_0}^{x} a(s, y_0)\, ds + \frac{d}{dx} \int_{y_0}^{y} b(x, t)\, dt \\ &= a(x, y_0) + \int_{y_0}^{y} b_x(x, t)\, dt \\ &= a(x, y_0) + \int_{y_0}^{y} a_y(x, t)\, dt \\ &= a(x, y_0) + (a(x, y) - a(x, y_0)) \\ &= a(x, y). \end{aligned}$$

Hence, we have shown that $F_x = a$ and $F_y = b$, which was to be proved. ■

Example. Consider again the differential from $2xydx + (x^2 + y^2)\, dy$ in $\Omega = \mathbb{R}^2$. Since

$$a_y = (2xy)_y = 2x = \left(x^2 + y^2\right)_x = b_x,$$

we conclude by Theorem 1.4 that the given form is exact. The integral F can be found by (1.16) taking $x_0 = y_0 = 0$:

$$F(x, y) = \int_0^x 2s0ds + \int_0^y \left(x^2 + t^2\right) dt = x^2 y + \frac{y^3}{3},$$

as it was observed above.

Example. Consider the differential form

$$\frac{-ydx+xdy}{x^2+y^2} \tag{1.17}$$

in $\Omega=\mathbb{R}^2\setminus\{0\}$. This form satisfies the condition $a_y=b_x$ because

$$a_y=-\left(\frac{y}{x^2+y^2}\right)_y=-\frac{(x^2+y^2)-2y^2}{(x^2+y^2)^2}=\frac{y^2-x^2}{(x^2+y^2)^2}$$

and

$$b_x=\left(\frac{x}{x^2+y^2}\right)_x=\frac{(x^2+y^2)-2x^2}{(x^2+y^2)^2}=\frac{y^2-x^2}{(x^2+y^2)^2}.$$

By Theorem 1.4 we conclude that the given form is exact in any rectangular domain in Ω. However, we'll show that the form is inexact in Ω.

Consider the function $\theta(x,y)$ which is the polar angle that is defined in the domain

$$\Omega'=\mathbb{R}^2\setminus\{(x,0):x\leq 0\}$$

by the conditions

$$\sin\theta=\frac{y}{r},\quad \cos\theta=\frac{x}{r},\quad \theta\in(-\pi,\pi),$$

where $r=\sqrt{x^2+y^2}$. Let us show that in Ω'

$$d\theta=\frac{-ydx+xdy}{x^2+y^2}. \tag{1.18}$$

In the half-plane $\{x>0\}$ we have $\tan\theta=\frac{y}{x}$ and $\theta\in(-\pi/2,\pi/2)$ whence

$$\theta=\arctan\frac{y}{x}.$$

Then (1.18) follows by differentiation of the arctan. In the half-plane $\{y>0\}$ we have $\cot\theta=\frac{x}{y}$ and $\theta\in(0,\pi)$ whence

$$\theta=\operatorname{arccot}\frac{x}{y}$$

and (1.18) follows again. Finally, in the half-plane $\{y<0\}$ we have $\cot\theta=\frac{x}{y}$ and $\theta\in(-\pi,0)$ whence

$$\theta=-\operatorname{arccot}\left(-\frac{x}{y}\right),$$

and (1.18) follows again. Since Ω' is the union of the three half-planes $\{x>0\}$, $\{y>0\}$, $\{y<0\}$, we conclude that (1.18) holds in Ω' and, hence, the form (1.17) is exact in Ω'.

Why the form (1.17) is inexact in Ω? Assume from the contrary that the form (1.17) is exact in Ω and that F is its integral in Ω, that is,

$$dF=\frac{-ydx+xdy}{x^2+y^2}.$$

Then $dF = d\theta$ in Ω' whence it follows that $d(F - \theta) = 0$ and, hence[1] $F = \theta + \text{const}$ in Ω'. It follows from this identity that function θ can be extended from Ω' to a continuous function on Ω, which however is not true, because the limits of θ when approaching the point $(-1, 0)$ (or any other point $(x, 0)$ with $x < 0$) from above and below are different.

The moral of this example is that the statement of Theorem 1.4 is not true for an arbitrary open set Ω. It is possible to show that the statement of Theorem 1.4 is true if and only if the set Ω is *simply connected*, that is, if any closed curve in Ω can be continuously shrunk to a point. Obviously, the rectangles are simply connected, while the set $\mathbb{R}^2 \setminus \{0\}$ is not.

1.4 Integrating factor

Consider again the quasilinear equation

$$a(x, y) + b(x, y) y' = 0 \tag{1.19}$$

and assume that it is *inexact*.

Write this equation in the form

$$adx + bdy = 0.$$

After multiplying by a non-zero function $M(x, y)$, we obtain equivalent equation

$$Madx + Mbdy = 0,$$

which may become exact, provided function M is suitably chosen.

Definition. A function $M(x, y)$ is called the *integrating factor* for the differential equation (1.19) in Ω if M is a non-zero function in Ω such that the form $Madx + Mbdy$ is exact in Ω.

If one has found an integrating factor then multiplying (1.19) by M we reduce the problem to the case of Theorem 1.3.

Example. Consider the ODE

$$y' = \frac{y}{4x^2 y + x},$$

and write it in the form

$$\frac{y}{4x^2 y + x} dx - dy = 0.$$

Clearly, this equation is not exact. However, multiplying by $4x^2 y + x$ and dividing by x^2, we obtain the equation

$$\frac{y}{x^2} dx - \left(4y + \frac{1}{x}\right) dy = 0,$$

[1] We use the following fact that is contained in Exercise 58 from Analysis II: if the differential of a function is identical zero in a connected open set $U \subset \mathbb{R}^n$ then the function is constant in this set. Recall that the set U is called connected if any two points from U can be connected by a polygonal line that is contained in U.

The set Ω' is obviously connected.

which is already exact in any rectangular domain because

$$\left(\frac{y}{x^2}\right)_y = \frac{1}{x^2} = -\left(4y+\frac{1}{x}\right)_x.$$

The integral of this form is obtained by (1.16) with $y_0 = 0$ and any $x_0 \neq 0$:

$$F(x,y) = \int_{x_0}^{x} \frac{y_0}{s^2} ds - \int_0^y \left(4t+\frac{1}{x}\right) dt = -2y^2 - \frac{y}{x}.$$

By Theorem 1.3, the general solution is given by the identity

$$2y^2 + \frac{y}{x} = C.$$

1.5 Second order ODE

A general second order ODE, resolved with respect to y'' has the form

$$y'' = f(x,y,y'),$$

where f is a given function of three variables and $y = y(x)$ is an unknown function. We consider here some problems that amount to a second order ODE.

1.5.1 Newtons' second law

Consider movement of a point particle along a straight line and let the coordinate at time t be $x(t)$. The velocity of the particle is $v(t) = x'(t)$ and the acceleration $a(t) = x''(t)$. The Newton's second law says that at any time $mx'' = F$ where m is the mass of the particle and F is the sum of all forces acting on the particle. In general, F may depend on t, x, x' so that we get a second order ODE for $x(t)$.

Assume that the force $F = F(x)$ depends only on the position x. Let U be a primitive function of $-F$; the function U is called the *potential* of the force F. Multiplying the equation $mx'' = F$ by x' and integrating in t, we obtain

$$m\int x''x'dt = \int F(x)\,x'dt,$$

$$\frac{m}{2}\int \frac{d}{dt}(x')^2\,dt = \int F(x)\,dx,$$

$$\frac{mv^2}{2} = -U(x) + C$$

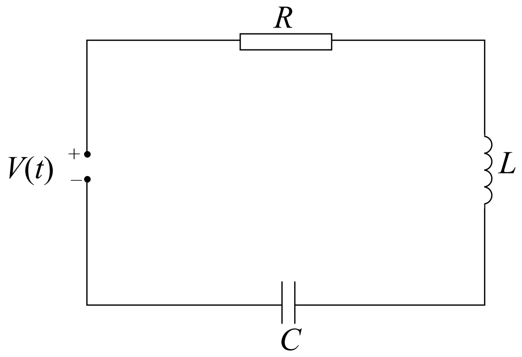
and

$$\frac{mv^2}{2} + U(x) = C.$$

The sum $\frac{mv^2}{2} + U(x)$ is called the *energy* of the particle (which is the sum of the kinetic energy and the potential energy). Hence, we have obtained the conservation law of the energy: the energy of the particle moving in a potential field remains constant.

1.5.2 Electrical circuit

Consider an RLC-circuit that is, an electrical circuit where a resistor, an inductor and a capacitor are connected in a series:



Denote by R the resistance of the resistor, by L the inductance of the inductor, and by C the capacitance of the capacitor. Let the circuit contain a power source with the voltage $V(t)$, where t is time. Denote by $I(t)$ the current in the circuit at time t. Using the laws of electromagnetism, we obtain that the potential difference v_R on the resistor R is equal to

$$v_R = RI$$

(Ohm's law), and the potential difference v_L on the inductor is equal to

$$v_L = L\frac{dI}{dt}$$

(Faraday's law). The potential difference v_C on the capacitor is equal to

$$v_C = \frac{Q}{C},$$

where Q is the charge of the capacitor; also we have $Q' = I$. By Kirchhoff's law, we have

$$v_R + v_L + v_C = V(t)$$

whence

$$RI + LI' + \frac{Q}{C} = V(t).$$

Differentiating in t, we obtain

$$LI'' + RI' + \frac{I}{C} = V', \tag{1.20}$$

which is a second order ODE with respect to $I(t)$. We will come back to this equation after having developed the theory of linear ODEs.

2 Existence and uniqueness theorems

2.1 1st order ODE

We change notation, denoting the independent variable by t and the unknown function by $x(t)$. Hence, we write an ODE in the form

$$x' = f(t,x),$$

where f is a real value function on an open set $\Omega \subset \mathbb{R}^2$ and a pair (t,x) is considered as a point in $\mathbb{R}^2$.

Let us associate with the given ODE the *initial value problem* (IVP), that is, the problem to find a solution that satisfies in addition the *initial condition* $x(t_0) = x_0$ where (t_0, x_0) is a given point in Ω. We write shortly IVP as follows:

$$\begin{cases} x' = f(t,x), \\ x(t_0) = x_0. \end{cases}$$

A solution to IVP is a differentiable function $x(t) : I \to \mathbb{R}$ where I is an open interval containing t_0, such that $(t, x(t)) \in \Omega$ for all $t \in I$, which satisfies the ODE in I and the initial condition. Geometrically, the graph of function $x(t)$ is contained in Ω and goes through the point (t_0, x_0).

In order to state the main result, we need the following definition.

Definition. We say that a function $f : \Omega \to \mathbb{R}$ is *locally Lipschitz* in x if, for any point $(t_0, x_0) \in \Omega$ there exist positive constants ε, δ, L such that the rectangle

$$R = [t_0 - \delta, t_0 + \delta] \times [x_0 - \varepsilon, x_0 + \varepsilon] \tag{2.1}$$

is contained in Ω and

$$|f(t,x) - f(t,y)| \le L|x - y|,$$

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$.

Lemma 2.1 *If the partial derivative f_x exists and is continuous in Ω then f is locally Lipschitz in Ω.*

Proof. Fix a point $(t_0, x_0) \in \Omega$ and choose positive ε, δ so that the rectangle R defined by (2.1) is contained in Ω (which is possible just because Ω is an open set). Then, for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$, we have by the mean value theorem

$$f(t,x) - f(t,y) = f_x(t,\xi)(x - y),$$

for some $\xi \in [x, y]$. Since R is a bounded closed set and f_x is continuous on R, the maximum of $|f_x|$ on R exists, so that

$$L := \sup_R |f_x| < \infty.$$

Since $(t,\xi) \in R$, we obtain $|f_x(t,\xi)| \leq L$ and, hence,

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

which finishes the proof. ∎

The next theorem is one of the main results of this course.

Theorem 2.2 (The Picard - Lindelöf theorem) *Let Ω be an open set in $\mathbb{R}^2$ and $f(t,x)$ be a continuous function in Ω that is locally Lipschitz in x. Then, for any point $(t_0, x_0) \in \Omega$, the initial value problem IVP has a solution. Furthermore, if there are two solutions $x_1(t)$ and $x_2(t)$ of the same IVP then $x_1(t) = x_2(t)$ in their common domain.*

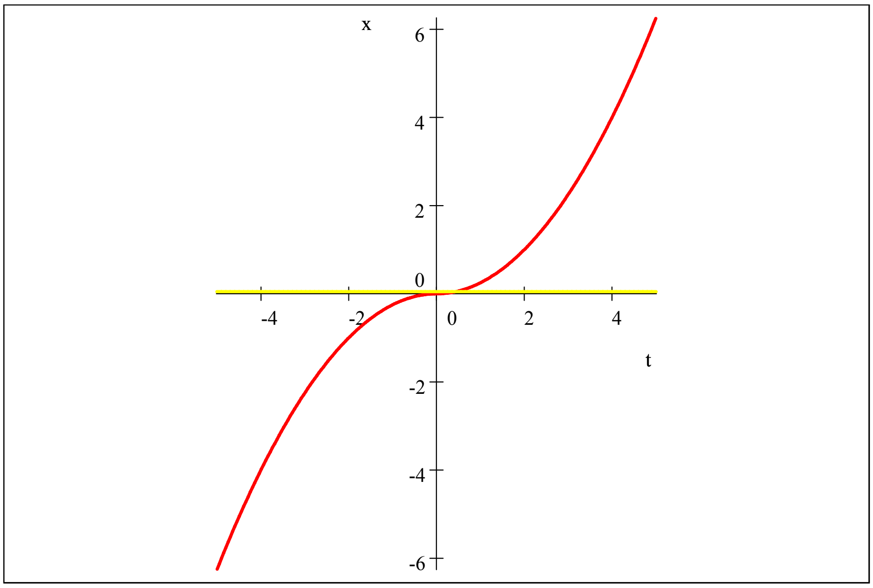
Remark. By Lemma 2.1, the hypothesis of Theorem 2.2 that f is locally Lipschitz in x could be replaced by a simpler hypotheses that f_x is continuous. However, there are simple examples of functions that are Lipschitz but not differentiable, as for example $f(x) = |x|$, and Theorem 2.2 applies for such functions.

If we completely drop the Lipschitz condition and assume only that f is continuous in (t,x) then the existence of a solution is still the case (Peano's theorem) while the uniqueness fails in general as will be seen in the next example.

Example. Consider the equation $x' = \sqrt{|x|}$ which was already solved before by separation of variables. The function $x(t) \equiv 0$ is a solution, and the following two functions

$$\begin{aligned} x(t) &= \frac{1}{4}t^2, \ \ t > 0, \\ x(t) &= -\frac{1}{4}t^2, t < 0 \end{aligned}$$

are also solutions (this can also be trivially verified by substituting them into the ODE). Gluing together these two functions and extending the resulting function to $t = 0$ by setting $x(0) = 0$, we obtain a new solution defined for all real t (see the diagram below). Hence, there are at least two solutions that satisfy the initial condition $x(0) = 0$.



The uniqueness breaks down because the function $\sqrt{|x|}$ is not Lipschitz near 0.

Proof of existence in Theorem 2.2. We start with the following observation.

Claim. *A function $x(t)$ solves IVP if and only if $x(t)$ is a continuous function on an open interval I such that $t_0 \in I$, $(t, x(t)) \in \Omega$ for all $t \in I$, and*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{2.2}$$

Indeed, if x solves IVP then (2.2) follows from $x' = f(t, x(t))$ just by integration:

$$\int_{t_0}^{t} x'(s)\, ds = \int_{t_0}^{t} f(s, x(s))\, ds$$

whence

$$x(t) - x_0 = \int_{t_0}^{t} f(s, x(s))\, ds.$$

Conversely, if x is a continuous function that satisfies (2.2) then the right hand side of (2.2) is differentiable in t whence it follows that $x(t)$ is differentiable. It is trivial that $x(t_0) = x_0$, and after differentiation (2.2) we obtain the ODE $x' = f(t, x)$.

Fix a point $(t_0, x_0) \in \Omega$ and let ε, δ be the parameter from the the local Lipschitz condition at this point, that is, there is a constant L such that

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

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$. Set

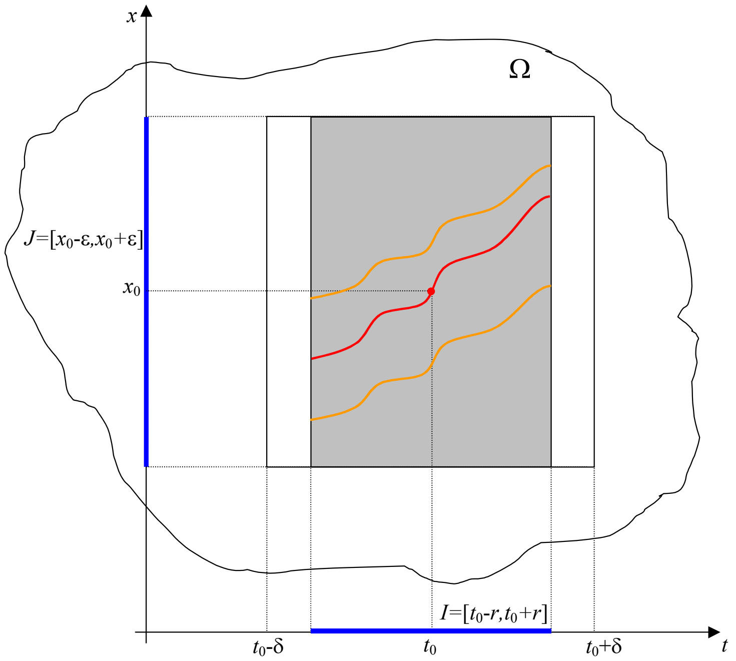
$$J = [x_0 - \varepsilon, x_0 + \varepsilon] \text{ and } I = [t_0 - r, t_0 + r],$$

were $0 < r \leq \delta$ is a new parameter, whose value will be specified later on.

Denote by X be the family of all continuous functions $x(t) : I \to J$, that is,

$$X = \{x : I \to J : x \text{ is continuous}\}$$

(see the diagram below).



We are going to consider the integral operator A defined on functions $x(t)$ by

$$Ax(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

which is obviously motivated by (2.2). To be more precise, we would like to ensure that $x \in X$ implies $Ax \in X$. Note that, for any $x \in X$, the point $(s, x(s))$ belongs to Ω so that the above integral makes sense and the function Ax is defined on I. This function is obviously continuous. We are left to verify that the image of Ax is contained in J. Indeed, the latter condition means that

$$|Ax(t) - x_0| \le \varepsilon \text{ for all } t \in I. \tag{2.3}$$

We have, for any $t \in I$,

$$|Ax(t) - x_0| = \left| \int_{t_0}^{t} f(s, x(s))\, ds \right| \le \sup_{s\in I, x\in J} |f(s,x)|\, |t - t_0| \le Mr,$$

where

$$M = \sup_{\substack{s\in[t_0-\delta, t_0+\delta] \\ x\in[x_0-\varepsilon, x_0+\varepsilon]}} |f(s,x)| < \infty.$$

Hence, if r is so small that $Mr \le \varepsilon$ then (2.3) is satisfied and, hence, $Ax \in X$.

To summarize the above argument, we have defined a function family X and a mapping $A : X \to X$. By the above Claim, a function $x \in X$ will solve the IVP if function x is a fixed point of the mapping A, that is, if $x = Ax$. The existence of a fixed point can be obtained, for example, using the Banach fixed point theorem. In order to be able to apply this theorem, we must introduce a distance function d on X so that (X, d) is a complete metric space, and A is a contraction mapping with respect to this distance.

Let d be the sup-distance, that is, for any two functions $x, y \in X$, set

$$d(x, y) = \sup_{t\in I} |x(t) - y(t)|.$$

Recall that, by a theorem from Analysis II, the space $C(I)$ of all continuous functions on I with the sup-distance is a complete metric space. The family X is a subset of $C(I)$ defined by the additional condition that the images of all functions from X are contained in J. Clearly, the set X is closed whence it follows that the metric space (X, d) is complete.

How to ensure that the mapping $A : X \to X$ is a contraction? For any two functions $x, y \in X$ and any $t \in I$, we have $x(t), y(t) \in J$ whence by the Lipschitz condition

$$\begin{aligned}
|Ax(t) - Ay(t)| &= \left| \int_{t_0}^{t} f(s, x(s))\, ds - \int_{t_0}^{t} f(s, y(s))\, ds \right| \\
&\le \left| \int_{t_0}^{t} |f(s, x(s)) - f(s, y(s))|\, ds \right| \\
&\le \left| \int_{t_0}^{t} L\, |x(s) - y(s)|\, ds \right| \\
&\le L\, |t - t_0| \sup_{I} |x - y| \\
&\le Lrd(x, y).
\end{aligned}$$

Taking sup in $t \in I$, we obtain

$$d(Ax, Ay) \leq Lrd(x, y).$$

Hence, choosing $r < 1/L$, we obtain that A is a contraction, which finishes the proof of the existence. ■

Remark. Let us summarize the proof of the existence of solutions as follows. Let ε, δ, L be the parameters from the the local Lipschitz condition at the point (t_0, x_0), that is,

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

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$. Let

$$M = \sup \{|f(t,x)| : t \in [t_0 - \delta, t_0 + \delta], \ x \in [x_0 - \varepsilon, x_0 + \varepsilon]\}.$$

Then the IVP has a solution on an interval $[t_0 - r, t_0 + r]$ provided r is a positive number that satisfies the following conditions:

$$r \leq \delta, \ r \leq \frac{\varepsilon}{M}, \ r < \frac{1}{L}. \tag{2.4}$$

For some applications, it is important that r can be determined as a function of $\varepsilon, \delta, M, L$.

For the proof of the uniqueness, we need the following two lemmas.

Lemma 2.3 (Gronwall inequality) *Let $z(t)$ be a non-negative continuous function on $[t_0, t_1]$ where $t_0 < t_1$. Assume that there are constants $C, L \geq 0$ such that*

$$z(t) \leq C + L \int_{t_0}^{t} z(s)\, ds \tag{2.5}$$

for all $t \in [t_0, t_1]$. Then

$$z(t) \leq C \exp(L(t - t_0)) \tag{2.6}$$

for all $t \in [t_0, t]$.

Proof. We can assume that C is strictly positive. Indeed, if (2.5) holds with $C = 0$ then it holds with any $C > 0$. Therefore, (2.6) holds with any $C > 0$, whence it follows that it holds with $C = 0$. Hence, assume in the sequel that $C > 0$. This implies that the right hand side of (2.5) is positive. Set

$$F(t) = C + L \int_{t_0}^{t} z(s)\, ds$$

and observe that F is differentiable and $F' = Lz$. It follows from (2.5) that

$$F' = Lz \leq LF.$$

This is a differential inequality for F that can be solved similarly to the separable ODE. Since $F > 0$, dividing by F we obtain

$$\frac{F'}{F} \leq L,$$

whence by integration

$$\ln \frac{F(t)}{F(t_0)} = \int_{t_0}^{t} \frac{F'(s)}{F(s)} ds \leq \int_{t_0}^{t} L ds = L(t - t_0).$$

It follows that

$$F(t) \leq F(t_0) \exp(L(t - t_0)) = C \exp(L(t - t_0)).$$

Using again (2.5), that is, $z \leq F$, we obtain (2.6). ■

Lemma 2.4 *If S is a subset of an interval $U \subset \mathbb{R}$ that is both open and closed in U then either S is empty or $S = U$.*

Proof. Set $S^c = U \setminus S$ so that S^c is closed in U. Assume that both S and S^c are non-empty and choose some points $a_0 \in S$, $b_0 \in S^c$. Set $c = \frac{a_0+b_0}{2}$ so that $c \in U$ and, hence, c belongs to S or S^c. Out of the intervals $[a_0, c]$, $[c, b_0]$ choose the one whose endpoints belong to different sets $S, S^c$and rename it by $[a_1, b_1]$, say $a_1 \in S$ and $b_1 \in S^c$. Considering the point $c = \frac{a_1+b_1}{2}$, we repeat the same argument and construct an interval $[a_2, b_2]$ being one of two halfs of $[a_1, b_1]$ such that $a_2 \in S$ and $b_2 \in S^c$. Contintue further, we obtain a nested sequence $\{[a_k, b_k]\}_{k=0}^{\infty}$ of intervals such that $a_k \in S$, $b_k \in S^c$ and $|b_k - a_k| \to 0$. By the principle of nested intervals, there is a common point $x \in [a_k, b_k]$ for all k. Note that $x \in U$. Since $a_k \to x$, we must have $x \in S$, and since $b_k \to x$, we must have $x \in S^c$, because both sets S and S^c are closed in U. This contradiction finishes the proof. ■

Proof of the uniqueness in Theorem 2.7. Assume that $x_1(t)$ and $x_2(t)$ are two solutions of the same IVP both defined on an open interval $U \subset \mathbb{R}$ and prove that they coincide on U.

We first prove that the two solution coincide in some interval around t_0. Let ε and δ be the parameters from the Lipschitz condition at the point (t_0, x_0) as above. Choose $0 < r < \delta$ so small that the both functions $x_1(t)$ and $x_2(t)$ restricted to $I = [t_0 - r, t_0 + r]$ take values in $J = [x_0 - \varepsilon, x_0 + \varepsilon]$ (which is possible because both $x_1(t)$ and $x_2(t)$ are continuous functions). As in the proof of the existence, the both solutions satisfies the integral identity

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

for all $t \in I$. Hence, for the difference $z(t) := |x_1(t) - x_2(t)|$, we have

$$z(t) = |x_1(t) - x_2(t)| \le \int_{t_0}^{t} |f(s, x_1(s)) - f(s, x_2(s))|\, ds,$$

assuming for certainty that $t_0 \le t \le t_0 + r$. Since the both points $(s, x_1(s))$ and $(s, x_2(s))$ in the given range of s are contained in $I \times J$, we obtain by the Lipschitz condition

$$|f(s, x_1(s)) - f(s, x_2(s))| \le L\, |x_1(s) - x_2(s)|$$

whence

$$z(t) \le L \int_{t_0}^{t} z(s)\, ds.$$

Appling the Gronwall inequality with $C = 0$ we obtain $z(t) \le 0$. Since $z \ge 0$, we conclude that $z(t) \equiv 0$ for all $t \in [t_0, t_0 + r]$. In the same way, one gets that $z(t) \equiv 0$ for $t \in [t_0 - r, t_0]$, which proves that the solutions $x_1(t)$ and $x_2(t)$ coincide on the interval I.

Now we prove that they coincide on the full interval U. Consider the set

$$S = \{t \in U : x_1(t) = x_2(t)\}$$

and let us show that the set S is both closed and open in I. The closedness is obvious: if $x_1(t_k) = x_2(t_k)$ for a sequence $\{t_k\}$ and $t_k \to t \in U$ as $k \to \infty$ then passing to the limit and using the continuity of the solutions, we obtain $x_1(t) = x_2(t)$, that is, $t \in S$.

Let us prove that the set S is open. Fix some $t_1 \in S$. Since $x_1(t_1) = x_2(t_1)$, the both functions $x_1(t)$ and $x_2(t)$ solve the same IVP with the initial condition at t_1. By the above argument, $x_1(t) = x_2(t)$ in some interval $I = [t_1 - r, t_1 + r]$ with $r > 0$. Hence, $I \subset S$, which implies that S is open.

Since the set S is non-empty (it contains t_0) and is both open and closed in U, we conclude by Lemma 2.4 that $S = U$, which finishes the proof of uniqueness. ■

Example. The method of the proof of the existence of the solution suggest the following iteration procedure for computation of the solution. Recall that finding a solution amounts to solving the equation $x = Ax$ where A is the integral operator

$$Ax(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

defined on functions $x \in X$, where X is the class of all continuous functions from I to J. By the proof of the Banach fixed point theorem, we can start with any function in X and construct a sequence $\{x_k\}$ of functions from X such that $x_{k+1} = Ax_k$. Then $x_k(t)$ converges to the solution $x(t)$ uniformly in $t \in I$. Choose the initial function $x_0(t)$ to be the constant x_0. In general, one cannot compute x_k explicitly, but for a particular choice of f this is possible. Namely, take $f(t, x) = x$, $t_0 = 0$, $x_0 = 1$, which corresponds to the the IVP

$$\begin{cases} x' = x, \\ x(0) = 1. \end{cases}$$

Then we have

$$Ax(t) = 1 + \int_0^t x(s)\, ds$$

whence

$$x_1(t) = 1 + \int_0^t x_0 ds = 1 + t,$$

$$x_2(t) = 1 + \int_0^t x_1 ds = 1 + t + \frac{t^2}{2}$$

$$x_3(t) = 1 + \int_0^t x_2 dt = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!}$$

and by induction

$$x_k(t) = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + ... + \frac{t^k}{k!}.$$

Clearly, $x_k \to e^t$ as $k \to \infty$, and the function $x(t) = e^t$ indeed solves the above IVP.

2.2 Dependence on the initial value

Consider the IVP

$$\begin{cases} x' = f(t, x) \\ x(t_0) = s \end{cases}$$

where the initial value is denoted by s instead of x_0 to emphasize that we allow now s to vary. Hence, the solution is can be considered as a function of two variables: $x = x(t, s)$. Our aim is to investigate the dependence on s.

As before, assume that f is continuous in an open set $\Omega \subset \mathbb{R}^2$ and is locally Lipschitz in this set in x. Fix a point $(t_0, x_0) \in \Omega$ and let ε,δ, L be the parameters from the local Lipschitz condition at this point, that is,

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

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$. Let M be the supremum of $|f(t,x)|$ in the rectangle $[t_0 - \delta, t_0 + \delta] \times [x_0 - \varepsilon, x_0 + \varepsilon]$.

As we know by the proof of Theorem 2.2, the solution with the initial condition $x(t_0) = x_0$ is defined in the interval $[t_0 - r, t_0 + r]$ where r is any positive number that satisfies (2.4), and $x(t)$ takes values in $[x_0 - \varepsilon, x_0 + \varepsilon]$. Now consider the IVP with the condition $x(t_0) = s$ where s is close enough to x_0, say

$$s \in [x_0 - \varepsilon/2, x_0 + \varepsilon/2]\,. \tag{2.7}$$

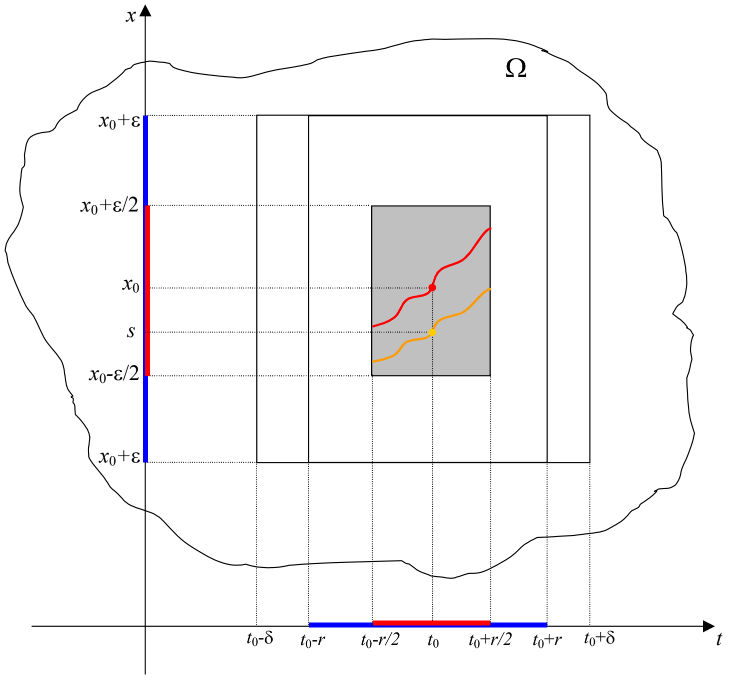
Then the interval $[s - \varepsilon/2, s + \varepsilon/2]$ is contained in $[x_0 - \varepsilon, x_0 + \varepsilon]$ so that the above Lipschitz condition holds if we replace the interval $[x_0 - \varepsilon, x_0 + \varepsilon]$ by $[s - \varepsilon/2, s + \varepsilon/2]$. Also, the supremum of $|f(t,x)|$ in $[t_0 - \delta, t_0 + \delta] \times [s - \varepsilon/2, s + \varepsilon/2]$ is bounded by M. Hence, the solution $x(t,s)$ is defined for all $t \in [t_0 - r(s), t_0 + r(s)]$ provided $r(s)$ satisfies the conditions.

$$r(s) \leq \delta, \ \ r(s) \leq \frac{\varepsilon}{2M}, \ \ r(s) < \frac{1}{L}. \tag{2.8}$$

Note that in comparison with (2.4) we use here $\varepsilon/2$ instead of ε to ensure that the solution takes values in $[s - \varepsilon/2, s + \varepsilon/2]$. Hence, if r satisfies (2.4) then we can take $r(s) = r/2$, which then satisfies (2.8). Hence, for any s as in (2.7), the solution $x(t,s)$ is defined in the interval

$$t \in [t_0 - r/2, t_0 + r/2] \tag{2.9}$$

and takes values in the interval $[x_0 - \varepsilon, x_0 + \varepsilon]$. In particular, we can compare solutions with different s since they have the common domain (2.9).



Theorem 2.5 (Continuous dependence on the initial value) *Let Ω be an open set in $\mathbb{R}^2$ and $f(t,x)$ be a continuous function in Ω that is locally Lipschitz in x. Let (t_0,x_0) be a point in Ω and let ε, r be as above. Then, for all $s', s'' \in [x_0-\varepsilon/2, x_0+\varepsilon/2]$ and $t \in [t_0-r/2, t_0+r/2]$,*

$$|x(t,s')-x(t,s'')| \le 2|s'-s''|. \tag{2.10}$$

Consequently, the function $x(t,s)$ is continuous in (t,s).

Proof. Consider again the integral equations

$$x(t,s') = s' + \int_{t_0}^{t} f(\tau, x(\tau, s'))\, d\tau$$

and

$$x(t,s'') = s'' + \int_{t_0}^{t} f(\tau, x(\tau, s''))\, d\tau.$$

Setting $z(t) = |x(t,s') - x(t,s'')|$ and assuming $t \in [t_0, t_0 + r/2]$, we obtain, using the Lipschitz condition

$$\begin{aligned} z(t) &\leq |s' - s''| + \int_{t_0}^{t} |f(\tau, x(\tau, s')) - f(\tau, x(\tau, s''))|\, d\tau \\ &\leq |s' - s''| + L \int_{t_0}^{t} z(\tau)\, d\tau. \end{aligned}$$

By the Gronwall inequality, we conclude that

$$z(t) \leq |s' - s''| \exp(L(t - t_0)).$$

Since $t - t_0 \leq r/2$ and $L \leq \frac{1}{2r}$we see that $L(t - t_0) \leq \frac{1}{4}$ and $\exp(L(t - t_0)) \leq e^{1/4} < 2$, which proves (2.10) for $t \geq t_0$. Similarly one obtains the same for $t \leq t_0$.

Let us prove that $x(t,s)$ is continuous in (t,s). Fix a point $(t_0, x_0) \in \Omega$ and prove that $x(t,s)$ is continuous at this point, that is,

$$x(t_n, x_n) \to x(t_0, x_0)$$

if $(t_n, x_n) \to (t_0, x_0)$. Choosing ε and r as above and taking n large enough, we can assume that $x_n \in [x_0 - \varepsilon/2, x_0 + \varepsilon/2]$ and $t_n \in [t_0 - r/2, t_0 + r/2]$. Then by (2.10)

$$\begin{aligned} |x(t_n, x_n) - x(t_0, x_0)| &\leq |x(t_n, x_n) - x(t_n, x_0)| + |x(t_n, x_0) - x(t_0, x_0)| \\ &\leq 2|x_n - x_0| + |x(t_n, x_0) - x(t, x_0)|, \end{aligned}$$

and this goes to 0 as $n \to \infty$. ■

2.3 Higher order ODE and reduction to the first order system

A general ODE of the order n resolved with respect to the highest derivative can be written in the form

$$y^{(n)} = F\left(t, y, ..., y^{(n-1)}\right), \tag{2.11}$$

where t is an independent variable and $y(t)$ is an unknown function. It is sometimes more convenient to replace this equation by a system of ODEs of the 1^{st} order.

Let $x(t)$ be a vector function of a real variable t, which takes values in $\mathbb{R}^n$. Denote by x_k the components of x. Then the derivative $x'(t)$ is defined component-wise by

$$x' = (x_1', x_2', ..., x_n').$$

Consider now a *vector ODE of the first order*

$$x' = f(t, x) \tag{2.12}$$

where f is a given function of $n+1$ variables, which takes values in $\mathbb{R}^n$, that is, $f : \Omega \to \mathbb{R}^n$ where Ω is an open subset of $\mathbb{R}^{n+1}$ (so that the couple (t, x) is considered as a point in Ω). Denoting by f_k the components of f, we can rewrite the vector equation (2.12) as a system of n scalar equations

$$\begin{cases} x_1' = f_1(t, x_1, ..., x_n) \\ ... \\ x_k' = f_k(t, x_1, ..., x_n) \\ ... \\ x_n' = f_n(t, x_1, ..., x_n) \end{cases} \tag{2.13}$$

Let us show how the equation (2.11) can be reduced to the system (2.13). Indeed, with any function $y(t)$ let us associate the vector-function

$$x = \left(y, y', ..., y^{(n-1)}\right),$$

which takes values in $\mathbb{R}^n$. That is, we have

$$x_1 = y, \ x_2 = y', \ ..., \ x_n = y^{(n-1)}.$$

Obviously,

$$x' = \left(y', y'', ..., y^{(n)}\right),$$

and using (2.11) we obtain a system of equations

$$\begin{cases} x_1' = x_2 \\ x_2' = x_3 \\ ... \\ x_{n-1}' = x_n \\ x_n' = F(t, x_1, ...x_n) \end{cases} \tag{2.14}$$

Obviously, we can rewrite this system as a vector equation (2.12) where

$$f(t, x) = (x_2, x_3, ..., x_n, F(t, x_1, ..., x_n)). \tag{2.15}$$

Conversely, the system (2.14) implies

$$x_1^{(n)} = x_n' = F\left(t, x_1, x_1', .., x_1^{(n-1)}\right)$$

so that we obtain equation (2.11) with respect to $y = x_1$. Hence, the equation (2.11) is equivalent to the vector equation (2.12) with function f defined by (2.15).

Example. For example, consider the second order equation

$$y'' = F(t, y, y').$$

Setting $x = (y, y')$ we obtain

$$x' = (y', y'')$$

whence

$$\begin{cases} x_1' = x_2 \\ x_2' = F(t, x_1, x_2) \end{cases}$$

Hence, we obtain the vector equation (2.12) with

$$f(t,x) = (x_2, F(t, x_1, x_2)).$$

What initial value problem is associated with the vector equation (2.12) and the scalar higher order equation (2.11)? Motivated by the study of the 1st order ODE, one can presume that it makes sense to consider the following IVP for the vector 1st order ODE

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases}$$

where $x_0 \in \mathbb{R}^n$ is a given initial value of $x(t)$. For the equation (2.11), this means that the initial conditions should prescribe the value of the vector $x = \left(y, y', ..., y^{(n-1)}\right)$ at some t_0, which amounts to n scalar conditions

$$\begin{cases} y(t_0) = y_0 \\ y'(t_0) = y_1 \\ ... \\ y^{(n-1)}(t_0) = y_{n-1} \end{cases}$$

where $y_0, ..., y_{n-1}$ are given values. Hence, the initial value problem IVP for the scalar equation of the order n can be stated as follows:

$$\begin{cases} y' = F\left(t, y, y', ..., y^{(n-1)}\right) \\ y(t_0) = y_0 \\ y'(t_0) = y_1 \\ ... \\ y^{(n-1)}(t_0) = y_{n-1}. \end{cases}$$

2.4 Existence and uniqueness for a system

Let Ω be an open subset of $\mathbb{R}^{n+1}$ and f be a mapping from Ω to $\mathbb{R}^n$. Denote a point in $\mathbb{R}^{n+1}$ by (t,x) where $t \in \mathbb{R}$ and $x \in \mathbb{R}^n$. Then we write $f = f(t,x)$.

Definition. Function f is called locally Lipschitz in x if for any point $(t_0 x_0) \in \Omega$ there exists positive constants ε, δ, L such that

$$\|f(t,x) - f(t,y)\| \leq L \|x - y\| \tag{2.16}$$

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in \overline{B}(x_0, \varepsilon)$.

Here $\|\cdot\|$ denotes some norms in $\mathbb{R}^n$ and $\mathbb{R}^{n+1}$ (arbitrary, but fixed) and $\overline{B}(x_0, \varepsilon)$ is the closed ball in $\mathbb{R}^n$, that is,

$$\overline{B}(x_0, \varepsilon) = \{y \in \mathbb{R}^n : \|x - y\| \leq \varepsilon\}.$$

Note that the value of the Lipschitz constant L depends on the choice of the norms, but the property of f to be locally Lipschitz is independent of the choice of the norms.

Lemma 2.6 *If all partial derivatives $\frac{\partial f_k}{\partial x_j}$ exists in Ω and are continuous then f is locally Lipschitz in Ω.*

Proof. Given a point (t_0, x_0) choose ε and δ so that the cylinder

$$K = [t_0 - \delta, t_0 + \delta] \times \overline{B}(x_0, \varepsilon)$$

is contained in Ω, which is possible by the openness of Ω. Since K is a closed bounded set, all functions $\left|\frac{\partial f_k}{\partial x_j}\right|$ are bounded on K. Set

$$C = \max_{k,j} \sup_K \left|\frac{\partial f_k}{\partial x_j}\right|.$$

Fix an index $k = 1, ..., n$, $t \in [t_0 - \delta, t_0 + \delta]$, and consider $f_k(t, x)$ as a function of x only (that is, as a mapping from a subset of $\mathbb{R}^n$ to $\mathbb{R}$). For any two points $x, y \in \overline{B}(x_0, \varepsilon)$, we have by the mean value theorem in $\mathbb{R}^n$

$$f_k(t, x) - f_k(t, y) = (f_k)_x(t, \xi)(x - y), \tag{2.17}$$

where ξ is a point in the interval $[x, y]$ and, hence, in $\overline{B}(x_0, \varepsilon)$, and $(f_k)_x$ is the full derivative of f_k in x. In fact, since the partial derivatives $\frac{\partial f_k}{\partial x_j}$ are continuous, the full derivative coincides with the Jacobian matrix, that is, $(f_k)_x$ is the $1 \times n$ matrix

$$(f_k)_x = \left(\frac{\partial f_k}{\partial x_1}, ..., \frac{\partial f_k}{\partial x_n}\right).$$

The right hand side of (2.17) is the product of this row and the column-vector $x - y$, that is,

$$f_k(t, x) - f_k(t, y) = \sum_{j=1}^{n} \frac{\partial f_k}{\partial x_j}(t, \xi)(x_j - y_j).$$

Since $(t, \xi) \in K$, we obtain by the definition of C

$$|f_k(t, x) - f_k(t, y)| \le C \sum_{j=1}^{n} |x_j - y_j| = C\|x - y\|_1.$$

Taking max in k, we obtain

$$\|f(t, x) - f(t, y)\|_\infty \le C\|x - y\|_1.$$

Switching to the fixed norm $\|\cdot\|$ in $\mathbb{R}^n$ and using the fact that any two norms have bounded ratio, we obtain (2.16). ■

Definition. Given a function $f : \Omega \to \mathbb{R}^n$, where Ω is an open set in $\mathbb{R}^{n+1}$, consider the IVP

$$\begin{cases} x' = f(t, x) \\ x(t_0) = x_0, \end{cases} \tag{2.18}$$

where (t_0, x_0) is a given point in Ω. A solution to IVP is a function $x(t) : I \to \mathbb{R}^n$ (where I is an open interval containing t_0) such that $(t, x(t)) \in \Omega$ for all $t \in I$ and $x(t)$ satisfies the ODE $x' = f(t, x)$ in I and the initial condition $x(t_0) = x_0$.

The graph of function $x(t)$, that is, the set of points $(t, x(t))$, is hence a curve in Ω that goes through the point (t_0, x_0).

Theorem 2.7 (Picard - Lindelöf Theorem) *Consider the equation*

$$x' = f(t, x)$$

where $f : \Omega \to \mathbb{R}^n$ is a mapping from an open set $\Omega \subset \mathbb{R}^{n+1}$ to $\mathbb{R}^n$. Assume that f is continuous on Ω and locally Lipschitz in x. Then, for any point $(t_0, x_0) \in \Omega$, the initial value problem IVP (2.18) *has a solution.*

Furthermore, if $x(t)$ and $y(t)$ are two solutions to the same IVP then $x(t) = y(t)$ in their common domain.

Proof. The proof is very similar to the case $n = 1$ considered in Theorem 2.2. We start with the following claim.

Claim. *A function $x(t)$ solves IVP if and only if $x(t)$ is a continuous function on an open interval I such that $t_0 \in I$, $(t, x(t)) \in \Omega$ for all $t \in I$, and*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{2.19}$$

Here the integral of the vector valued function is understood component-wise. If x solves IVP then (2.19) follows from $x_k' = f_k(t, x(t))$ just by integration:

$$\int_{t_0}^{t} x_k'(s)\, ds = \int_{t_0}^{t} f_k(s, x(s))\, ds$$

whence

$$x_k(t) - (x_0)_k = \int_{t_0}^{t} f_k(s, x(s))\, ds$$

and (2.19) follows. Conversely, if x is a continuous function that satisfies (2.19) then

$$x_k = (x_0)_k + \int_{t_0}^{t} f_k(s, x(s))\, ds.$$

The right hand side here is differentiable in t whence it follows that $x_k(t)$ is differentiable. It is trivial that $x_k(t_0) = (x_0)_k$, and after differentiation we obtain $x_k' = f_k(t, x)$ and, hence, $x' = f(t, x)$.

Fix a point $(t_0, x_0) \in \Omega$ and let ε, δ be the parameter from the the local Lipschitz condition at this point, that is, there is a constant L such that

$$\|f(t, x) - f(t, y)\| \le L \|x - y\|$$

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in \overline{B}(x_0, \varepsilon)$. Set $I = [t_0 - r, t_0 + r]$,where $0 < r \le \delta$ is a new parameter, whose value will be specified later on, and $J = \overline{B}(x_0, \varepsilon)$.

Denote by X be the family of all continuous functions $x(t) : I \to J$, that is,

$$X = \{x : I \to J : x \text{ is continuous}\}.$$

We are going to consider the integral operator A defined on functions $x(t)$ by

$$Ax(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

and we would like to ensure that $x \in X$ implies $Ax \in X$. Note that, for any $x \in X$, the point $(s, x(s))$ belongs to Ω so that the above integral makes sense and the function Ax is defined on I. This function is obviously continuous. We are left to verify that the image of Ax is contained in J. Indeed, the latter condition means that

$$\|Ax(t) - x_0\| \leq \varepsilon \text{ for all } t \in I. \tag{2.20}$$

We have, for any $t \in I$,

$$\begin{aligned} \|Ax(t) - x_0\| &= \left\| \int_{t_0}^{t} f(s, x(s))\, ds \right\| \\ &\leq \int_{t_0}^{t} \|f(s, x(s))\|\, ds \\ &\leq \sup_{s \in I, x \in J} \|f(s, x)\| \, |t - t_0| \leq Mr, \end{aligned}$$

where

$$M = \sup_{\substack{s \in [t_0 - \delta, t_0 + \delta] \\ x \in \overline{B}(x_0, \varepsilon).}} |f(s, x)| < \infty.$$

Hence, if r is so small that $Mr \leq \varepsilon$ then (2.3) is satisfied and, hence, $Ax \in X$.

Define a distance function on the function family X as follows: if $x, y \in X$ then

$$d(x,y) = \sup_{t \in I} \|x(t) - y(t)\|.$$

We claim that (X, d) is a complete metric space (see Exercises).

We are left to ensure that the mapping $A : X \to X$ is a contraction. For any two functions $x, y \in X$ and any $t \in I$, $t \geq t_0$, we have $x(t), y(t) \in J$ whence by the Lipschitz condition

$$\begin{aligned} \|Ax(t) - Ay(t)\| &= \left\| \int_{t_0}^{t} f(s, x(s))\, ds - \int_{t_0}^{t} f(s, y(s))\, ds \right\| \\ &\leq \int_{t_0}^{t} \|f(s, x(s)) - f(s, y(s))\|\, ds \\ &\leq \int_{t_0}^{t} L\, \|x(s) - y(s)\|\, ds \\ &\leq L(t - t_0) \sup_{s \in I} \|x(s) - y(s)\| \\ &\leq Lr d(x,y). \end{aligned}$$

The same inequality holds for $t \leq t_0$. Taking sup in $t \in I$, we obtain

$$d(Ax, Ay) \leq Lr d(x,y).$$

Hence, choosing $r < 1/L$, we obtain that A is a contraction. By the Banach fixed point theorem, we conclude that the equation $Ax = x$ has a solution $x \in X$, which hence solves the IVP.

Assume that $x(t)$ and $y(t)$ are two solutions of the same IVP both defined on an open interval $U \subset \mathbb{R}$ and prove that they coincide on U. We first prove that the two solution coincide in some interval around t_0. Let ε and δ be the parameters from the Lipschitz condition at the point (t_0, x_0) as above. Choose $0 < r < \delta$ so small that the both functions $x(t)$ and $y(t)$ restricted to $I = [t_0 - r, t_0 + r]$ take values in $J = \overline{B}(x_0, \varepsilon)$ (which is possible because both $x(t)$ and $y(t)$ are continuous functions). As in the proof of the existence, the both solutions satisfies the integral identity

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

for all $t \in I$. Hence, for the difference $z(t) := \|x(t) - y(t)\|$, we have

$$z(t) = \|x(t) - y(t)\| \leq \int_{t_0}^{t} \|f(s, x(s)) - f(s, y(s))\|\, ds,$$

assuming for certainty that $t_0 \leq t \leq t_0 + r$. Since the both points $(s, x(s))$ and $(s, y(s))$ in the given range of s are contained in $I \times J$, we obtain by the Lipschitz condition

$$|f(s, x(s)) - f(s, y(s))| \leq L\, \|x(s) - y(s)\|$$

whence

$$z(t) \leq L \int_{t_0}^{t} z(s)\, ds.$$

Appling the Gronwall inequality with $C = 0$ we obtain $z(t) \leq 0$. Since $z \geq 0$, we conclude that $z(t) \equiv 0$ for all $t \in [t_0, t_0 + r]$. In the same way, one gets that $z(t) \equiv 0$ for $t \in [t_0 - r, t_0]$, which proves that the solutions $x(t)$ and $y(t)$ coincide on the interval I.

Now we prove that they coincide on the full interval U. Consider the set

$$S = \{t \in U : x(t) = y(t)\}$$

and let us show that the set S is both closed and open in I. The closedness is obvious: if $x(t_k) = y(t_k)$ for a sequence $\{t_k\}$ and $t_k \to t \in U$ as $k \to \infty$ then passing to the limit and using the continuity of the solutions, we obtain $x(t) = y(t)$, that is, $t \in S$.

Let us prove that the set S is open. Fix some $t_1 \in S$. Since $x(t_1) = y(t_1) =: x_1$, the both functions $x(t)$ and $y(t)$ solve the same IVP with the initial data (t_1, x_1). By the above argument, $x(t) = y(t)$ in some interval $I = [t_1 - r, t_1 + r]$ with $r > 0$. Hence, $I \subset S$, which implies that S is open.

Since the set S is non-empty (it contains t_0) and is both open and closed in U, we conclude by Lemma 2.4 that $S = U$, which finishes the proof of uniqueness. ■

Remark. Let us summarize the proof of the existence part of Theorem 2.7 as follows. For any point $(t_0, x_0) \in \Omega$, we first choose positive constants ε, δ, L from the Lipschitz condition, that is, the cylinder

$$G = [t_0 - \delta, t_0 + \delta] \times \overline{B}(x_0, \varepsilon)$$

is contained in Ω and, for any two points (t, x) and (t, y) from G with the same t,

$$\|f(t,x) - f(t,y)\| \leq L \|x - y\|.$$

Let

$$M = \sup_G \|f(t,x)\|$$

and choose any positive r to satisfy

$$r \leq \delta,\ r \leq \frac{\varepsilon}{M},\ r < \frac{1}{L}. \tag{2.21}$$

Then there exists a solution $x(t)$ to the IVP, which is defined on the interval $[t_0 - r, t_0 + r]$ and takes values in $\overline{B}(x_0, \varepsilon)$.

The fact that the domain of the solution admits the explicit estimates (2.21) can be used as follows.

Corollary. *Under the conditions of Theorem 2.7 for any point $(t_0, x_0) \in \Omega$ there are positive constants ε and r such that, for any $t_1 \in [t_0 - r, t_0 + r]$ and $x_1 \in \overline{B}(x_0, \varepsilon/2)$ the IVP*

$$\begin{cases} x' = f(t,x), \\ x(t_1) = x_1 \end{cases} \tag{2.22}$$

has a solution $x(t)$ defined for $t \in [t_1 - r, t_1 + r]$ and taking values in $\overline{B}(x_1, \varepsilon/2)$.

In particular, if $t_1 \in [t_0 - r/2, t_0 + r/2]$ then $x(t)$ is defined for all $t \in [t_0 - r/2, t_0 + r/2]$ and takes values in $\overline{B}(x_0, \varepsilon)$.

Proof. Let $\varepsilon, \delta, L, M$ be as above. Assuming that $t_1 \in [t_0 - \delta/2, t_0 + \delta/2]$ and $x_1 \in \overline{B}(x_0, \varepsilon/2)$, we obtain that the cylinder

$$G_1 = [t_1 - \delta/2, t_1 + \delta/2] \times \overline{B}(x_1, \varepsilon/2)$$

is contained in G. Hence, the values of L and M for the cylinder G_1 can be taken the same as those for G. Hence, the IVP (2.22) has solution $x(t)$ in the interval $[t_1 - r, t_1 + r]$ taking values in $\overline{B}(x_1, \varepsilon/2)$ provided

$$r \leq \delta/2, \; r \leq \frac{\varepsilon}{2M}, \; r < \frac{1}{L}.$$

In particular, r can be taken to depend only on $\varepsilon, \delta, L, M$, that is, r is a function of (t_0, x_0). We are left to observe that, for this choice of r, the condition $t_1 \in [t_0 - r/2, t_0 + r/2]$ implies $t_1 \in [t_0 - \delta/2, t_0 + \delta/2]$.

The second claim follows from the observations that $\overline{B}(x_0, \varepsilon) \supset \overline{B}(x_1, \varepsilon/2)$ and $[t_0 - r/2, t_0 + r/2] \subset [t_1 - r, t_1 + r]$ provided $t_1 \in [t_0 - r/2, t_0 + r/2]$. ■

2.5 Maximal solutions

Consider again the ODE

$$x' = f(t, x)$$

where $f : \Omega \to \mathbb{R}^n$ is a mapping from an open set $\Omega \subset \mathbb{R}^{n+1}$ to $\mathbb{R}^n$, which is continuous on Ω and locally Lipschitz in x.

Although the uniqueness part of Theorem 2.7 says that any two solutions are the same in their common interval, still there are many different solutions to the same IVP because strictly speaking, the functions that are defined on different domains are different, despite they coincide in the intersection of the domains. The purpose of what follows is to define the maximal possible domain where the solution to the IVP exists.

We say that a solution $y(t)$ of the ODE is an extension of a solution $x(t)$ if the domain of $y(t)$ contains the domain of $x(t)$ and the solutions coincide in the common domain.

Definition. A solution $x(t)$ of the ODE is called *maximal* if it is defined on an open interval and cannot be extended to any larger open interval.

Theorem 2.8 *Assume that the conditions of Theorem* 2.7 *are satisfied. Then the following is true.*

(a) *Any IVP has is a unique maximal solution.*

(b) *If $x(t)$ and $y(t)$ are two maximal solutions to the same ODE and $x(t) = y(t)$ for some value of t, then x and y are identically equal, including the identity of their domains.*

(c) *If $x(t)$ is a maximal solution with the domain (a, b), then $x(t)$ leaves any compact set $K \subset \Omega$ as $t \to a$ and as $t \to b$.*

Here the phrase "$x(t)$ leaves any compact set K as $t \to b$" means the follows: there is $T \in (a, b)$ such that for any $t \in (T, b)$, the point $(t, x(t))$ does not belong to K. Similarly, the phrase "$x(t)$ leaves any compact set K as $t \to a$" means that there is $T \in (a, b)$ such that for any $t \in (a, T)$, the point $(t, x(t))$ does not belong to K.

Example. 1. Consider the ODE $x' = x^2$ in the domain $\Omega = \mathbb{R}^2$. This is separable equation and can be solved as follows. Obviously, $x \equiv 0$ is a constant solution. In the domains where $x \neq 0$ we have

$$\int \frac{x' dt}{x^2} = \int dt$$

whence

$$-\frac{1}{x} = \int \frac{dx}{x^2} = \int dt = t + C$$

and $x(t) = -\frac{1}{t-C}$ (where we have replaced C by $-C$). Hence, the family of all solutions consists of a straight line $x(t) = 0$ and hyperbolas $x(t) = -\frac{1}{x-C}$ with the maximal domains $(C, +\infty)$ and $(-\infty, C)$. Each of these solutions leaves any compact set K, but in different ways: the solutions $x(t) = 0$ leaves K as $t \to \pm\infty$ because K is bounded, while $x(t) = -\frac{1}{x-C}$ leaves K as $t \to C$ because $x(t) \to \pm\infty$.

2. Consider the ODE $x' = \frac{1}{x}$ in the domain $\Omega = \{t \in \mathbb{R} \text{ and } x > 0\}$. By separation of variables, we obtain

$$\frac{x^2}{2} = \int x dx = \int x x' dt = \int dt = t + C$$

whence

$$x(t) = \sqrt{2(t-C)}\ ,\ t > C$$

(where we have changed the constant C). Obviously, the solution is maximal in the domain $(C, +\infty)$. It leaves any compact $K \subset \Omega$ as $t \to C$ because $(t, x(t))$ tends to the point $(C, 0)$ at the boundary of Ω.

The proof of Theorem 2.8 will be preceded by a lemma.

Lemma 2.9 *Let $\{x_\alpha(t)\}_{\alpha \in A}$ be a family of solutions to the same IVP where A is any index set, and let the domain of x_α be an open interval I_α. Set $I = \bigcup_{\alpha \in A} I_\alpha$ and define a function $x(t)$ on I as follows:*

$$x(t) = x_\alpha(t) \text{ if } t \in I_\alpha. \tag{2.23}$$

Then I is an open interval and $x(t)$ is a solution to the same IVP on I.

The function $x(t)$ defined by (2.23) is referred to as the *union* of the family $\{x_\alpha(t)\}$.

Proof. First of all, let us verify that the identity (2.23) defines $x(t)$ correctly, that is, the right hand side does not depend on the choice of α. Indeed, if also $t \in I_\beta$ then t belongs to the intersection $I_\alpha \cap I_\beta$ and by the uniqueness theorem, $x_\alpha(t) = x_\beta(t)$. Hence, the value of $x(t)$ is independent of the choice of the index α. Note that the graph of $x(t)$ is the union of the graphs of all functions $x_\alpha(t)$.

Set $a = \inf I$, $b = \sup I$ and show that $I = (a,b)$. Let us first verify that $(a,b) \subset I$, that is, any $t \in (a,b)$ belongs also to I. Assume for certainty that $t \geq t_0$. Since $b = \sup I$, there is $t_1 \in I$ such that $t < t_1 < b$. There exists an index α such that $t_1 \in I_{\ \alpha}$. Since also $t_0 \in I_\alpha$, the entire interval $[t_0, t_1]$ is contained in I_α. Since $t \in [t_0, t_1]$, we conclude that $t \in I_\alpha$ and, hence, $t \in I$.

It follows that I is an interval with the endpoints a and b. Since I is the union of open intervals, I is an open subset of $\mathbb{R}$, whence it follows that I is an open interval, that is, $I = (a,b)$.

Finally, let us verify why $x(t)$ solves the given IVP. We have $x(t_0) = x_0$ because $t_0 \in I_\alpha$ for any α and

$$x(t_0) = x_\alpha(t_0) = x_0$$

so that $x(t)$ satisfies the initial condition. Why $x(t)$ satisfies the ODE at any $t \in I$? Any given $t \in I$ belongs to some I_α. Since x_α solves the ODE in I_α and $x \equiv x_\alpha$ on I_α, we conclude that x satisfies the ODE at t, which finishes the proof. ■

Proof of Theorem 2.8. (a) Consider the IVP

$$\begin{cases} x' = f(t,x), \\ x(t_0) = x_0 \end{cases} \tag{2.24}$$

and let S be the set of all possible solutions to this IVP defined on open intervals. Let $x(t)$ be the union of all solutions from S. By Lemma 2.9, the function $x(t)$ is also a solution to the IVP and, hence, $x(t) \in S$. Moreover, $x(t)$ is a maximal solution because the domain of $x(t)$ contains the domains of all other solutions from S and, hence, $x(t)$ cannot be extended to a larger open interval. This proves the existence of a maximal solution.

Let $y(t)$ be another maximal solution to the IVP and let $z(t)$ be the union of the solutions $x(t)$ and $y(t)$. By Lemma 2.9, $z(t)$ solves the IVP and extends both $x(t)$ and $y(t)$, which implies by the maximality of x and y that z is identical to both x and y. Hence, x and y are identical (including the identity of the domains), which proves the uniqueness of a maximal solution.

(b) Let $x(t)$ and $y(t)$ be two maximal solutions that coincide at some t, say $t = t_1$. Set $x_1 = x(t_1) = y(t_1)$. Then both x and y are solutions to the same IVP with the initial point (t_1, x_1) and, hence, they coincide by part (a).

(c) Let $x(t)$ be a maximal solution defined on (a,b) and assume that $x(t)$ does not leave a compact $K \subset \Omega$ as $t \to a$. Then there is a sequence $t_k \to a$ such that $(t_k, x_k) \in K$ where $x_k = x(t_k)$. By a property of compact sets, any sequence in K has a convergent subsequence whose limit is in K. Hence, passing to a subsequence, we can assume that the sequence $\{(t_k, x_k)\}_{k=1}^\infty$ converges to a point $(t_0, x_0) \in K$ as $k \to \infty$. Clearly, we have $t_0 = a$, which in particular implies that a is finite.

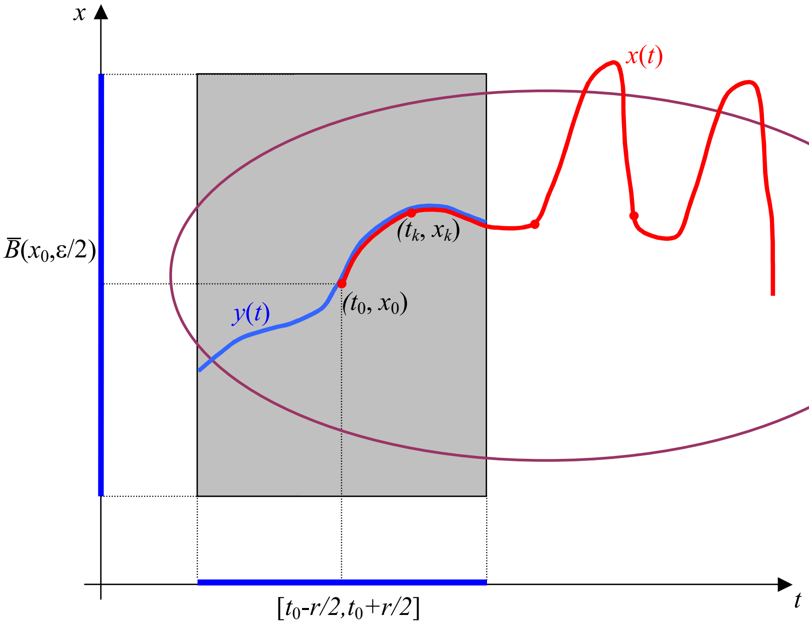
By Corollary to Theorem 2.7, for the point (t_0, x_0), there exist $r, \varepsilon > 0$ such that the IVP with the initial point inside the cylinder

$$G = [t_0 - r/2, t_0 + r/2] \times \overline{B}(x_0, \varepsilon/2)$$

has a solution defined for all $t \in [t_0 - r/2, t_0 + r/2]$. In particular, if k is large enough then $(t_k, x_k) \in G$, which implies that the solution $y(t)$ to the following IVP

$$\begin{cases} y' = f(t, y), \\ y(t_k) = x_k, \end{cases}$$

is defined for all $t \in [t_0 - r/2, t_0 + r/2]$ (see the diagram below).



Since $x(t)$ also solves this IVP, the union $z(t)$ of $x(t)$ and $y(t)$ solves the same IVP. Note that $x(t)$ is defined only for $t > t_0$ while $z(t)$ is defined also for $t \in [t_0 - r/2, t_0]$. Hence, the solution $x(t)$ can be extended to a larger interval, which contradicts the maximality of $x(t)$. ■

Remark. By definition, a maximal solution $x(t)$ is defined on an open interval, say (a, b), and it cannot be extended to a larger open interval. One may wonder if $x(t)$ can be extended at least to the endpoints $t = a$ or $t = b$. It turns out that this is never the case (unless the domain Ω of the function $f(t, x)$ can be enlarged). Indeed, if $x(t)$ can be defined as a solution to the ODE also for $t = a$ then $(a, x(a)) \in \Omega$ and, hence, there is ball B in $\mathbb{R}^{n+1}$ centered at the point $(a, x(a))$ such that $B \subset \Omega$. By shrinking the radius of B, we can assume that the corresponding closed ball $\overline{B}$ is also contained in Ω. Since $x(t) \to x(a)$ as $t \to a$, we obtain that $(t, x(t)) \in \overline{B}$ for all t close enough to a. Therefore, the solution $x(t)$ does not leave the compact set $\overline{B} \subset \Omega$ as $t \to a$, which contradicts part (c) of Theorem 2.8.

2.6 Continuity of solutions with respect to $f(t,x)$

Consider the IVP

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases}$$

In one of the previous sections, we have considered in the one dimensional case the question how the solution $x(t)$ depends on the initial value x_0 thus allowing x_0 to vary. This question can be is a particular case of a more general question how the solution $x(t)$ depends on the right hand side $f(t,x)$. Indeed, consider the function $y(t) = x(t) - x_0$, which obviously solves the IVP

$$\begin{cases} y' = f(t, y + x_0), \\ y(t_0) = 0. \end{cases}$$

Hence, for $y(t)$, the initial value does not change while the right hand side does change when x_0 varies.

Consider now a more general question. Let Ω be an open set in $\mathbb{R}^{n+1}$ and f, g be two functions from Ω to $\mathbb{R}^n$. Assume that both f, g are continuous and locally Lipschitz in x, and consider two initial value problems

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} \tag{2.25}$$

and

$$\begin{cases} y' = g(t,y) \\ y(t_0) = x_0 \end{cases} \tag{2.26}$$

where (t_0, x_0) is a fixed point in Ω.

Assume that the function f as fixed and $x(t)$ is a fixed solution of (2.25). However, the function g can be chosen. Our purpose is to show that if g is chosen close enough to f then the solution $y(t)$ of (2.26) is close enough to $x(t)$. Apart from the theoretical interest, this question has significant practical consequences. For example, if one knows the function $f(t,x)$ only approximately then solving (2.25) approximately means solving another problem (2.26) where g is an approximation to f. Hence, it is important to know that the solution $y(t)$ is actually an approximation of $x(t)$.

Theorem 2.10 *Let $x(t)$ be a solution to the IVP (2.25) defined on an interval (a,b). Then, for all $\alpha < \beta$ such that $t_0 \in [\alpha, \beta] \subset (a,b)$, and for any $\varepsilon > 0$, there is $\eta > 0$ such that, for any function $g : \Omega \to \mathbb{R}^n$ with the property*

$$\sup_{\Omega} \|f - g\| \leq \eta, \tag{2.27}$$

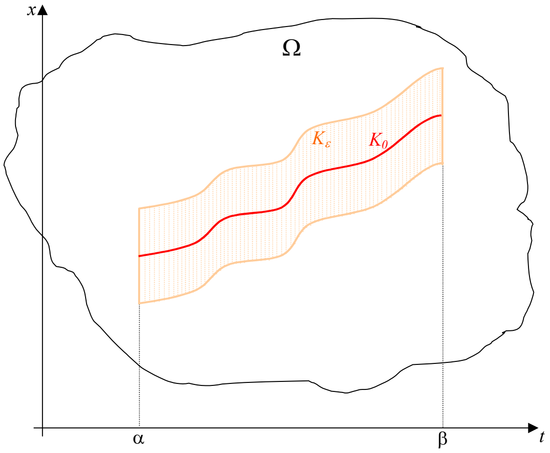
there is a solution $y(t)$ of the IVP (2.26) defined in $[\alpha, \beta]$, and this solution satisfies

$$\sup_{[\alpha,\beta]} \|x(t) - y(t)\| \leq \varepsilon.$$

Proof. For any $\varepsilon \geq 0$, consider the set

$$K_\varepsilon = \left\{(t,x) \in \mathbb{R}^{n+1} : \alpha \leq t \leq \beta,\ \ \|x - x(t)\| \leq \varepsilon\right\} \tag{2.28}$$

which can be regarded as the ε-neighborhood in $\mathbb{R}^{n+1}$ of the graph of the function $t \mapsto x(t)$ where $t \in [\alpha, \beta]$. In particular, K_0 is the graph of this function (see the diagram below).



It is easy to see that K_ε is bounded and closed; hence, K_ε is a compact subset of $\mathbb{R}^{n+1}$

Claim 1. *There are positive ε and L such that $K_\varepsilon \subset \Omega$ and*

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

for all $(t,x), (t,y) \in K_\varepsilon$. That is, f is Lipschitz in x on the set K_ε.

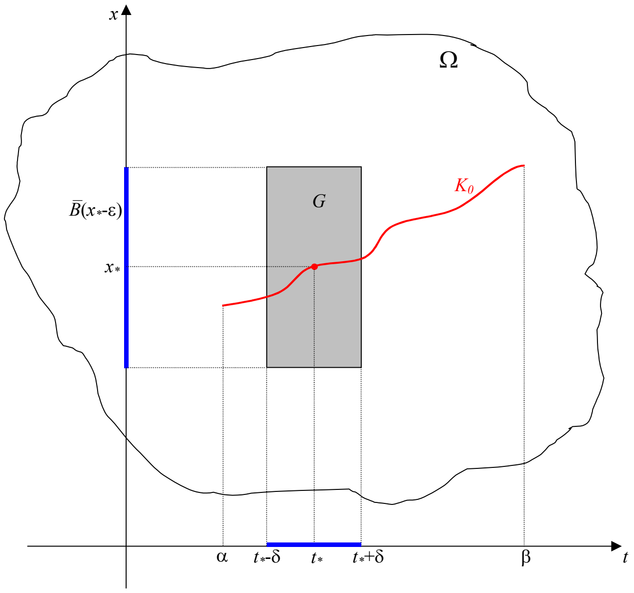
By the local Lipschitz condition, for any point $(t_*, x_*) \in \Omega$ (in particular, for any $(t_*, x_*) \in K_0$), there are constants ε, δ, L such that the cylinder

$$G = [t_* - \delta, t_* + \delta] \times \overline{B}(x_*, \varepsilon)$$

is contained in Ω and

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

for all $(t,x), (t,y) \in G$ (see the diagram below).



Varying the point (t_*, x_*) in K_0, we obtain a cover of K_0 by open cylinders of the type $(t_* - \delta, t_* + \delta) \times B(x_*, \varepsilon/2)$ where ε, δ (and L) depend on (t_*, x_*). Since K_0 is compact, there is a finite subcover, that is, a finite number of points $\{(t_i, x_i)\}_{i=1}^N$ on K_0 and the corresponding numbers $\varepsilon_i, \delta_i, L_i$ such that the cylinders $G_i = (t_i - \delta_i, t_i + \delta_i) \times B(x_i, \varepsilon_i/2)$ cover all K_0 and

$$\|f(t,x) - f(t,y)\| \le L_i \|x - y\|$$

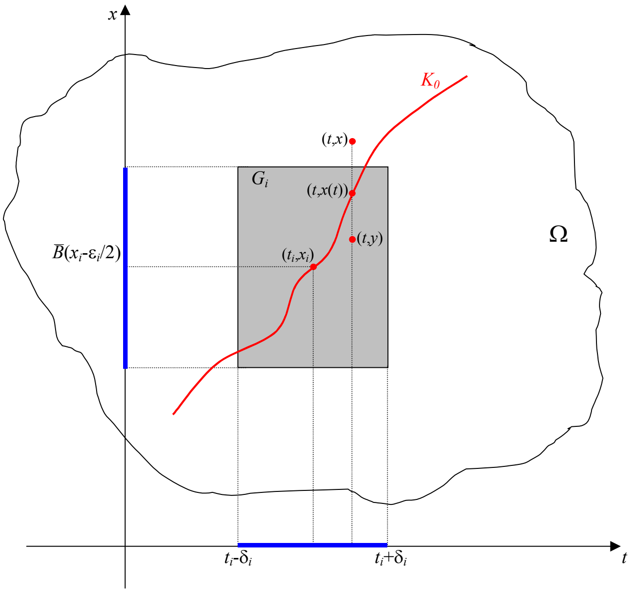
for all $t \in [t_i - \delta_i, t_i + \delta_i]$ and $x, y \in \overline{B}(x_i, \varepsilon_i)$. Set

$$\varepsilon = \frac{1}{2} \min_i \varepsilon_i \text{ and } L = \max_i L_i$$

and prove that the Lipschitz condition holds in K_ε with the constant L. For any two points $(t,x), (t,y) \in K_\varepsilon$, we have $t \in [\alpha, \beta]$, $(t, x(t)) \in K_0$ and

$$\|x - x(t)\| \le \varepsilon \text{ and } \|y - x(t)\| \le \varepsilon.$$

The point $(t, x(t))$ belongs to one of the cylinders G_i so that $t \in (t_i - \delta_i, t_i + \delta_i)$ and $\|x(t) - x_i\| < \varepsilon_i/2$ (see the diagram below).



By the triangle inequality, we have

$$\|x - x_i\| \le \|x - x(t)\| + \|x(t) - x_i\| < \varepsilon + \varepsilon_i/2 \le \varepsilon_i,$$

where we have used that $\varepsilon \le \varepsilon_i/2$. In the same way one proves that $\|y - x_i\| < \varepsilon_i$. Therefore, x and y belong to the ball $B(x_i, \varepsilon_i)$ whence it follows, by the choice of ε_i and δ_i, that

$$\|f(t,x) - f(t,y)\| \le L_i \|x - y\| \le L \|x - y\|,$$

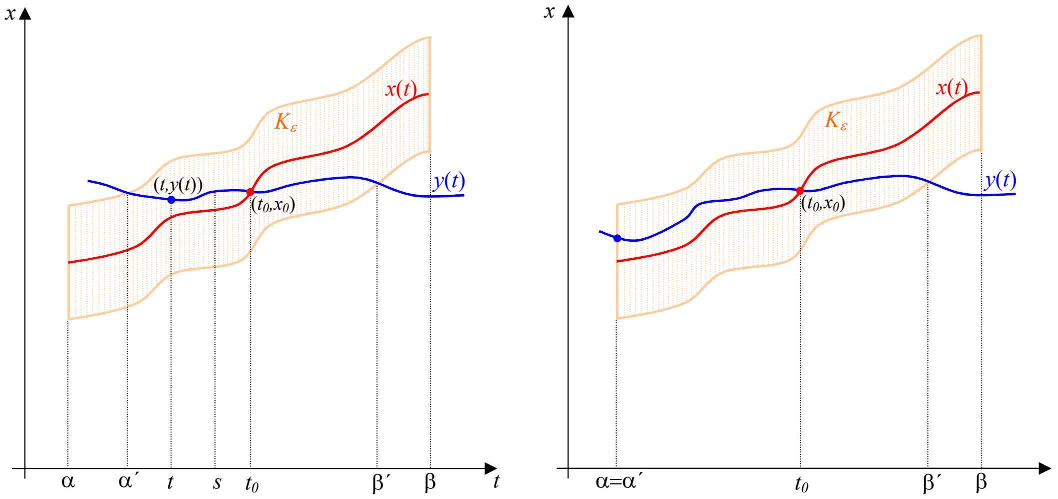
which finishes the proof of Claim 1.

Observe that if the statement of Claim 1 holds for some value of ε then it holds for all smaller values of ε as well, with the same L. Hence, we can assume that the value of ε from Theorem 2.10 is small enough so that it satisfies the statement of Claim 1.

Let now $y(t)$ be the maximal solution to the IVP (2.26), and let (a', b') be its domain. By Theorem 2.8, the graph of $y(t)$ leaves K_ε when $t \to a'$ and when $t \to b'$. Let (α', β') be the maximal interval such that the graph of $y(t)$ on this interval is contained in K_ε, that is,

$$\alpha' = \inf \{t \in (\alpha, \beta) \cap (a', b') : (t, y(t)) \in K_\varepsilon \text{ and } (s, y(s)) \in K_\varepsilon \text{ for all } s \in (t, t_0)\} \quad (2.29)$$

and β' is defined similarly with inf replaced by sup (see the diagrams below for the cases $\alpha' > \alpha$ and $\alpha' = \alpha$, respectively).



This definition implies that (α', β') is contained in $(a', b') \cap (\alpha, \beta)$, function $y(t)$ is defined on (α', β') and by (2.29)

$$(t, y(t)) \in K_\varepsilon \text{ for all } t \in (\alpha', \beta'). \quad (2.30)$$

Claim 2. *We have $[\alpha', \beta'] \subset (a', b')$. In particular, $y(t)$ is defined on $[\alpha', \beta']$. Moreover, the following is true: either $\alpha' = \alpha$ or $\alpha' > \alpha$ and*

$$\|x(t) - y(t)\| = \varepsilon \text{ for } t = \alpha'. \quad (2.31)$$

A similar statement holds for β' and β.

By Theorem 2.8, $y(t)$ leaves K_ε as $t \to a'$. Hence, for all values of t close enough to a' we have $(t, y(t)) \notin K_\varepsilon$. For any such t we have by (2.29) $t \le \alpha'$ whence $a' < t \le \alpha$ and $a' < \alpha'$. Similarly, one shows that $b' > \beta'$, whence $[\alpha', \beta'] \subset [a', b']$.

To prove the second part, assume that $\alpha' \ne \alpha$ that is, $\alpha' > \alpha$, and prove that

$$\|x(t) - y(t)\| = \varepsilon \text{ for } t = \alpha'.$$

The condition $\alpha' > \alpha$ together with $\alpha' > a'$ implies that α' belongs to the open interval $(\alpha, \beta) \cap (a', b')$. It follows that, for $\tau > 0$ small enough,

$$(\alpha' - \tau, \alpha' + \tau) \subset (\alpha, \beta) \cap (a', b'). \quad (2.32)$$

For any $t \in (\alpha', \beta')$, we have

$$\|x(t) - y(t)\| \le \varepsilon.$$

By the continuity, this inequality extends also to $t = \alpha'$. We need to prove that, for $t = \alpha'$, equality is attained here. Indeed, a strict inequality

$$\|x(t) - y(t)\| < \varepsilon$$

for $t = \alpha'$ implies by the continuity of $x(t)$ and $y(t)$, that the same inequality holds for all $t \in (\alpha' - \tau, \alpha' + \tau)$ provided $\tau > 0$ is small enough. Choosing τ to satisfy also (2.32), we obtain that $(t, y(t)) \in K_\varepsilon$ for all $t \in (\alpha' - \tau, \alpha']$, which contradicts the definition of α'.

Claim 3. *For any given $\alpha, \beta, \varepsilon, L$ as above, there exists $\eta > 0$ such that if*

$$\sup_{K_\varepsilon} \|f - g\| \leq \eta, \tag{2.33}$$

then $[\alpha', \beta'] = [\alpha, \beta]$.

In fact, Claim 3 will finish the proof of Theorem 2.10. Indeed, Claims 2 and 3 imply that $y(t)$ is defined on $[\alpha, \beta]$, and by (2.30) $(t, y(t)) \in K_\varepsilon$ for all $t \in (\alpha, \beta)$. By continuity, the latter inclusion extends to $t \in [\alpha, \beta]$. By (2.28), this means

$$\|y(t) - x(t)\| \leq \varepsilon \text{ for all } t \in [\alpha, \beta],$$

which was the claim of Theorem 2.10.

To prove Claim 3, for any $t \in [\alpha', \beta']$ use the integral identities

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

and

$$y(t) = x_0 + \int_{t_0}^{t} g(s, y(s))\, ds$$

whence

$$\begin{aligned} \|x(t) - y(t)\| &= \left\| \int_{t_0}^{t} (f(s, x(s)) - g(s, y(s)))\, ds \right\| \\ &\leq \left\| \int_{t_0}^{t} (f(s, x(s)) - f(s, y(s)))\, ds \right\| + \left\| \int_{t_0}^{t} (f(s, y(s)) - g(s, y(s)))\, ds \right\|. \end{aligned}$$

Assuming for simplicity that $t \geq t_0$ and noticing that the points $(s, x(s))$ and $(s, y(s))$ are in K_ε, we obtain by the Lipschitz condition in K_ε (Claim 1) and (2.33)

$$\|x(t) - y(t)\| \leq \int_{t_0}^{t} L \|x(s) - y(s)\|\, ds + \eta(\beta - \alpha). \tag{2.34}$$

Hence, by the Gronwall lemma applied to the function $z(t) = \|x(t) - y(t)\|$,

$$\begin{aligned} \|x(t) - y(t)\| &\leq \eta(\beta - \alpha) \exp L(t - t_0) \\ &\leq \eta(\beta - \alpha) \exp L(\beta - \alpha). \end{aligned}$$

Now choose η by

$$\eta = \frac{\varepsilon}{2(\beta - \alpha)} e^{-L(\beta - \alpha)}$$

so that

$$\|x(t) - y(t)\| \leq \varepsilon/2 \text{ for all } t \in [\alpha', \beta']. \tag{2.35}$$

It follows from Claim 2 that $\alpha' = \alpha$ because otherwise we would have (2.31), which contradicts (2.35). In the same way, $\beta' = \beta$, which finishes the proof. ■

Using the proof of Theorem 2.10, we can refine the statement of Theorem 2.10 as follows.

Theorem 2.10 ′ *Under conditions of Theorem* 2.10, *let* $x(t)$ *be a solution to the IVP* (2.25) *defined on an interval* (a,b), *and let* $[\alpha,\beta]$ *be an interval such that* $t_0 \in [\alpha,\beta] \subset (a,b)$. *Let* $\varepsilon > 0$ *be sufficiently small so that the Lipschitz condition holds in* K_ε *with a constant* L, *and set*

$$C = 2(\beta-\alpha)e^{L(\beta-\alpha)}.$$

Then the solution $y(t)$ *of the IVP* (2.26) *is defined on* $[\alpha,\beta]$ *and*

$$\sup_{[\alpha,\beta]}\|x(t)-y(t)\| \leq C\sup_{K_\varepsilon}\|f-g\|, \tag{2.36}$$

provided $\sup_{K_\varepsilon}\|f-g\|$ *is sufficiently small.*

Proof. Fix ε as above and introduce one more parameter $\varepsilon' \leq \varepsilon$. Then $K_{\varepsilon'} \subset K_\varepsilon$ and the Lipschitz condition holds in $K_{\varepsilon'}$ with the same constant L. Using Claim 3 from the proof of Theorem 2.10 with ε' instead of ε, we conclude that if

$$\sup_{K_{\varepsilon'}}\|f-g\| \leq \eta \tag{2.37}$$

where η satisfies

$$\eta(\beta-\alpha)\exp(L(\beta-\alpha)) = \varepsilon'/2,$$

that is, $C\eta = \varepsilon'$, then the maximal solution $y(t)$ of the IVP (2.26) is defined on $[\alpha,\beta]$ and

$$\sup_{[\alpha,\beta]}\|x(t)-y(t)\| \leq \varepsilon'.$$

Replacing $K_{\varepsilon'}$ in (2.37) by a larger set K_ε, we obtain, in particular, that if $\sup_{K_\varepsilon}\|f-g\|$ is sufficiently small then $y(t)$ is defined on $[\alpha,\beta]$. Furthermore, replacing η by $C^{-1}\varepsilon'$, we obtain that

$$\sup_{K_\varepsilon}\|f-g\| \leq C^{-1}\varepsilon' \tag{2.38}$$

implies

$$\sup_{[\alpha,\beta]}\|x(t)-y(t)\| \leq \varepsilon'.$$

Choosing ε' so that equality holds in (2.38), we obtain (2.36). ■

2.7 Continuity of solutions with respect to a parameter

Consider the IVP with a parameter $s \in \mathbb{R}^m$

$$\begin{cases} x' = f(t,x,s) \\ x(t_0) = x_0 \end{cases} \tag{2.39}$$

where $f : \Omega \to \mathbb{R}^n$ and Ω is an open subset of $\mathbb{R}^{n+m+1}$. Here the triple (t,x,s) is identified as a point in $\mathbb{R}^{n+m+1}$ as follows:

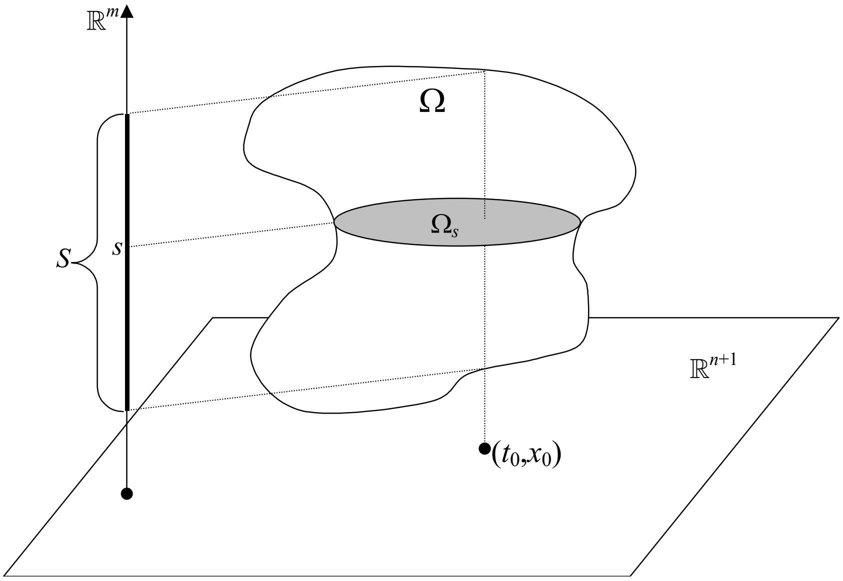
$$(t,x,s) = (t,x_1,..,x_n,s_1,...,s_m).$$

How do we understand (2.39)? For any $s \in \mathbb{R}^m$, consider the open set

$$\Omega_s = \left\{(t, x) \in \mathbb{R}^{n+1} : (t, x, s) \in \Omega\right\}.$$

Denote by S the set of those s, for which Ω_s contains (t_0, x_0), that is,

$$\begin{aligned} S &= \{s \in \mathbb{R}^m : (t_0, x_0) \in \Omega_s\} \\ &= \{s \in \mathbb{R}^m : (t_0, x_0, s) \in \Omega\} \end{aligned}$$



Then the IVP (2.39) can be considered in the domain Ω_s for any $s \in S$. We always assume that the set S is non-empty. Assume also in the sequel that $f(t, x, s)$ is a continuous function in $(t, x, s) \in \Omega$ and is locally Lipschitz in x for any $s \in S$. For any $s \in S$, denote by $x(t, s)$ the maximal solution of (2.39) and let I_s be its domain (that is, I_s is an open interval on the axis t). Hence, $x(t, s)$ as a function of (t, s) is defined in the set

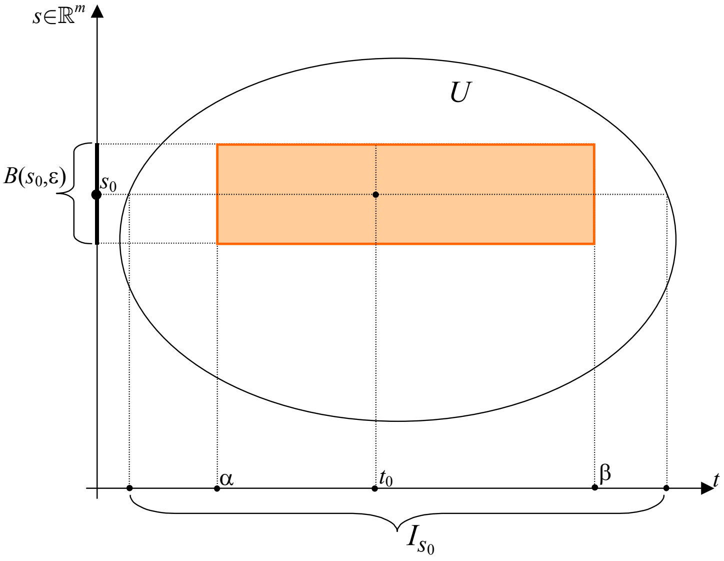
$$U = \left\{(t, s) \in \mathbb{R}^{m+1} : s \in S, t \in I_s\right\}.$$

Theorem 2.11 *Under the above assumptions, the set U is an open subset of $\mathbb{R}^{n+1}$ and the function $x(t, s) : U \to \mathbb{R}^n$ is continuous.*

Proof. Fix some $s_0 \in S$ and consider solution $x(t) = x(t, s_0)$ defined for $t \in I_{s_0}$. Choose some interval $[\alpha, \beta] \subset I_{s_0}$ such that $t_0 \in [\alpha, \beta]$. We will prove that there is $\varepsilon > 0$ such that

$$[\alpha, \beta] \times B(s_0, \varepsilon) \subset U, \tag{2.40}$$

which will imply that U is open. Here $B(s_0, \varepsilon)$ is a ball in $\mathbb{R}^m$ with respect to ∞-norm (we can assume that all the norms in various spaces $\mathbb{R}^k$ are the ∞-norms).



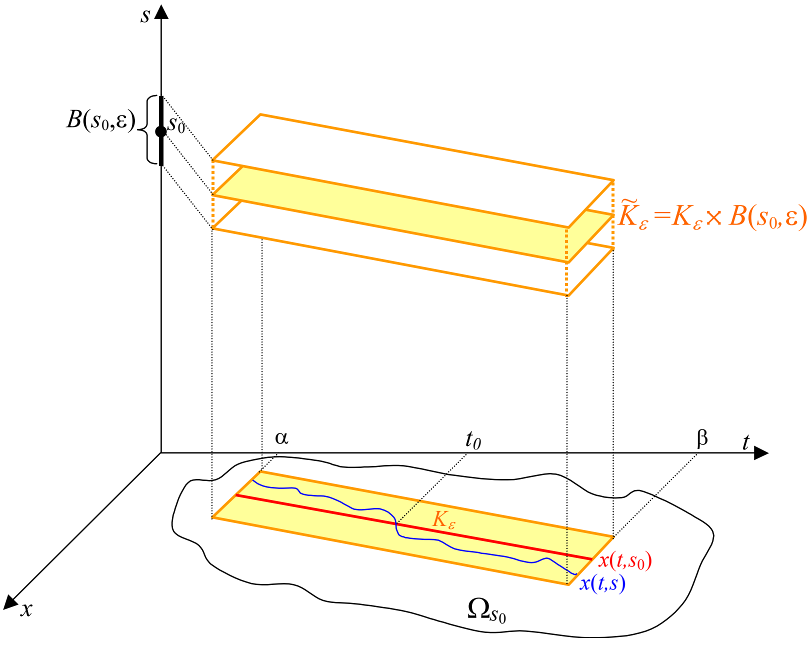
As in the proof of Theorem 2.10, consider a set

$$K_\varepsilon = \left\{(t,x) \in \mathbb{R}^{n+1} : \alpha \le t \le \beta,\ \ \|x - x(t)\| \le \varepsilon\right\}$$

and its extension in $\mathbb{R}^{n+m+1}$ defined by

$$\begin{aligned} \widetilde{K}_\varepsilon &= \left\{(t,x,s) \in \mathbb{R}^{n+m+1} : \alpha \le t \le \beta, \|x - x(t)\| \le \varepsilon, \|s - s_0\| \le \varepsilon\right\} \\ &= K_\varepsilon \times B(s_0,\varepsilon) \end{aligned}$$

(see the diagram below).



If ε is small enough then $\widetilde{K}_\varepsilon$ is contained in Ω (cf. the proof of Theorem 2.10 and Exercise 26). Hence, for any $s \in B(s_0,\varepsilon)$, the function $f(t,x,s)$ is defined for all $(t,x) \in$

K_ε. Since the function f is continuous on Ω, it is uniformly continuous on the compact set $\widetilde{K}_\varepsilon$, whence it follows that

$$\sup_{(t,x)\in K_\varepsilon} \|f(t,x,s_0) - f(t,x,s)\| \to 0 \text{ as } s \to s_0.$$

Using Theorem 2.10$'$with[2] $f(t,x) = f(t,x,s_0)$ and $g(t,x) = f(t,x,s)$ where $s \in B(s_0,\varepsilon)$, we obtain that if

$$\sup_{(t,x)\in K_\varepsilon} \|f(t,x,s) - f(t,x,s_0)\|$$

is small enough then then the solution $y(t) = x(t,s)$ is defined on $[\alpha,\beta]$. In particular, this implies (2.40) for small enough ε. Furthermore, by Theorem 2.10$'$ we also obtain that

$$\sup_{t\in[\alpha,\beta]} \|x(t,s) - x(t,s_0)\| \le C \sup_{(t,x)\in K_\varepsilon} \|f(t,x,s_0) - f(t,x,s)\|,$$

where the constant C depending only on α,β,ε and the Lipschitz constant L of the function $f(t,x,s_0)$ in K_ε. Letting $s \to s_0$, we obtain that

$$\sup_{t\in[\alpha,\beta]} \|x(t,s) - x(t,s_0)\| \to 0 \text{ as } s \to s_0,$$

so that $x(t,s)$ is continuous in s uniformly in $t \in [\alpha,\beta]$. Since $x(t,s)$ is continuous in t for any fixed s, we conclude that x is continuous in (t,s) (see Exercise 27), which finishes the proof. ■

2.8 Global existence

Theorem 2.12 *Let I be an open interval in $\mathbb{R}$. Assume that a function $f(t,x) : I\times\mathbb{R}^n \to \mathbb{R}^n$ is continuous, locally Lipschitz in x, and satisfies the inequality*

$$\|f(t,x)\| \le a(t)\|x\| + b(t) \tag{2.41}$$

for all $t \in I$ and $x \in \mathbb{R}^n$, where $a(t)$ and $b(t)$ are some continuous non-negative functions of t. Then, for all $t_0 \in I$ and $x_0 \in \mathbb{R}^n$, the initial value problem

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} \tag{2.42}$$

has a (unique) solution $x(t)$ on I.

Proof. Let $x(t)$ be the maximal solution to the problem (2.42), and let $J = (\alpha,\beta)$ be the open interval where $x(t)$ is defined. We will show that $J = I$. Assume from the contrary that this is not the case. Then one of the points α,β is contained in I, say $\beta \in I$. What can happen to $x(t)$ when $t \to \beta$? By Theorem 2.8, $(t,x(t))$ leaves any compact $K \subset \Omega := I \times \mathbb{R}^n$. Consider a compact set $K = [\beta-\varepsilon,\beta] \times \overline{B}(0,r)$ where $\varepsilon > 0$ is so small that $[\beta-\varepsilon,\beta] \subset I$. Clearly, $K \subset \Omega$. If t is close enough to β then $t \in [\beta-\varepsilon,\beta]$. Since $(t,x(t))$ must be outside K, we conclude that $x \notin \overline{B}(0,r)$, that is, $\|x(t)\| > r$. In other words, we see that $\|x(t)\| \to \infty$ as $t \to \beta$.

[2] Since the common domain of the functions $f(t,x,s)$ and $f(t,x,s_0)$ is $(t,s) \in \Omega_{s_0} \cap \Omega_s$, Theorem 2.10 should be applied with this domain.

On the other hand, let us show that the solution $x(t)$ remains bounded when $t \to \beta$. From the integral equation

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

we obtain, for any $t \in [t_0, \beta)$

$$\begin{aligned} \|x(t)\| &\leq \|x_0\| + \int_{t_0}^{t} \|f(s, x(s))\|\, ds \\ &\leq \|x_0\| + \int_{t_0}^{t} (a(s)\|x(s)\| + b(s))\, ds \\ &\leq C + A\int_{t_0}^{t} \|x(s)\|\, ds, \end{aligned}$$

where

$$A = \sup_{[t_0,\beta]} a(s) \quad \text{and} \quad C = \|x_0\| + \int_{t_0}^{\beta} b(s)\, ds.$$

Since $[t_0, \beta] \subset I$ and functions $a(s)$ and $b(s)$ are continuous in $[t_0, \beta]$, the values of A and C are finite. The Gronwall lemma yields

$$\|x(t)\| \leq C\exp(A(t - t_0)) \leq C\exp(A(\beta - t_0)).$$

Since the right hand side here does not depend on t, we conclude that the function $\|x(t)\|$ remains bounded as $t \to \beta$, which finishes the proof. ■

Example. We have considered above the ODE $x' = x^2$ defined in $\mathbb{R} \times \mathbb{R}$ and have seen that the solution $x(t) = \frac{1}{C-t}$ cannot be defined on full $\mathbb{R}$. The same occurs for the equation $x' = x^\alpha$ for $\alpha > 1$. The reason is that the function $f(t,x) = x^\alpha$ does not admit the estimate (2.41) for large x, due to $\alpha > 1$. This example also shows that the condition (2.41) is rather sharp.

A particularly important application of Theorem 2.12 is the case of the *linear* equation

$$x' = A(t)x + B(t),$$

where $x \in \mathbb{R}^n$, $t \in I$ (where I is an open interval in $\mathbb{R}$), $B : I \to \mathbb{R}^n$, $A : I \to \mathbb{R}^{n\times n}$. Here $\mathbb{R}^{n\times n}$ is the space of all $n \times n$ matrices (that can be identified with $\mathbb{R}^{n^2}$). In other words, for each $t \in I$, $A(t)$ is an $n \times n$ matrix, and $A(t)x$ is the product of the matrix $A(t)$ and the column vector x. In the coordinate form, one has a system of linear equations

$$x_k' = \sum_{i=1}^{n} A_{ki}(t)\, x_i + B_k(t),$$

for any $k = 1, ..., n$.

Theorem 2.13 *Let $A(t)$ and $B(t)$ be continuous in an open interval $I \subset \mathbb{R}$. Then, for any $t_0 \in I$ and $x_0 \in \mathbb{R}^n$, the IVP*

$$\begin{cases} x' = A(t)x + B(t) \\ x(t_0) = x_0 \end{cases}$$

has a (unique) solution $x(t)$ defined on I.

Proof. It suffices to check that the function $f(t,x) = A(t)x + B(t)$ satisfies the conditions of Theorem 2.12. This function is obviously continuous in (t,x). Let us show that $\|A(t)x\| \leq a(t)\|x\|$ for a continuous function $a(t)$. Indeed, using the ∞-norm, we have

$$\|A(t)x\| = \max_k |(A(t)x)_k| = \max_k \left|\sum_l A_{kl}(t)x_l\right| \leq \max_k \left|\sum_l A_{kl}(t)\right| \max_l |x_l| = a(t)\|x\|$$

where $a(t) = \max_k |\sum_l A_{kl}(t)|$ is a continuous function. Setting also $b(t) = \|B(t)\|$, we obtain

$$\|f(t,x)\| \leq \|A(t)x\| + \|B(t)\| \leq a(t)\|x\| + b(t).$$

Since function $f(t,x)$ is continuously differentiable in x, it is locally Lipschitz by Lemma 2.6. Alternatively, let us show that $f(t,x)$ is Lipschitz in x in any set of the form $[\alpha,\beta]\times\mathbb{R}^n$ where $[\alpha,\beta]$ is a closed bounded interval in I. Indeed, for any $t \in [\alpha,\beta]$ and $x,y \in \mathbb{R}^n$, we have

$$\|f(t,x) - f(t,y)\| = \|A(t)(x-y)\| \leq a(t)\|x-y\| \leq L\|x-y\|$$

where

$$L = \sup_{t\in[\alpha,\beta]} a(t).$$

■

2.9 Differentiability of solutions in parameter

Before we can state and prove the main result, let us prove a lemma from Analysis.

Definition. A set $K \subset \mathbb{R}^n$ is called *convex* if for any two points $x,y \in K$, also the full interval $[x,y]$ is contained in K, that is, the point $(1-\lambda)x + \lambda y$ belong to K for any $\lambda \in [0,1]$.

Example. Let us show that any ball $B(z,r)$ in $\mathbb{R}^n$ with respect to any norm is convex. Indeed, it suffices to treat $z = 0$. If $x,y \in B(0,r)$ that is, $\|x\|$ and $\|y\|$ are smaller than r then also

$$\|(1-\lambda)x + \lambda y\| \leq (1-\lambda)\|x\| + \lambda\|y\| < r$$

so that $(1-\lambda)x + \lambda y \in B(0,r)$.

If $f(x,u)$ is a function of $x \in \mathbb{R}^n$ and some parameter u, and f takes values in $\mathbb{R}^l$ then denote by f_x the Jacobian matrix of f with respect to x, that is, the $l \times n$ matrix defined by

$$f_x = \frac{\partial f}{\partial x} = \left(\frac{\partial f_k}{\partial x_j}\right),$$

where $k = 1,...,l$ is the row index and $j = 1,...,n$ is the column index. In particular, if $n = l = 1$ then f_x is just the partial derivative of f in x.

Lemma 2.14 (The Hadamard lemma) *Let $f(t,x)$ be a continuous mapping from Ω to $\mathbb{R}^l$ where Ω is an open subset of $\mathbb{R}^{n+1}$ such that, for any $t \in \mathbb{R}$, the set*

$$\Omega_t = \{x \in \mathbb{R}^n : (t,x) \in \Omega\}$$

is convex (see the diagram below). Assume that $f_x(t,x)$ exists and is also continuous in Ω. Consider the domain

$$\begin{aligned}\Omega' &= \left\{(t,x,y)\in\mathbb{R}^{2n+1} : t\in\mathbb{R},\ x,y\in\Omega_t\right\} \\ &= \left\{(t,x,y)\in\mathbb{R}^{2n+1} : (t,x)\ \textit{and}\ (t,y)\in\Omega\right\}.\end{aligned}$$

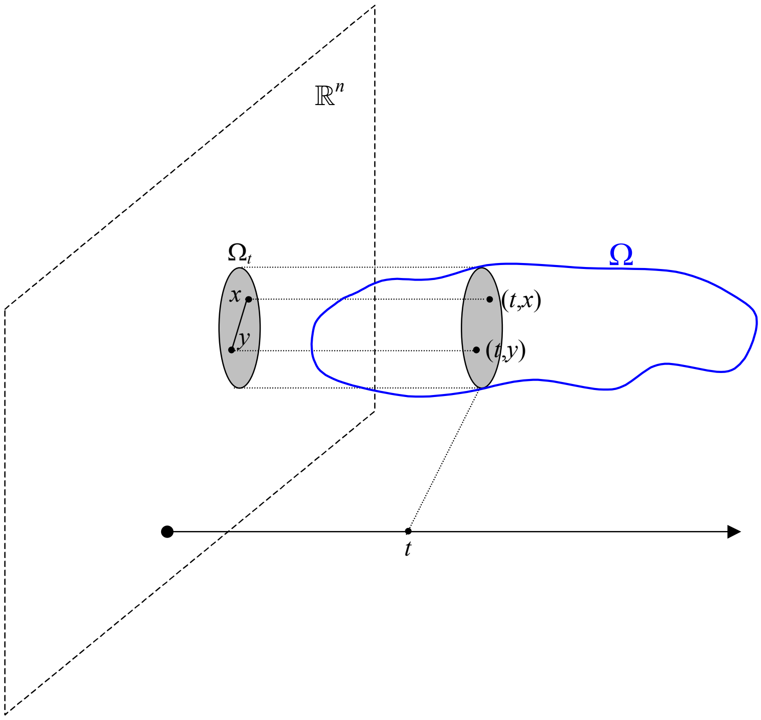
Then there exists a continuous mapping $\varphi(t,x,y):\Omega'\to\mathbb{R}^{l\times n}$ such that the following identity holds:

$$f(t,y)-f(t,x)=\varphi(t,x,y)(y-x)$$

for all $(t,x,y)\in\Omega'$ (here $\varphi(t,x,y)(y-x)$ is the product of the $l\times n$ matrix and the column-vector).

Furthermore, we have for all $(t,x)\in\Omega$ the identity

$$\varphi(t,x,x)=f_x(t,x). \tag{2.43}$$



Remark. The variable t can be higher dimensional, and the proof goes through without changes.

Since $f(t,x)$ is continuously differentiable at x, we have

$$f(t,y)-f(t,x)=f_x(t,x)(y-x)+o(\|y-x\|)\ \text{as}\ y\to x.$$

The point of the above Lemma is that the term $o(\|x-y\|)$ can be eliminated if one replaces $f_x(t,x)$ by a continuous function $\varphi(t,x,y)$.

Example. Consider some simple examples of functions $f(x)$ with $n=l=1$ and without dependence on t. Say, if $f(x)=x^2$ then we have

$$f(y)-f(x)=(y+x)(y-x)$$

so that $\varphi(x,y)=y+x$. In particular, $\varphi(x,x)=2x=f'(x)$. Similar formula holds for $f(x)=x^k$ with any $k\in\mathbb{N}$:

$$f(y)-f(x)=\left(x^{k-1}+x^{k-2}y+...+y^{k-1}\right)(y-x).$$

For any continuously differentiable function $f(x)$, one can define $\varphi(x,y)$ as follows:

$$\varphi(x,y)=\begin{cases}\frac{f(y)-f(x)}{y-x}, & y\neq x,\\ f'(x), & y=x.\end{cases}$$

It is obviously continuous in (x,y) for $x\neq y$, and it is continuous at (x,x) because if $(x_k,y_k)\to(x,x)$ as $k\to\infty$ then

$$\frac{f(y_k)-f(x_k)}{y_k-x_k}=f'(\xi_k)$$

where $\xi_k\in(x_k,y_k)$, which implies that $\xi_k\to x$ and hence, $f'(\xi_k)\to f'(x)$, where we have used the continuity of the derivative $f'(x)$.

Clearly, this argument will not work in the higher dimensional case, so one needs a different approach.

Proof of Lemma 2.14. It suffices to prove this lemma for each component f_i separately. Hence, we can assume that $l=1$ so that φ is a row $(\varphi_1,...,\varphi_n)$. Hence, we need to prove the existence of n real valued continuous functions $\varphi_1,...,\varphi_n$ of (t,x,y) such that the following identity holds:

$$f(t,y)-f(t,x)=\sum_{i=1}^n\varphi_i(t,x,y)(y_i-x_i).$$

Fix a point $(t,x,y)\in\Omega'$ and consider a function

$$F(\lambda)=f(t,x+\lambda(y-x))$$

on the interval $\lambda\in[0,1]$. Since $x,y\in\Omega_t$ and Ω_t is convex, the point $x+\lambda(y-x)$ belongs to Ω_t. Therefore, $(t,x+\lambda(y-x))\in\Omega$ and the function $F(\lambda)$ is indeed defined for all $\lambda\in[0,1]$. Clearly, $F(0)=f(t,x)$, $F(1)=f(t,y)$. By the chain rule, $F(\lambda)$ is continuously differentiable and

$$F'(\lambda)=\sum_{i=1}^n f_{x_i}(t,x+\lambda(y-x))(y_i-x_i).$$

By the fundamental theorem of calculus, we obtain

$$\begin{aligned}f(t,y)-f(t,x) &= F(1)-F(0)\\ &= \int_0^1 F'(\lambda)\,d\lambda\\ &= \sum_{i=1}^n\int_0^1 f_{x_i}(t,x+\lambda(y-x))(y_i-x_i)\,d\lambda\\ &= \sum_{i=1}^n\varphi_i(t,x,y)(y_i-x_i)\end{aligned}$$

where

$$\varphi_i(t,x,y)=\int_0^1 f_{x_i}(t,x+\lambda(y-x))\,d\lambda. \tag{2.44}$$

We are left to verify that φ_i is continuous. Observe first that the domain Ω' of φ_i is an open subset of $\mathbb{R}^{2n+1}$. Indeed, if $(t,x,y) \in \Omega'$ then (t,x) and $(t,y) \in \Omega$ which implies by the openness of Ω that there is $\varepsilon > 0$ such that the balls $B\left((t,x),\varepsilon\right)$ and $B\left((t,y),\varepsilon\right)$ in $\mathbb{R}^{n+1}$ are contained in Ω. Assuming the norm in all spaces in question is the ∞-norm, we obtain that $B\left((t,x,y),\varepsilon\right) \subset \Omega'$. The continuity of φ_i follows from the following general statement.

Lemma 2.15 *Let $f(\lambda,u)$ be a continuous real-valued function on $[a,b] \times U$ where U is an open subset of $\mathbb{R}^k$, $\lambda \in [a,\beta]$ and $u \in U$. Then the function*

$$\varphi(u) = \int_a^b f(\lambda,u)\,d\lambda$$

is continuous in $u \in U$.

The proof of Lemma 2.14 is then finished as follows. Consider $f_{x_i}\left(t, x+\lambda(y-x)\right)$ as a function of $(\lambda,t,x,y) \in [0,1] \times \Omega'$. This function is continuous in (λ,t,x,y), which implies by Lemma 2.15 that also $\varphi_i(t,x,y)$ is continuous in (t,x,y).

Finally, if $x=y$ then $f_{x_i}\left(t,x+\lambda(y-x)\right) = f_{x_i}(t,x)$ which implies by (2.44) that

$$\varphi_i(t,x,x) = f_{x_i}(t,x)$$

and, hence, $\varphi(t,x) = f_x(t,x)$, that is, (2.43). ■

Proof of Lemma 2.15. (This was Exercise 62 from Analysis II). The fact that $f(\lambda,u)$ is continuous in $[a,b] \times U$ implies that it is uniformly continuous on any compact set in this domain, in particular, in any set of the form $[a,b] \times K$ where K is a compact subset of U. In particular, if we have a convergent sequence in U

$$u_k \to u \text{ as } k \to \infty$$

then all u_k with large enough k can be put in a compact set K (say, a closed ball), whence it follows that the convergence

$$f(\lambda,u_k) \to f(\lambda,u) \text{ as } k \to \infty$$

is uniform in λ. Since one can exchange the operations of integration and uniform convergence, we conclude that also

$$\varphi(u_k) \to \varphi(u),$$

which proves the continuity of φ. ■

Consider again the initial value problem with parameter

$$\begin{cases} x' = f(t,x,s), \\ x(t_0) = x_0, \end{cases} \tag{2.45}$$

where $f : \Omega \to \mathbb{R}^n$ is a continuous function defined on an open set $\Omega \subset \mathbb{R}^{n+m+1}$ and where $(t,x,s) = (t,x_1,...,x_n,s_1,...,s_m)$. As above, denote by f_x the Jacobian matrix of f with respect to x, which is an $n \times n$ matrix. Similarly, denote by f_s the Jacobian matrix of f with respect to s, that is, f_s is the $n \times m$ matrix

$$f_s = \frac{\partial f}{\partial s} = \partial_s f = \left(\frac{\partial f_k}{\partial s_i}\right),$$

where k is the row index and i is the column index. If f_x is continuous in Ω then by Lemma 2.6 f is locally Lipschitz in x so that all the existence result apply. Let $x(t,s)$ be the maximal solution to (2.45). Recall that, by Theorem 2.11, the domain U of $x(t,s)$ is an open subset of $\mathbb{R}^{m+1}$ and $x : U \to \mathbb{R}^n$ is continuous.

Theorem 2.16 *Assume that function $f(t,x,s)$ is continuous and f_x and f_s exist and are also continuous in Ω. Then $x(t,s)$ is continuously differentiable in $(t,s) \in U$ and the Jacobian matrix $y = \partial_s x$ solves the initial value problem*

$$\begin{cases} y' = f_x(t, x(t,s), s)\, y + f_s(t, x(t,s), s), \\ y(t_0) = 0. \end{cases} \tag{2.46}$$

The linear ODE in (2.46) is called the *variational equation* for (2.45) along the solution $x(t,s)$ (or the equation in variations). Note that $y(t,s)$ is an $n \times m$ matrix and, hence, can be considered also as a vector in $\mathbb{R}^{nm}$. All terms in (2.46) are also $n \times m$ matrices. For example, $f_x y$ is the product of $n \times n$ matrix f_x by the $n \times m$ matrix y, which is hence an $n \times m$ matrix.

Let for a fixed s the domain of $x(t,s)$ be an interval I_s. Then the right hand side in (2.46) is defined in $I_s \times \mathbb{R}^{nm}$. Since this is a linear equation and its coefficients $f_x(t, x(t,s), s)$ and $f_s(t, x(t,s), s)$ are continuous in $t \in I_s$, we conclude by Theorem 2.13 that solution $y(t)$ exists in the full interval I_s. Hence, Theorem 2.16 can also be stated as follows: if $x(t,s)$ is the solution of (2.45) on I_s and $y(t)$ is the solution of (2.46) on I_s then the identity $y(t) = \partial_s x(t,s)$ takes place for all $t \in I_s$.

Example. Consider the IVP with parameter

$$\begin{cases} x' = x^2 + 2s/t \\ x(1) = -1 \end{cases}$$

in the domain $(0, +\infty) \times \mathbb{R} \times \mathbb{R}$ (that is, $t > 0$ and x, s are arbitrary real). The task is to find x and $\partial_s x$ for $s = 0$. Obviously, the function $f(t,x,s) = x^2 + 2s/t$ is continuously differentiable in (x,s) whence it follows that the solution $x(t,s)$ is continuously differentiable in (t,s).

For $s = 0$ we have the IVP

$$\begin{cases} x' = x^2 \\ x(1) = -1 \end{cases}$$

whence we obtain $x(t,0) = -\frac{1}{t}$. Setting $y = \partial_s x(t,0)$ and noticing that

$$f_x = 2x \text{ and } f_s = 2/t$$

we obtain the variational equation for y:

$$y' = \left(f_x|_{x=-\frac{1}{t}, s=0} \right) y + \left(f_s|_{x=-\frac{1}{t}, s=0} \right) = -\frac{2}{t} y + \frac{2}{t}.$$

This is the linear equation of the form $y' = a(t)\, y + b(t)$ which is solved by the formula

$$y = e^{A(t)} \int e^{-A(t)} b(t)\, dt,$$

where $A(t)$ is a primitive of $a(t)$, for example $A(t) = -2\ln t$. Hence,

$$y(t) = t^{-2}\int t^2 \frac{2}{t}dt = t^{-2}\left(t^2 + C\right) = 1 + Ct^{-2}.$$

The initial condition $y(1) = 0$ is satisfied for $C = -1$ so that $y(t) = 1 - t^{-2}$.

Expanding $x(t,s)$ as a function of s by the Taylor formula of the order 1, we obtain

$$x(t,s) = x(t,0) + \partial_s x(t,0)\,s + o(s) \text{ as } s \to 0,$$

whence

$$x(t,s) = -\frac{1}{t} + \left(1 - \frac{1}{t^2}\right)s + o(s) \text{ as } s \to 0.$$

Hence, the function

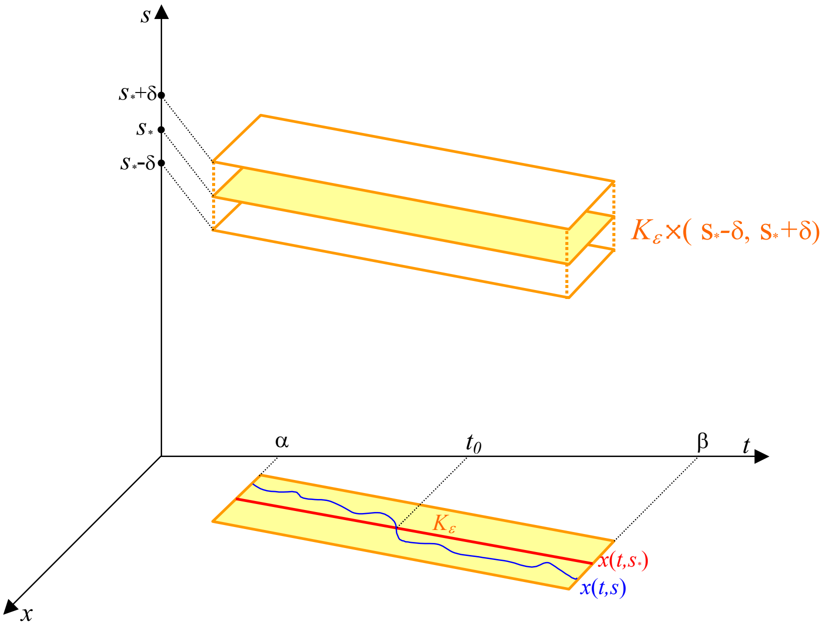
$$u(t) = -\frac{1}{t} + \left(1 - \frac{1}{t^2}\right)s$$

can be considered as an approximation for $x(t,s)$ for small s. Later on, we'll be able to obtain more terms in the Taylor formula and, hence, to get a better approximation for $x(t,s)$.

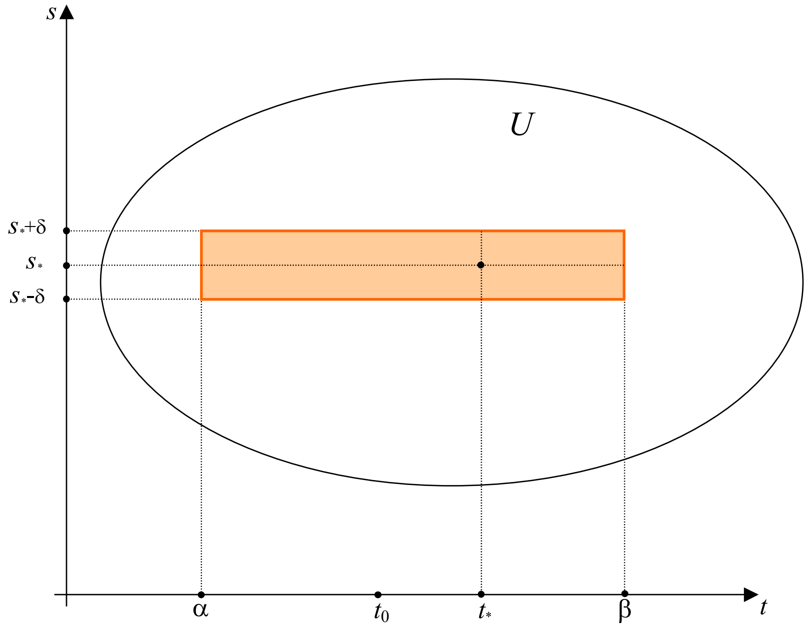
Proof of Theorem 2.16. In the main part of the proof, we show that the partial derivative $\partial_{s_i}x$ exists. Since this can be done separately for any component s_i, in this part we can and will assume that s is one-dimensional (that is, $m = 1$).

Fix some $(t_*, s_*) \in U$ and prove that $\partial_s x$ exists at this point. Since the differentiability is a local property, we can restrict the domain of the variables (t,s) as follows. Choose $[\alpha,\beta]$ to be any interval in I_{s_*}containing both t_0 and t_*. Then choose $\varepsilon, \delta > 0$ so small that the following conditions are satisfied (cf. the proof of Theorem 2.11):

1. The set
$$K_\varepsilon = \left\{(t,x) \in \mathbb{R}^{n+1} : \alpha < t < \beta, \|x - x(t,s_*)\| < \varepsilon\right\}$$
is contained in Ω_{s_*} and
$$K_\varepsilon \times (s_* - \delta, s_* + \delta) \subset \Omega.$$



2. The rectangle $(a,\beta)\times(s_*-\delta,s_*+\delta)$ is contained in U and, for all $s\in(s_*-\delta,s_*+\delta)$,
$$\sup_{t\in(\alpha,\beta)}\|x(t,s)-x(t,s_*)\|<\varepsilon,$$
that is, $(t,x(t,s))\in K_\varepsilon$.



In what follows, we restrict the domain of the variables (t,x,s) to $K_\varepsilon\times(s_*-\delta,s_*+\delta)$. Note that this domain is convex with respect to the variable (x,s), for any fixed t. Indeed, for a fixed t, x varies in the ball $B(x(t,s_*),\varepsilon)$ and s varies in the interval $(s_*-\delta,s_*+\delta)$, which are both convex sets.

Applying the Hadamard lemma to the function $f(t,x,s)$ in this domain and using the fact that f is continuously differentiable with respect to (x,s), we obtain the identity
$$f(t,y,s)-f(t,x,\sigma)=\varphi(t,x,\sigma,y,s)(y-x)+\psi(t,x,\sigma,y,s)(s-\sigma),$$
where φ and ψ are continuous functions on the appropriate domains. In particular, substituting $\sigma=s_*$, $x=x(t,s_*)$ and $y=x(t,s)$, we obtain
$$\begin{aligned}f(t,x(t,s),s)-f(t,x(t,s_*),s_*) &= \varphi(t,x(t,s_*),s_*,x(t,s),s)(x(t,s)-x(t,s_*))\\ &\quad+\psi(t,x(t,s_*),s_*,x(t,s),s)(s-s_*)\\ &= a(t,s)(x(t,s)-x(t,s_*))+b(t,s)(s-s_*),\end{aligned}$$
where the functions
$$a(t,s)=\varphi(t,x(t,s_*),s_*,x(t,s),s)\ \text{ and }\ b(t,s)=\psi(t,x(t,s_*),s_*,x(t,s),s)\qquad(2.47)$$
are continuous in $(t,s)\in(\alpha,\beta)\times(s_*-\delta,s_*+\delta)$ (the dependence on s_* is suppressed because s_* is fixed).

Set for any $s \in (s_* - \delta, s_* + \delta) \setminus \{s_*\}$

$$z(t,s) = \frac{x(t,s) - x(t,s_*)}{s - s_*}$$

and observe that

$$\begin{aligned} z' &= \frac{x'(t,s) - x'(t,s_*)}{s - s_*} = \frac{f(t, x(t,s), s) - f(t, x(t,s_*), s_*)}{s - s_*} \\ &= a(t,s) z + b(t,s). \end{aligned}$$

Note also that $z(t_0, s) = 0$ because both $x(t,s)$ and $x(t,s_*)$ satisfy the same initial condition. Hence, function $z(t,s)$ solves for any fixed $s \in (s_* - \delta, s_* + \delta) \setminus \{s_*\}$ the IVP

$$\begin{cases} z' = a(t,s) z + b(t,s) \\ z(t_0, s) = 0. \end{cases} \tag{2.48}$$

Since this ODE is linear and the functions a and b are continuous in $t \in (\alpha, \beta)$, we conclude by Theorem 2.13 that the solution to this IVP exists for all $s \in (s_* - \delta, s_* + \delta)$ and $t \in (\alpha, \beta)$ and, by Theorem 2.11, the solution is continuous in $(t,s) \in (\alpha, \beta) \times (s_* - \delta, s_* + \delta)$. Hence, we can define $z(t,s)$ also at $s = s_*$ as the solution of the IVP (2.48). In particular, using the continuity of $z(t,s)$ in s, we obtain

$$\lim_{s \to s_*} z(t,s) = z(t,s_*),$$

that is,

$$\partial_s x(t,s_*) = \lim_{s \to s_*} \frac{x(t,s) - x(t,s_*)}{s - s_*} = \lim_{s \to s_*} z(t,s) = z(t,s_*).$$

Hence, the derivative $y(t) = \partial_s x(t,s_*)$ exists and is equal to $z(t,s_*)$, that is, $y(t)$ satisfies the IVP

$$\begin{cases} y' = a(t,s_*) y + b(t,s_*), \\ y(t_0) = 0. \end{cases}$$

Note that by (2.47) and Lemma 2.14

$$a(t,s_*) = \varphi(t, x(t,s_*), s_*, x(t,s_*), s_*) = f_x(t, x(t,s_*), s_*)$$

and

$$b(t,s_*) = \psi(t, x(t,s_*), s_*, x(t,s_*), s_*) = f_s(t, x(t,s_*), s_*)$$

Hence, we obtain that $y(t)$ satisfies (2.46).

To finish the proof, we have to verify that $x(t,s)$ is continuously differentiable in (t,s). Here we come back to the general case $s \in \mathbb{R}^m$. The derivative $\partial_s x = y$ satisfies the IVP (2.46) and, hence, is continuous in (t,s) by Theorem 2.11. Finally, for the derivative $\partial_t x$ we have the identity

$$\partial_t x = f(t, x(t,s), s), \tag{2.49}$$

which implies that $\partial_t x$ is also continuous in (t,s). Hence, x is continuously differentiable in (t,s). ■

Remark. It follows from (2.49) that $\partial_t x$ is differentiable in s and, by the chain rule,

$$\partial_s\left(\partial_t x\right) = \partial_s\left[f\left(t, x\left(t,s\right), s\right)\right] = f_x\left(t, x\left(t,s\right), s\right)\partial_s x + f_s\left(t, x\left(t,s\right), s\right). \tag{2.50}$$

On the other hand, it follows from (2.46) that

$$\partial_t\left(\partial_s x\right) = \partial_t y = f_x\left(t, x\left(t,s\right), s\right)\partial_s x + f_s\left(t, x\left(t,s\right), s\right), \tag{2.51}$$

whence we conclude that

$$\partial_s\partial_t x = \partial_t\partial_s x.$$

Hence, the derivatives ∂_s and ∂_t commute[3] on x. If one knew this identity in advance then the derivation of (2.46) would be easy. Indeed, by differentiating in s the equation (2.49), we obtain (2.50). Interchanging then ∂_t and ∂_s, we obtain (2.46). Although this argument is not a proof of (2.46), it allows one to memorize the equation in (2.46).

For the next statement, introduce the following terminology. Let $f\left(u,v\right)$ be a function of two (vector) variables u, v, defined in an open set Ω. We write $f \in C^k\left(u\right)$ if all the partial derivatives of f up to the order k with respect to all components u_i exist and are continuous in Ω. That is, each partial derivative

$$\partial_u^\alpha f = \partial_{u_1}^{\alpha_1}\partial_{u_2}^{\alpha_2}...f$$

exists and is continuous in $\left(u,v\right) \in \Omega$ provided $|\alpha| = \alpha_1 + \alpha_2 + ... \leq k$. Previously we have used the notation $f \in C^k$ to say that f has the partial derivatives up to the order k with respect to *all* variables, in this case, u_i and v_j.

Theorem 2.17 *Under the conditions of* Theorem 2.16, *assume that, for some* $k \in \mathbb{N}$, $f\left(t,x,s\right) \in C^k\left(x,s\right)$. *Then the maximal solution* $x\left(t,s\right)$ *belongs to* $C^k\left(s\right)$. *Moreover, for any multiindex* α *of the order* $|\alpha| \leq k$ *and of the dimension* m *(the same as that of* s*), we have*

$$\partial_t\partial_s^\alpha x = \partial_s^\alpha\partial_t x. \tag{2.52}$$

Proof. Induction in k. If $k = 1$ then the fact that $x \in C^1\left(s\right)$ is the claim of Theorem 2.16, and the equation (2.52) with $|\alpha| = 1$ was also proved above. Let us make inductive step from $k-1$ to k, for $k \geq 2$. Assume $f \in C^k\left(x,s\right)$. Since also $f \in C^{k-1}\left(x,s\right)$, by the inductive hypothesis we have $x \in C^{k-1}\left(s\right)$. Set $y = \partial_s x$ and recall that by Theorem 2.16

$$\begin{cases} y' = f_x\left(t,x,s\right)y + f_s\left(t,x,s\right), \\ y\left(t_0\right) = 0, \end{cases} \tag{2.53}$$

where $x = x\left(t,s\right)$. Since f_x and f_s belong to $C^{k-1}\left(x,s\right)$ and $x\left(t,s\right) \in C^{k-1}\left(s\right)$, we obtain that the composite functions $f_x\left(t, x\left(t,s\right), s\right)$ and $f_s\left(t, x\left(t,s\right), s\right)$ are of the class $C^{k-1}\left(s\right)$. Hence, the right hand side in (2.53) is of the class $C^{k-1}\left(y,s\right)$ and, by the inductive hypothesis, we conclude that $y \in C^{k-1}\left(s\right)$. It follows that $x \in C^k\left(s\right)$.

[3] The equality of the mixed derivatives can be concluded by a theorem from Analysis II if one knows that both $\partial_s\partial_t x$ and $\partial_t\partial_s x$ are continuous. Their continuity follows from the identities (2.50) and (2.51), which prove at the same time also their equality.

To prove (2.52), choose some index i so that $\alpha_i \geq 1$. Set $\beta = \alpha - (0, ...1, ..0)$ where the only 1 is at the position i. Since by the first part of the proof $f(t, x(t,s), s) \in C^k(s)$, we obtain, using $\partial_s^\alpha = \partial_s^\beta \partial_{s_i}$, the chain rule, and the equation (2.46) for the column y_i,

$$\begin{aligned}\partial_s^\alpha \partial_t x &= \partial_s^\alpha f(t,x,s) = \partial_s^\beta \partial_{s_i} f(t,x,s) = \partial_s^\beta \left(f_{x_i}(t,x,s)\,\partial_{s_i} x + f_{s_i}\right) \\ &= \partial_s^\beta \left(f_{x_i}(t,x,s)\, y_i + f_{s_i}\right) = \partial_s^\beta \partial_t y_i.\end{aligned}$$

Since $|\beta| = k-1$, we can apply the inductive hypothesis to the IVP (2.53) and conclude that

$$\partial_s^\beta \partial_t y_i = \partial_t \partial_s^\beta y_i,$$

whence it follows that

$$\partial_s^\alpha \partial_t x = \partial_t \partial_s^\beta y_i = \partial_t \partial_s^\beta \partial_{s_i} x = \partial_t \partial_s^\alpha x.$$

■

How can one find the higher derivatives of $x(t,s)$ in s? Let us show how to find the ODE for the second derivative $z = \partial_{ss} x$ assuming for simplicity that $n = m = 1$, that is, both x and s are one-dimensional. For the derivative $y = \partial_s x$ we have the IVP (2.53), which we write in the form

$$\begin{cases} y' = g(t,y,s) \\ y(t_0) = 0 \end{cases} \tag{2.54}$$

where

$$g(t,y,s) = f_x(t,x(t,s),s)\,y + f_s(t,x(t,s),s).$$

If $f \in C^2(x,s)$ then $x(t,s) \in C^2(s)$ which implies that $g \in C^1(y,s)$. Noticing that $z = \partial_s y$ and applying the variational equation for the problem (2.54), we obtain the equation for z

$$z' = g_y(t,y(t,s),s)\,z + g_s(t,y(t,s),s)$$

(alternatively, differentiating in s the equation $y' = g(t,y,s)$ and interchanging the derivatives ∂_t and ∂_s, we obtain the same equation). Since $g_y = f_x(t,x,s)$ and

$$g_s(t,y,s) = f_x(t,x,s)\,z + f_{xx}(t,x,s)(\partial_s x)\,y + f_{xs}(t,x,s)\,y + f_{sx}(t,x,s)\,\partial_s x + f_{ss}(t,x,s)$$

and $\partial_s x = y$, we conclude that

$$\begin{cases} z' = f_x(t,x,s)\,z + f_{xx}(t,x,s)\,y^2 + 2f_{xs}(t,x,s)\,y + f_{ss}(t,x,s) \\ z'(t_0) = 0. \end{cases} \tag{2.55}$$

Note that here $x(t,s)$ must be substituted for x and $y(t,s)$ – for y. The equation (2.55) is called the variational equation of the second order, or the second variational equation. Note that it is a linear equation and it has the same coefficient $f_x(t,x(t,s),s)$ in front of the unknown function as the first variational equation. Similarly one finds the variational equations of the higher orders.

Example. This is a continuation of the previous example of the IVP with parameter

$$\begin{cases} x' = x^2 + 2s/t \\ x(1) = -1 \end{cases}$$

where we have computed that

$$x(t,0)=-\frac{1}{t}\ \text{ and }\ y(t):=\partial_s x(t,0)=1-\frac{1}{t^2}.$$

Obviously, the function $f(t,x,s)=x^2+2s/t$ belongs to $C^\infty(x,s)$ whence it follows by Theorem 2.17 that $x(t,s)\in C^\infty(s)$. Let us compute $z=\partial_{ss}x(t,0)$. Since

$$f_{xx}=2,\ f_{xs}=0,\ f_{ss}=0,$$

we obtain the second variational equation

$$z'=-\frac{2}{t}z+\left(f_{xx}|_{x=-\frac{1}{t},s=0}\right)y^2=-\frac{2}{t}z+2\left(1-t^{-2}\right)^2.$$

Solving similarly to the first variational equation with the same $a(t)=-\frac{2}{t}$ and with $b(t)=2\left(1-t^{-2}\right)^2$, we obtain

$$\begin{aligned} z(t) &= e^{A(t)}\int e^{-A(t)}b(t)\,dt=t^{-2}\int 2t^2\left(1-t^{-2}\right)^2dt \\ &= t^{-2}\left(\frac{2}{3}t^3-\frac{2}{t}-4t+C\right)=\frac{2}{3}t-\frac{2}{t^3}-\frac{4}{t}+\frac{C}{t^2}. \end{aligned}$$

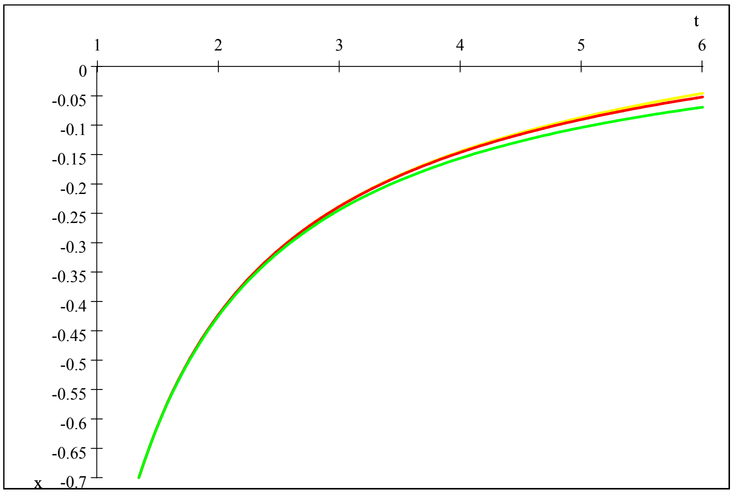
The initial condition $z(1)=0$ yields $C=\frac{16}{3}$ whence

$$z(t)=\frac{2}{3}t-\frac{2}{t^3}-\frac{4}{t}+\frac{16}{3t^2}.$$

Expanding $x(t,s)$ at $s=0$ by the Taylor formula of the second order, we obtain as $s\to 0$

$$\begin{aligned} x(t,s) &= x(t)+y(t)s+\frac{1}{2}z(t)s^2+o\left(s^2\right) \\ &= -\frac{1}{t}+\left(1-t^{-2}\right)s+\left(\frac{1}{3}t-\frac{2}{t}+\frac{8}{3t^2}-\frac{1}{t^3}\right)s^2+o\left(s^2\right). \end{aligned}$$

For comparison, the plots below show for $s=0.1$ the solution $x(t,s)$ (yellow) found by numerical methods (MAPLE), the first order approximation $u(t)=-\frac{1}{t}+\left(1-t^{-2}\right)s$ (green) and the second order approximation $v(t)=-\frac{1}{t}+\left(1-t^{-2}\right)s+\left(\frac{1}{3}t-\frac{2}{t}+\frac{8}{3t^2}-\frac{1}{t^3}\right)s^2$ (red).



Let us discuss an alternative method of obtaining the equations for the derivatives of $x(t,s)$ in s. As above, let $x(t)$, $y(t)$, $z(t)$ be respectively $x(t,0)$, $\partial_s x(t,0)$ and $\partial_{ss} x(t,0)$ so that by the Taylor formula

$$x(t,s) = x(t) + y(t)s + \frac{1}{2}z(t)s^2 + o\left(s^2\right). \tag{2.56}$$

Let us write a similar expansion for $x' = \partial_t x$. Since by Theorem 2.17 the derivatives ∂_t and ∂_s commute on x, we have

$$\partial_s x' = \partial_t \partial_s x = y'$$

and in the same way $\partial_{ss} x' = z'$. Hence,

$$x'(t,s) = x'(t) + y'(t)s + \frac{1}{2}z'(t)s^2 + o\left(s^2\right).$$

Substituting this into the equation

$$x' = x^2 + 2s/t$$

we obtain

$$x'(t) + y'(t)s + \frac{1}{2}z'(t)s^2 + o\left(s^2\right) = \left(x(t) + y(t)s + \frac{1}{2}z(t)s^2 + o\left(s^2\right)\right)^2 + 2s/t$$

whence

$$x'(t) + y'(t)s + \frac{1}{2}z'(t)s^2 = x^2(t) + 2x(t)y(t)s + \left(y(t)^2 + x(t)z(t)\right)s^2 + 2s/t + o\left(s^2\right).$$

Equating the terms with the same powers of s (which can be done by the uniqueness of the Taylor expansion), we obtain the equations

$$\begin{aligned} x'(t) &= x^2(t) \\ y'(t) &= 2x(t)y(t) + 2s/t \\ z'(t) &= 2x(t)z(t) + 2y^2(t). \end{aligned}$$

From the initial condition $x(1,s) = -1$ we obtain

$$-1 = x(1) + sy(1) + \frac{s^2}{2}z(1) + o\left(s^2\right),$$

whence $x(t) = -1$, $y(1) = z(1) = 0$. Solving successively the above equations with these initial conditions, we obtain the same result as above.

2.10 Differentiability of solutions in the initial conditions

Theorems 2.16 and 2.17 can be applied to the case when the parameter enters the initial condition, say, for the IVP

$$\begin{cases} x' = f(t,x), \\ x(t_0) = s, \end{cases} \tag{2.57}$$

where x and s are n-dimensional. As we already know, the change $\widetilde{x} = x - s$ reduces this problem to

$$\begin{cases} \widetilde{x}' = f(t, \widetilde{x} + s) \\ \widetilde{x}(t_0) = 0. \end{cases}$$

Hence, if $f \in C^k(x)$ then $f(t, \widetilde{x} + s) \in C^k(\widetilde{x}, s)$ and by Theorem 2.17 we obtain $\widetilde{x} \in C^k(s)$ and, hence, $x \in C^k(s)$. It follows from $x' = f(t, x)$ that also $x' \in C^k(s)$.

To conclude this Chapter, let us emphasize that the main results are the existence, uniqueness, continuity and differentiability in parameters for the systems of ODEs of the first order. Recall that a higher order scalar ODE $x^{(n)} = f\left(t, x, x', ..., x^{(n-1)}\right)$ can be reduced to a system of the first order. Hence, all the results for the systems can be transferred to the higher order ODE.

3 Linear equations and systems

A linear (system of) ODE of the first order is a (vector) ODE of the form

$$x' = A(t)\,x + B(t)$$

where $A(t) : I \to \mathbb{R}^{n \times n}$ and $B : I \to \mathbb{R}^n$ and I being an open interval in $\mathbb{R}$. If $A(t)$ and $B(t)$ are continuous in t then by Theorem 2.13 the IVP

$$\begin{cases} x' = A(t)\,x + B(t) \\ x(t_0) = v \end{cases} \tag{3.1}$$

has, for any $t_0 \in I$ and $v \in \mathbb{R}^n$, a unique solution defined on the full interval I. In the sequel, we always assume that $A(t)$ and $B(t)$ are continuous on I and consider only solutions defined on the entire interval I. Denote by $x(t, v)$ the solution to the IVP (3.1), where t_0 will be considered as fixed, while v may vary. When v varies in $\mathbb{R}^n$, $x(t, v)$ runs over all solutions to the ODE $x' = A(t)\,x + B(t)$ because any solution has some value at t_0. Hence, $x(t, v)$ with a parameter v is the general solution to the ODE.

The linear ODE is called homogeneous if $B(t) \equiv 0$, and inhomogeneous otherwise.

3.1 Space of solutions of a linear system

Denote by $\mathcal{A}$ the set of all solutions of the ODE $x' = A(t)\,x$ and by $\mathcal{B}$ - the set of all solutions of the ODE $x' = A(t)\,x + B(t)$.

Theorem 3.1 (*a*) *The set $\mathcal{A}$ is a linear space and $\mathcal{B} = \mathcal{A} + x_0$ for any $x_0 \in \mathcal{B}$*

(*b*) $\dim \mathcal{A} = n$. *Consequently, if $x_1(t), ..., x_n(t)$ is a sequence of n linearly independent solutions of $x' = Ax$ then the general solution of this equation is given by*

$$x(t) = C_1 x_1(t) + ... + C_n x_n(t) \tag{3.2}$$

where $C_1, ..., C_n$ are arbitrary constants. Furthermore, the general solution to the equation $x' = Ax + B$ is given by

$$x(t) = x_0(t) + C_1 x_1(t) + ... + C_n x_n(t) \tag{3.3}$$

where $x_0(t)$ is one of the solutions of this equation.

Proof of Theorem 3.1(a). All $\mathbb{R}^n$-valued functions on I form a linear space with respect to operations addition and multiplication by constant. Zero element is the function which is constant 0 on I. We need to prove that $\mathcal{A}$ is a linear subspace of the space of all functions. It suffices to show that $\mathcal{A}$ is closed under operations of addition and multiplication by constant.

If x and $y \in \mathcal{A}$ then also $x+y \in \mathcal{A}$ because

$$(x+y)' = x' + y' = Ax + Ax = A(x+y)$$

and similarly $\lambda x \in \mathcal{A}$ for any $\lambda \in \mathbb{R}$. Hence, $\mathcal{A}$ is a linear space.

Let $x \in \mathcal{A}$. Then

$$(x_0 + x)' = Ax_0 + B + Ax = A(x_0 + x) + B$$

so that $x_0 + x \in \mathcal{B}$. Conversely, any solution $y \in \mathcal{B}$ can be represented in the from $x_0 + x$ where $x \in \mathcal{A}$. Indeed, just set $x = y - x_0$ and observe that $x \in \mathcal{A}$ because

$$x' = y' - x_0' = (Ay + B) - (Ax_0 + B) = A(y - x_0) = Ax.$$

Hence, we have shown that $\mathcal{B} = \mathcal{A} + x_0$. ■

For part (b), we need first a lemma.

Lemma 3.2 *If $x(t,v)$ solves in the interval I the IVP*

$$\begin{cases} x' = Ax \\ x(t_0) = v \end{cases}$$

then, for any $t \in I$, the mapping $v \mapsto x(t,v)$ is a linear isomorphism of $\mathbb{R}^n$.

Proof. Fix $t \in I$ and show that $x(t,v)$ is a linear function of v. Indeed, the function

$$y(t) = x(t,u) + x(t,v)$$

is the solution as the sum of two solutions, and satisfies the initial condition

$$y(t_0) = x(t_0,u) + x(t_0,v) = u + v.$$

Hence, $y(t)$ solves the same IVP as $x(t,u+v)$ and, hence, by the uniqueness, $y(t) = x(t,u+v)$, that is,

$$x(t,u) + x(t,v) = x(t,u+v).$$

In the same way,

$$x(t,\lambda v) = \lambda x(t,v).$$

Hence, the mapping $v \mapsto x(t,v)$ is a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Let us show that the mapping $v \mapsto x(t,v)$ is injective, that is,

$$x(t,v) = 0 \implies v = 0.$$

Indeed, assume $v \neq 0$ but $x(t,v) = 0$ for some $t \in I$. Then the solutions $x \equiv 0$ and $x(\cdot,v)$ have the same value 0 at time t. Therefore, they solve the same IVP with the initial condition at time t, which implies that they must be equal. In particular, this implies $v = x(t_0, v) = 0$.

The mapping $v \mapsto x(t, v)$ is surjective by a general property of linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^n$ that the injectivity implies surjectivity. Another way to see it is as follows. For any $u \in \mathbb{R}^n$, we can find a solution that takes the value u at time t and define v as the value of this solution at t_0. Then $u = x(t, v)$.

Hence, the mapping $v \mapsto x(t, v)$ is a linear bijection of $\mathbb{R}^n$ onto $\mathbb{R}^n$, that is, an isomorphism. ■

Proof of Theorem 3.1(*b*). Consider the mapping $\Phi : \mathbb{R}^n \to \mathcal{A}$ such that for any $v \in \mathbb{R}^n$, $\Phi(v)$ is the solution $x(t, v)$ (unlike the statement of Lemma 3.2, here we do not fix t so that $\Phi(v)$ is a function of t rather than a vector in $\mathbb{R}^n$). It follows from Lemma 3.2 that Φ is a linear mapping. Since any function from $\mathcal{A}$ has the form $x(t, v)$ for some v, the mapping Φ is surjective. Mapping Φ is injective because $x(t, v) \equiv 0$ implies $v = x(t_0, v) = 0$. Hence, Φ is a linear isomorphism of $\mathbb{R}^n$ and $\mathcal{A}$, whence $\dim \mathcal{A} = \dim \mathbb{R}^n = n$.

Consequently, if $x_1, ..., x_n$ are linearly independent functions from $\mathcal{A}$ then they form a basis in $\mathcal{A}$ because $n = \dim \mathcal{A}$. It follows that any element of $\mathcal{A}$ is a linear combination of $x_1, ..., x_n$, that is, any solution to $x' = Ax$ has the form (3.2). The fact that any solution to $x' = Ax + B$ has the form (3.3) follows from $\mathcal{B} = \mathcal{A} + x_0$. ■

Theorem 3.1 suggests that in order to find a general solution of the system $x' = Ax$, it suffices to find n linearly independent solutions. There are various methods for that, which will be discussed later in this Chapter. How to verify that the functions $x_1, ..., x_n$ are linearly independent? Note that the zero of the linear space $\mathcal{A}$ is the function which is identical zero on I. Therefore, functions $x_1, ..., x_n$ are linearly independent if

$$\lambda_1 x_1(t) + ... + \lambda_n x_n(t) \equiv 0 \implies \lambda_1 = ... = \lambda_n = 0$$

where $\lambda_1, ..., \lambda_n$ are real coefficients. The next statement gives a convenient criterion for linear independence.

Definition. Given a sequence of n vector functions $x_1, ..., x_n : I \to \mathbb{R}^n$, define their *Wronskian* $W(t)$ as a real valued function on I by

$$W(t) = \det\left(x_1(t) \mid x_2(t) \mid ... \mid x_n(t)\right),$$

where the matrix on the right hand side is formed by the column-vectors $x_1, ..., x_n$. Hence, $W(t)$ is the determinant of the $n \times n$ matrix.

Lemma 3.3 *Let $x_1, ..., x_n$ be the sequence of n solutions of $x' = Ax$ (that is, $x_i \in \mathcal{A}$ for all $i = 1, ..., n$). Then either $W(t) \equiv 0$ for all $t \in I$ and the functions $x_1, ..., x_n$ are linearly dependent or $W(t) \neq 0$ for all $t \in I$ and the functions $x_1, ..., x_n$ are linearly independent.*

Proof. For any fixed $t \in I$, the sequence $x_1(t), ..., x_n(t)$ is a sequence of vectors from $\mathbb{R}^n$. By Linear Algebra, this sequence is linearly dependent if and only if $W(t) = 0$.

If $W(t) = 0$ for some $t = t_0$ then sequence $x_1(t_0), ..., x_n(t_0)$ is linearly dependent so that

$$\lambda_1 x_1(t_0) + ... + \lambda_n x_n(t_0) = 0$$

for some constants $\lambda_1, ..., \lambda_n$ that are not all equal to 0. Then the function

$$y(t) = \lambda_1 x(t) + ... + \lambda_n x_n(t)$$

solves the ODE $y' = Ay$ and satisfies the condition $y(t_0) = 0$. Hence, $y(t) = 0$ for all $t \in I$, that is,

$$\lambda_1 x_1(t) + ... + \lambda_n x_n(t) = 0 \text{ for all } t \in I. \tag{3.4}$$

Therefore, the sequence of functions $x_1, ..., x_n$ is linearly dependent. The identity (3.4) obviously implies that $W(t) = 0$ for all t.

Hence, we have proved that if $W(t) = 0$ for some t then $W(t) = 0$ for all t and the functions $x_1, ..., x_n$ are linearly dependent. Obviously, if $W(t) \neq 0$ for all t then the functions $x_1, ..., x_n$ are linearly independent. Hence, we see that only two alternatives from the statement of Lemma 3.3 can occur. ■

Example. Consider two vector functions

$$x_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} \sin t \\ \cos t \end{pmatrix}.$$

The Wronskian of this sequence is

$$W(t) = \begin{pmatrix} \cos t & \sin t \\ \sin t & \cos t \end{pmatrix} = \cos^2 t - \sin^2 t = \cos 2t.$$

Clearly, $W(t)$ can vanish at some points (say at $t = \pi/4$) while $W(t) \neq 0$ at other points. This means that these two vector functions cannot be solutions of the same system of ODEs.

For comparison, the functions

$$x_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

have the Wronskian $W(t) \equiv 1$, and they both are solutions of the same system

$$x' = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x.$$

3.2 Space of solutions of a linear ODE of the order n

Consider an ODE of the order n

$$x^{(n)} = f\left(t, x, x', ..., x^{(n-1)}\right), \tag{3.5}$$

where $x(t)$ is a real-valued function and f is a real valued function in an open subset $\Omega \subset \mathbb{R}^{n+1}$. The initial conditions are

$$x(t_0) = v_0,\ \ x'(t_0) = v_1, ...,\ \ x^{(n-1)}(t_0) = v_{n-1} \tag{3.6}$$

where $(t_{0,}v_0, ..., v_{n-1}) \in \Omega$. Considering the vector function

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ ... \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} x(t) \\ x'(t) \\ ... \\ x^{(n-1)}(t) \end{pmatrix}, \tag{3.7}$$

rewrite the ODE (3.5) in the vector form

$$\mathbf{x}' = F(t, \mathbf{x})$$

where

$$F(t, \mathbf{x}) = \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \\ ... \\ f(t, \mathbf{x}_1, ..., \mathbf{x}_n) \end{pmatrix}.$$

The initial condition becomes $\mathbf{x}(t_0) = v = (v_0, ..., v_{n-1})$. The system $\mathbf{x}' = F(t, \mathbf{x})$ is called the *normal system* of the ODE (3.5).

Assuming that the function $f(t, \mathbf{x})$ is continuous and locally Lipschitz in $\mathbf{x}$, we obtain that the same is true for $F(t, \mathbf{x})$ so that we obtain the existence and uniqueness for the IVP

$$\begin{cases} \mathbf{x}' = F(t, \mathbf{x}) \\ \mathbf{x}(t_0) = v. \end{cases}$$

It follows that also the IVP (3.5)-(3.6) has a unique maximal solution $x(t) = \mathbf{x}_1(t)$ for any set of the initial conditions.

Consider now a higher order *linear* ODE

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = b(t) \tag{3.8}$$

where $a_i(t)$ and $b(t)$ are real-valued continuous functions on an interval I and $x(t)$ is the unknown real-valued function. Alongside (3.8), consider also the corresponding homogeneous ODE

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = 0 \tag{3.9}$$

In the both cases, the initial conditions are

$$x(t_0) = v_0,\ \ x'(t_0) = v_1, ...,\ \ x^{(n-1)}(t_0) = v_{n-1}$$

where $t_0 \in \mathbb{R}$ and $v = (v_0, ..., v_{n-1}) \in \mathbb{R}^n$.

Theorem 3.4 *(a) For any $t_0 \in I$ and $v \in \mathbb{R}^n$, the IVP for* (3.8) *has a unique solution defined on I.*

(b) Let $\mathcal{A}$ be the set of all solutions to (3.9) *and $\mathcal{B}$ be the set of all solutions to* (3.8)*. Then $\mathcal{A}$ is a linear space and $\mathcal{B} = \mathcal{A} + x_0$ where x_0 is any solution to* (3.8)*.*

(c) $\dim \mathcal{A} = n$. Consequently, if $x_1, ..., x_n$ are n linearly independent solutions of (3.9) *then the general solution of* (3.9) *has the form*

$$x = C_1 x_1 + ... + C_n x_n$$

where $C_1, ..., C_n$ are arbitrary real constants, and the general solution of (3.8) *has the form*

$$x = x_0 + C_1 x_1 + ... + C_n x_n,$$

where x_0 is one of the solutions of (3.8)*.*

Proof. (a) The linear equation (3.8) has the form $x^{(n)} = f\left(t, x', ..., x^{(n-1)}\right)$ with the function

$$f\left(t, x, .., x^{(n-1)}\right) = -a_1 x^{(n-1)} - ... - a_n x + b.$$

Hence, the function $F(t, \mathbf{x})$ for the normal system is

$$F(t,\mathbf{x}) = \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \\ ... \\ -a_n \mathbf{x}_1 - ... - a_1 \mathbf{x}_n + b \end{pmatrix} = A\mathbf{x} + B$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ 0 \\ ... \\ b \end{pmatrix}.$$

Hence, the initial value problem for the ODE (3.8) amounts to

$$\begin{cases} \mathbf{x}' = A\mathbf{x} + B \\ \mathbf{x}(t_0) = v \end{cases}$$

which is linear. By Theorem 2.13, it has a unique solution defined on the entire interval I. Therefore, the IVP for (3.8) has also a unique solution $x(t) = \mathbf{x}_1(t)$ defined on I.

(b) The facts that $\mathcal{A}$ is a linear space and $\mathcal{B} = \mathcal{A} + x_0$ are trivial and proved in the same way as Theorem 3.1.

(c) Let $\widehat{\mathcal{A}}$ be the space of all solutions to the normal system $\mathbf{x}' = A\mathbf{x}$ where A is as above. Then we have a bijection between $\widehat{\mathcal{A}}$ and $\mathcal{A}$ given by (3.7). Obviously, this bijection is linear, which implies that $\mathcal{A}$ and $\widehat{\mathcal{A}}$ are isomorphic as linear spaces, whence $\dim \mathcal{A} = \dim \widehat{\mathcal{A}} = n$. ■

Let $x_1, ..., x_n$ are n real-valued functions on an interval I of the class C^{n-1}. Then their Wronskian is defined by

$$W(t) = \det \begin{pmatrix} x_1 & x_2 & ... & x_n \\ x_1' & x_2' & ... & x_n' \\ ... & ... & ... & ... \\ x_1^{(n-1)} & x_2^{(n-1)} & ... & x_n^{(n-1)} \end{pmatrix}.$$

Lemma 3.5 *Let $x_1, ..., x_n$ be n functions from $\mathcal{A}$. Then either $W(t) \equiv 0$ for all $t \in I$ and the functions $x_1, ..., x_n$ are linearly dependent or $W(t) \neq 0$ for all $t \in I$ and the functions $x_1, ..., x_n$ are linearly independent.*

Proof. Define the vector function

$$\mathbf{x}_k = \begin{pmatrix} x_k \\ x_k' \\ ... \\ x_k^{(n-1)} \end{pmatrix}$$

so that $\mathbf{x}_1, ..., \mathbf{x}_k$ is the sequence of vector functions that solve the vector ODE $\mathbf{x}' = A\mathbf{x}$. The Wronskian of the sequence $\mathbf{x}_1, ..., \mathbf{x}_n$ is obviously the same as the Wronskian of $x_1, ..., x_n$, and the sequence $\mathbf{x}_1, ..., \mathbf{x}_n$ is linearly independent if and only so is $x_1, ..., x_n$. Hence, the rest follows from Lemma 3.3. ■

Example. Consider the ODE $x'' + x = 0$. Two obvious solutions are $x_1(t) = \cos t$ and $x_2(t) = \sin t$. Their Wronskian is

$$W(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} = 1.$$

Hence, we conclude that these solutions are linearly independent and, hence, the general solution is $x(t) = C_1 \cos t + C_2 \sin t$. This can be used to solve the IVP

$$\begin{cases} x'' + x = 0 \\ x(t_0) = v_0, \ x'(t_0) = v_1. \end{cases}$$

Indeed, the coefficients C_1 and C_2 can be determined from the initial conditions.

Of course, in order to use this method, one needs to find enough number of independent solutions. This can be done for certain classes of linear ODEs, for example, for linear ODEs with constant coefficients.

3.3 Linear homogeneous ODEs with constant coefficients

Consider the methods of finding n independent solutions to the ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = 0, \tag{3.10}$$

where $a_1, ..., a_n$ are constants.

It will be convenient to obtain the complex valued general solution $x(t)$ and then to extract the real valued general solution. The idea is very simple. Let us look for a solution in the form $x(t) = e^{\lambda t}$ where λ is a complex number to be determined. Substituting this function into (3.10) and noticing that $x^{(k)} = \lambda^k e^{\lambda t}$, we obtain the equation for λ (after cancellation by $e^{\lambda t}$):

$$\lambda^n + a_1 \lambda^{n-1} + ... + a_n = 0.$$

This equation is called the *characteristic equation* of (3.10) and the polynomial $P(\lambda) = \lambda^n + a_1 \lambda^{n-1} + ... + a_n$ is called the characteristic polynomial of (3.10). Hence, if λ is the root of the characteristic polynomial then $e^{\lambda t}$ solves (3.10). We try to obtain in this way n independent solutions.

Claim 1. *If $\lambda_1, ..., \lambda_n$ are distinct complex numbers then the functions $e^{\lambda_1 t}, ..., e^{\lambda_n t}$ are linearly independent.*

Proof. Let us prove this by induction in n. If $n = 1$ then the claim is trivial, just because the exponential function is not identical zero. Inductive step from $n - 1$ to n. Assume that for some complex constants $C_1, ..., C_n$ and all $t \in \mathbb{R}$,

$$C_1 e^{\lambda_1 t} + ... + C_n e^{\lambda_n t} = 0 \tag{3.11}$$

and prove that $C_1 = ... = C_n = 0$. Dividing (3.11) by $e^{\lambda_n t}$ and setting $\mu_i = \lambda_i - \lambda_n$, we obtain

$$C_1 e^{\mu_1 t} + ... + C_{n-1} e^{\mu_{n-1} t} + C_n = 0.$$

Differentiating in t, we obtain

$$C_1 \mu_1 e^{\mu_1 t} + ... + C_{n-1} \mu_{n-1} e^{\mu_{n-1} t} = 0.$$

By the inductive hypothesis, we conclude that $C_i \mu_i = 0$ when by $\mu_i \neq 0$ we conclude $C_i = 0$, for all $i = 1, ..., n - 1$. Substituting into (3.11), we obtain also $C_n = 0$. ■

Hence, we obtain the following result.

Theorem 3.6 *If the characteristic polynomial $P(\lambda)$ of* (3.10) *has n distinct complex roots $\lambda_1, ..., \lambda_n$ then the general complex solution to* (3.10) *is given by*

$$x(t) = C_1 e^{\lambda_1 t} + ... + C_n e^{\lambda_n t}.$$

Proof. Indeed, each function $e^{\lambda_i t}$ is a solution, the sequence $\left\{e^{\lambda_i t}\right\}_{i=1}^n$ is linearly independent by Claim 1, and by Theorem 3.4 the general solution is as claimed. ■

Example. Consider the ODE

$$x'' - 3x' + 2x = 0.$$

The characteristic polynomial is $P(\lambda) = \lambda^2 - 3\lambda + 2$, which has the roots $\lambda_1 = 2$ and $\lambda_2 = 1$. Hence, the linearly independent solutions are e^{2t} and e^t, and the general solution is $C_1 e^{2t} + C_2 e^t$.

Consider now the ODE $x'' + x = 0$. The characteristic polynomial is $P(\lambda) = \lambda^2 + 1$, which has the complex roots $\lambda_1 = i$ and $\lambda_2 = -i$. Hence, we obtain the complex solutions e^{it} and e^{-it}. Out of them, we can get also real linearly independent solutions. Indeed, just replace these two functions by their two linear combinations (which corresponds to a change of the basis in the space of solutions)

$$\frac{e^{it} + e^{-it}}{2} = \cos t \text{ and } \frac{e^{it} - e^{-it}}{2i} = \sin t.$$

Hence, we conclude that $\cos t$ and $\sin t$ are linearly independent solutions and the general solution is $C_1 \cos t + C_2 \sin t$.

If $\{v_k\}$ is a sequence of vectors in a linear space then by span $\{v_k\}$ we denote the set of all linear combinations (with complex coefficients) of these vectors. Clearly, span $\{v_k\}$ is a linear subspace. The argument in the above example can be stated as follows

$$\operatorname{span}\left\{e^{it}, e^{-it}\right\} = \operatorname{span}\{\cos t, \sin t\}.$$

Claim 2. *Let a polynomial $P(\lambda)$ with real coefficients have a complex root $\lambda = \alpha + i\beta$, where $\beta \neq 0$. Then also $\overline{\lambda} = \alpha - i\beta$ is a root, and*

$$\operatorname{span}\left(e^{\lambda t}, e^{\overline{\lambda} t}\right) = \operatorname{span}\left(e^{\alpha t}\cos\beta t, e^{\alpha t}\sin\beta t\right).$$

Proof. Since the complex conjugations commutes with addition and multiplication of numbers, the identity $P(\lambda) = 0$ implies $P\left(\overline{\lambda}\right) = 0$ (since a_k are real, we have $\overline{a}_k = a_k$).

Next, we have

$$e^{\lambda t} = e^{\alpha t}(\cos\beta t + i\sin\beta t) \quad \text{and} \quad e^{\overline{\lambda} t} = e^{\alpha t}(\cos\beta t - \sin\beta t)$$

so that $e^{\lambda t}$ and $e^{\overline{\lambda} t}$ are linear combinations of $e^{\alpha t}\cos\beta t$ and $e^{\alpha t}\sin\beta t$. The converse is true also, because

$$e^{\alpha t}\cos\beta t = \frac{1}{2}\left(e^{\lambda t} + e^{\overline{\lambda} t}\right) \quad \text{and} \quad e^{\alpha t}\sin\beta t = \frac{1}{2i}\left(e^{\lambda t} - e^{\overline{\lambda} t}\right).$$

■

Using Theorem 3.6 and Claim 2, we can obtain a real basis in the space of solutions provided the characteristic polynomial has n distinct complex roots. Indeed, it suffices in the sequence $e^{\lambda_1 t}, ..., e^{\lambda_n t}$ to replace every couple $e^{\lambda t}$, $e^{\overline{\lambda} t}$ by functions $e^{\alpha t}\cos \beta t$ and $e^{\alpha t}\sin \beta t$.

Example. Consider an ODE $x''' - x = 0$. The characteristic polynomial is $P(\lambda) = \lambda^3 - 1 = (\lambda - 1)\left(\lambda^2 + \lambda + 1\right)$ that has the roots $\lambda_1 = 1$ and $\lambda_{2,3} = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$. Hence, we obtain the three linearly independent real solutions

$$e^t,\ e^{-\frac{1}{2}t}\cos\frac{\sqrt{3}}{2}t,\ e^{-\frac{1}{2}t}\sin\frac{\sqrt{3}}{2}t,$$

and the real general solution is

$$C_1 e^t + e^{-\frac{1}{2}t}\left(C_2\cos\frac{\sqrt{3}}{2}t + C_3\sin\frac{\sqrt{3}}{2}t\right).$$

What to do when $P(\lambda)$ has fewer than n distinct roots? Recall the fundamental theorem of algebra (which is normally proved in a course of Complex Analysis): any polynomial $P(\lambda)$ of degree n with complex coefficients has exactly n complex roots counted with multiplicity. What is it the multiplicity of a root? If λ_0 is a root of $P(\lambda)$ then its multiplicity is the maximal natural number m such that $P(\lambda)$ is divisible by $(\lambda - \lambda_0)^m$, that is, the following identity holds

$$P(\lambda) = (\lambda - \lambda_0)^m Q(\lambda),$$

where $Q(\lambda)$ is another polynomial of λ. Note that $P(\lambda)$ is always divisible by $\lambda - \lambda_0$ so that $m \geq 1$. The fundamental theorem of algebra can be stated as follows: if $\lambda_1, ..., \lambda_k$ are all distinct roots of $P(\lambda)$ and the multiplicity of λ_i is m_i then $m_1 + ... + m_k = n$ and, hence,

$$P(\lambda) = (\lambda - \lambda_1)^{m_1} ... (\lambda - \lambda_k)^{m_k}.$$

In order to obtain n independent solutions to the ODE (3.10), each root λ_i should give us m_i independent solutions.

Theorem 3.7 *Let $\lambda_1, ..., \lambda_k$ be the distinct roots of the characteristic polynomial $P(\lambda)$ with the multiplicities $m_1, ..., m_k$. Then the following n functions are linearly independent solutions of* (3.10)*:*

$$\left\{t^{j-1}e^{\lambda_i t}\right\},\ i = 1, ..., k,\ j = 1, ..., m_i.$$

Consequently, the general solution of (3.10) *is*

$$x(t) = \sum_{i=1}^{k}\sum_{j=1}^{m_i} C_{ij} t^{j-1} e^{\lambda_i t} \tag{3.12}$$

where C_{ij} are arbitrary complex constants.

Remark. Denoting

$$P_i(t) = \sum_{j=1}^{m_i} C_{ij} t^{j-1}$$

we obtain from (3.12)

$$x(t) = \sum_{i=1}^{k} P_i(t) e^{\lambda_i t}. \tag{3.13}$$

Hence, any solution to (3.10) has the form (3.13) where P_i is an arbitrary polynomial of t of the degree at most $m_i - 1$.

Example. Consider the ODE $x'' - 2x' + x = 0$ which has the characteristic polynomial

$$P(\lambda) = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2.$$

Obviously, $\lambda = 1$ is the root of multiplicity 2. Hence, by Theorem 3.7, the functions e^t and te^t are linearly independent solutions, and the general solution is

$$x(t) = (C_1 + C_2 t) e^t.$$

Consider the ODE $x^V + x^{IV} - 2x''' - 2x'' + x' + x = 0$. The characteristic polynomial is

$$P(\lambda) = \lambda^5 + \lambda^4 - 2\lambda^3 - 2\lambda^2 + \lambda + 1 = (\lambda - 1)^2 (\lambda + 1)^3.$$

Hence, the roots are $\lambda_1 = 1$ with $m_1 = 2$ and $\lambda_2 = -1$ with $m_2 = 3$. We conclude that the following 5 function are linearly independent solutions:

$$e^t, \ te^t, \ e^{-t}, \ te^{-t}, \ t^2 e^{-t}.$$

The general solution is

$$x(t) = (C_1 + C_2 t) e^t + \left(C_3 + C_4 t + C_5 t^2\right) e^{-t}.$$

Proof of Theorem 3.7. We first verify that the function $t^{j-1} e^{\lambda_i t}$ is indeed a solution. Given a polynomial $Q(\lambda) = b_0 \lambda^l + b_1 \lambda^{l-1} + ... + b_0$ with complex coefficients, we can associate with it the differential operator

$$\begin{aligned} Q\left(\frac{d}{dt}\right) &= b_0 \left(\frac{d}{dt}\right)^l + b_1 \left(\frac{d}{dt}\right)^{l-1} + ... + b_0 \\ &= b_0 \frac{d^l}{dt^l} + b_1 \frac{d^{l-1}}{dt^{l-1}} + ... + b_0, \end{aligned}$$

where we use the convention that the "product" of differential operators is the composition. That is, the operator $Q\left(\frac{d}{dt}\right)$ acts on a smooth enough function $f(t)$ by the rule

$$Q\left(\frac{d}{dt}\right) f = b_0 f^{(l)} + b_1 f^{(l-1)} + ... + b_0 f$$

(here the constant term b_0 is understood as a multiplication operator). It follows from the definition that if $Q(\lambda)$ and $R(\lambda)$ are two polynomials then

$$(QR)\left(\frac{d}{dt}\right) = Q\left(\frac{d}{dt}\right) R\left(\frac{d}{dt}\right)$$

(because the product of two differential operators of the above kind is computed using the same rules as for the product of the polynomials).

Since λ_i is a root of $P(\lambda)$ of multiplicity m_i, we obtain for some polynomial $Q(\lambda)$

$$P(\lambda) = (\lambda - \lambda_i)^{m_i} Q(\lambda) = Q(\lambda)(\lambda - \lambda_i)^{m_i}$$

whence

$$P\left(\frac{d}{dt}\right) = Q\left(\frac{d}{dt}\right)\left(\frac{d}{dt} - \lambda_i\right)^{m_i}.$$

We would like to show that the function $f(t) = t^j e^{\lambda_i t}$ is a solution for any $j \le m_i$, that is,

$$P\left(\frac{d}{dt}\right) f(t) = 0,$$

and for that it suffices to prove that

$$\left(\frac{d}{dt} - \lambda_i\right)^{m_i} f(t) = 0.$$

To simplify notation, let us drop the index i and state this fact in a bit more general way:

Claim 3. *For all $\lambda \in \mathbb{C}$ and $j, m \in \mathbb{N}$ such that $j \le m$,*

$$\left(\frac{d}{dt} - \lambda\right)^m \left(t^{j-1} e^{\lambda t}\right) = 0.$$

It suffices to prove this for $m = j$ because for larger values of m this will follow trivially. Hence, let us prove by induction in j that

$$\left(\frac{d}{dt} - \lambda\right)^j \left(t^{j-1} e^{\lambda t}\right) = 0.$$

(Note that if $\lambda = 0$ then this amounts to the trivial identity $\left(\frac{d}{dt}\right)^j t^{j-1} = 0$). Inductive bases for $j = 1$ is verified as follows:

$$\left(\frac{d}{dt} - \lambda\right) e^{\lambda t} = \left(e^{\lambda t}\right)' - \lambda e^{\lambda t} = 0.$$

The inductive step from j to $j + 1$. We have by the product rule

$$\begin{aligned}
\left(\frac{d}{dt} - \lambda\right)^{j+1} \left(t^j e^{\lambda t}\right) &= \left(\frac{d}{dt} - \lambda\right)^j \left(\frac{d}{dt} - \lambda\right) \left(t^j e^{\lambda t}\right) \\
&= \left(\frac{d}{dt} - \lambda\right)^j \left(\left(t^j e^{\lambda t}\right)' - \lambda t^j e^{\lambda t}\right) \\
&= \left(\frac{d}{dt} - \lambda\right)^j \left(j t^{j-1} e^{\lambda t} + \lambda t^j e^{\lambda t} - \lambda t^j e^{\lambda t}\right) \\
&= j\left(\frac{d}{dt} - \lambda\right)^j \left(t^{j-1} e^{\lambda t}\right) = 0,
\end{aligned}$$

where the last identity is true by the inductive hypothesis. This finishes the proof of Claim 3.

We are left to show that the sequence of functions $\left\{t^{j-1}e^{\lambda_i t}\right\}_{i,j}$ is linearly independent. Assume from the contrary that they are linearly dependent, that is, some linear combination of them is identically equal to 0. Combining together the terms with the same λ_i, we obtain that $\sum_{i=1}^k P_i(t)\, e^{\lambda_t t} \equiv 0$ where $P_i(t)$ are some polynomials of t (that are obtained by taking linear combinations of the terms t^{j-1}). We would like to be able to conclude that all polynomials $P_i(t)$ are identical 0, which will imply that all the coefficients of the linear combination are zeros.

Claim 4. *If $\lambda_1, ..., \lambda_k$ are distinct complex numbers and if, for some polynomials $P_i(t)$,*

$$\sum_{i=1}^k P_i(t)\, e^{\lambda_t t} \equiv 0 \tag{3.14}$$

then $P_t(t) \equiv 0$ for all i.

For any non-zero polynomial P, define $\deg P$ as the maximal power of t that enters P with non-zero coefficient. If $P(t) \equiv 0$ then set $\deg P = 0$. We prove the claim by induction in a parameter s assuming that

$$\sum_{i=1}^k \deg P_i \leq s.$$

Inductive basis for $s = 0$. In this case, all $\deg P_i$ must be zero, that is, each P_i is just a constant. Then the identity $\sum_i P_i e^{\lambda_i t} = 0$ implies $P_i = 0$ because the functions $e^{\lambda_i t}$ are linearly independent by Claim 1.

Inductive step from $s-1$ to s where $s \geq 1$. If all $\deg P_i = 0$ then we are again in the case of the inductive basis. Assume that among P_i there is a polynomial of a positive degree, say $\deg P_k > 0$. Differentiating (3.14) in t we obtain

$$\sum_{i=1}^k \left(P_i' + \lambda_i P_i\right) e^{\lambda_i t} = 0.$$

Subtracting (3.14) times c where c is a constant, we obtain

$$\sum_{i=1}^k Q_i(t)\, e^{\lambda_i t} = 0,$$

where

$$Q_i = P_i' + (\lambda_i - c)\, P_i.$$

Note that always $\deg Q_i \leq \deg P_i$. Choose now $c = \lambda_k$. Then $Q_k = P_k'$ whence

$$\deg Q_k = \deg P_k' < \deg P_k.$$

Hence, the sum of all the degrees of the polynomials Q_i is at most $s-1$. By the inductive hypothesis we conclude that $Q_i(t) \equiv 0$, that is, for any index i,

$$P_i' + (\lambda_i - c)\, P_i = 0.$$

Solving this ODE we obtain

$$P_i(t) = C \exp(-(\lambda_i - c)t).$$

If $i < k$ then $\lambda_i \neq c$, and the above identity of the polynomial and exponential functions only possible if $C \equiv 0$ (indeed, the exponential function has all higher order derivatives non-zero while a high enough derivative of a polynomial vanishes identically). Hence, $P_i(t) \equiv 0$ for all $i < k$. Substituting into (3.14) we obtain that also $P_k(t) e^{\lambda_k t} \equiv 0$ whence $P_k(t) \equiv 0$. ■

Finally, let us show how to extract the real general solution from the complex general solution. The following lemma is a generalization of Claim 2.

Lemma 3.8 *Let a polynomial $P(\lambda)$ with real coefficients have a complex root $\lambda = \alpha + i\beta$ (where $\beta \neq 0$) of multiplicity m. Then also $\overline{\lambda} = \alpha - i\beta$ is a root of multiplicity m and, for any $j \leq m$,*

$$\operatorname{span}\left(t^{j-1}e^{\lambda t}, t^{j-1}e^{\overline{\lambda} t}\right) = \operatorname{span}\left(t^{j-1}e^{\alpha t}\cos\beta t, t^{j-1}e^{\alpha t}\sin\beta t\right). \tag{3.15}$$

Hence, in the family of n independent solutions the sequence

$$e^{\lambda t}, te^{\lambda t}, ..., t^{m-1}e^{\lambda t}, e^{\overline{\lambda} t}, te^{\overline{\lambda} t}, ..., t^{m-1}e^{\overline{\lambda} t}$$

can be replaced by

$$e^{\alpha t}\cos\beta t,\ te^{\alpha t}\cos\beta t, ...,\ t^{m-1}e^{\alpha t}\cos\beta t,\ e^{\alpha t}\sin\beta t,\ te^{\alpha t}\sin\beta t, ...,\ t^{m-1}e^{\alpha t}\sin\beta t.$$

Proof. If λ_0 is a root of multiplicity m, then we have the identity

$$P(\lambda) = (\lambda - \lambda_0)^m Q(\lambda)$$

for some polynomial Q. Applying the complex conjugation and using the fact that the coefficients of P are real, we obtain

$$P\left(\overline{\lambda}\right) = \left(\overline{\lambda} - \overline{\lambda}_0\right)^m \overline{Q}\left(\overline{\lambda}\right)$$

where $\overline{Q}$ is the polynomial whose coefficients are complex conjugate to those of Q. Replacing $\overline{\lambda}$ by λ, we obtain

$$P(\lambda) = \left(\lambda - \overline{\lambda}_0\right)^m \overline{Q}(\lambda).$$

Hence, $\overline{\lambda}_0$ is also a root of multiplicity $m_1 \geq m$. Applying the complex conjugation to $\overline{\lambda}_0$ we obtain as above that $m \geq m_1$, whence $m = m_1$.

The identity (3.15) is an immediate consequence of Claim 2: for example, knowing that $e^{\alpha t}\cos\beta t$ is a linear combination of $e^{\lambda t}$ and $e^{\overline{\lambda} t}$, we conclude that $t^{j-1}e^{\alpha t}\cos\beta t$ is the linear combination of $t^{j-1}e^{\lambda t}$ and $t^{j-1}e^{\overline{\lambda} t}$. ■

Example. Consider the ODE $x^V + 2x''' + x' = 0$. Its characteristic polynomial is

$$P(\lambda) = \lambda^5 + 2\lambda^3 + \lambda = \lambda\left(\lambda^2+1\right)^2 = \lambda(\lambda+i)^2(\lambda-i)^2,$$

and it has the roots $\lambda_1 = 0$, $\lambda_2 = i$ and $\lambda_3 = -i$, where λ_2 and λ_3 has multiplicity 2. The general complex solution is then

$$C_1 + (C_2 + C_3 t)\,e^{it} + (C_4 + C_5 t)\,e^{-it},$$

and the general real solution is

$$C_1 + (C_2 + C_3 t)\cos t + (C_4 + C_5 t)\sin t.$$

3.4 Linear inhomogeneous ODEs with constant coefficients

Here we consider the equation

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = f(t) \tag{3.16}$$

where the function $f(t)$ is a *quasi-polynomial*, that is, f has the form

$$f(t) = \sum_i R_i(t)\,e^{\mu_i t}$$

where $R_i(t)$ are polynomials, μ_i are complex numbers and the sum is finite. It is obvious that the sum and the product of two quasi-polynomials is again a quasi-polynomial.

In particular, the following functions are quasi-polynomials

$$t^k e^{\alpha t}\cos\beta t \quad \text{and} \quad t^k e^{\alpha t}\sin\beta t$$

(where k is a non-negative integer and $\alpha, \beta \in \mathbb{R}$) because

$$\cos\beta t = \frac{e^{i\beta t} + e^{-i\beta t}}{2} \quad \text{and} \quad \sin\beta t = \frac{e^{i\beta t} - e^{-i\beta t}}{2i}.$$

As we know, the general solution of the inhomogeneous equation (3.16) is obtained as a sum of the general solution of the homogeneous equation and a particular solution of (3.16). Hence, we focus on finding a particular solution of (3.16).

As before, denote by $P(\lambda)$ the characteristic polynomial of (3.16), that is

$$P(\lambda) = \lambda^n + a_1\lambda^{n-1} + ... + a_n.$$

Then the equation (3.16) can be written in the short form $P\left(\frac{d}{dt}\right)x = f$, which will be used below.

Claim 1. *If $f = c_1 f_1 + ... + c_k f_k$ and $x_1(t), ..., x_k(t)$ are solutions to the equation $P\left(\frac{d}{dt}\right)x_i = f_i$, then $x = c_1x_1 + ... + c_kx_k$ solves the equation $P\left(\frac{d}{dt}\right)x = f$.*

Proof. This is trivial because

$$P\left(\frac{d}{dt}\right)x=P\left(\frac{d}{dt}\right)\sum_i c_ix_i=\sum_i c_iP\left(\frac{d}{dt}\right)x_i=\sum_i c_if_i=f.$$

■

Hence, we can assume that the function f in (3.16) is of the form $f(t)=R(t)e^{\mu t}$ where $R(t)$ is a polynomial.

To illustrate the method, which will be used in this Section, consider first a particular case.

Claim 2. *If μ is not a root of the polynomial P then the equation*

$$P\left(\frac{d}{dt}\right)x=e^{\mu t}$$

has a solution $x(t)=ae^{\mu t}$ where a is a complex constant to be chosen.

Proof. Indeed, we have

$$P\left(\frac{d}{dt}\right)\left(e^{\mu t}\right)=\sum_{i=0}^{n}a_{n-i}\left(e^{\mu t}\right)^{(i)}=\sum_{i=0}^{n}a_{n-i}\mu^ie^{\mu t}=P(\mu)e^{\mu t}.$$

Therefore, setting $a=\frac{1}{P(\mu)}$, we obtain

$$P\left(\frac{d}{dt}\right)\left(ae^{\mu t}\right)=e^{\mu t}$$

that is, $x(t)=ae^{\mu t}$ is a solution. ■

Note that in this argument it is important that $P(\mu)\neq 0$.

Example. Find a particular solution to the equation:

$$x''+2x'+x=e^t.$$

Note that $P(\lambda)=\lambda^2+2\lambda+1$ and $\mu=1$ is not a root of P. Look for solution in the form $x(t)=ae^t$. Substituting into the equation, we obtain

$$ae^t+2ae^t+ae^t=e^t$$

whence we obtain the equation for a:

$$4a=1,\ a=\frac{1}{4}.$$

Alternatively, we can obtain a from

$$a=\frac{1}{P(\mu)}=\frac{1}{1+2+1}=\frac{1}{4}.$$

Hence, the answer is $x(t)=\frac{1}{4}e^t$.

Consider another equation:

$$x''+2x'+x=\sin t \tag{3.17}$$

Note that $\sin t$ is the imaginary part of e^{it}. So, we first solve

$$x'' + 2x' + x = e^{it}$$

and then take the imaginary part of the solution. Looking for solution in the form $x(t) = ae^{it}$, we obtain

$$a = \frac{1}{P(\mu)} = \frac{1}{i^2 + 2i + 1} = \frac{1}{2i} = -\frac{i}{2}.$$

Hence, the solution is

$$x = -\frac{i}{2}e^{it} = -\frac{i}{2}(\cos t + i\sin t) = \frac{1}{2}\sin t - \frac{i}{2}\cos t.$$

Therefore, its imaginary part $x(t) = -\frac{1}{2}\cos t$ solves the equation (3.17).

Consider yet another right hand side

$$x'' + 2x' + x = e^{-t}\cos t. \tag{3.18}$$

Here $e^{-t}\cos t$ is a real part of $e^{\mu t}$ where $\mu = -1 + i$. Hence, first solve

$$x'' + 2x' + x = e^{\mu t}.$$

Setting $x(t) = ae^{\mu t}$, we obtain

$$a = \frac{1}{P(\mu)} = \frac{1}{(-1+i)^2 + 2(-1+i) + 1} = -1.$$

Hence, the complex solution is $x(t) = -e^{(-1+i)t} = -e^{-t}\cos t - ie^{-t}\sin t$, and the solution to (3.18) is $x(t) = -e^{-t}\cos t$.

Finally, let us combine the above examples into one:

$$x'' + 2x' + x = 2e^t - \sin t + e^{-t}\cos t. \tag{3.19}$$

A particular solution is obtained by combining the above particular solutions:

$$\begin{aligned} x(t) &= 2\left(\frac{1}{4}e^t\right) - \left(-\frac{1}{2}\cos t\right) + \left(-e^{-t}\cos t\right) \\ &= \frac{1}{2}e^t + \frac{1}{2}\cos t - e^{-t}\cos t. \end{aligned}$$

Since the general solution to $x'' + 2x' + x = 0$ is

$$x(t) = (C_1 + C_2 t)\,e^{-t},$$

we obtain the general solution to (3.19)

$$x(t) = (C_1 + C_2 t)\,e^{-t} + \frac{1}{2}e^t + \frac{1}{2}\cos t - e^{-t}\cos t.$$

Consider one more equation

$$x'' + 2x' + x = e^{-t}.$$

This time $\mu = -1$ is a root of $P(\lambda) = \lambda^2 + 2\lambda + 1$ and the above method does not work. Indeed, if we look for a solution in the form $x = ae^{-t}$ then after substitution we get 0 in the left hand side because e^{-t} solves the homogeneous equation.

The case when μ is a root of $P(\lambda)$ is referred to as a *resonance*. This case as well as the case of the general quasi-polynomial in the right hand side is treated in the following theorem.

Theorem 3.9 *Let $R(t)$ be a non-zero polynomial of degree $k \geq 0$ and μ be a complex number. Let m be the multiplicity of μ if μ is a root of P and $m = 0$ if μ is not a root of P. Then the equation*

$$P\left(\frac{d}{dt}\right)x = R(t)\,e^{\mu t}$$

has a solution of the form

$$x(t) = t^m Q(t)\,e^{\mu t},$$

where $Q(t)$ is a polynomial of degree k (which is to be found).

Example. Come back to the equation

$$x'' + 2x' + x = e^{-t}.$$

Here $\mu = -1$ is a root of multiplicity $m = 2$ and $R(t) = 1$ is a polynomial of degree 0. Hence, the solution should be sought in the form

$$x(t) = at^2e^{-t}$$

where a is a constant that replaces Q (indeed, Q must have degree 0 and, hence, is a constant). Substituting this into the equation, we obtain

$$a\left(\left(t^2e^{-t}\right)'' + 2\left(t^2e^{-t}\right)' + t^2e^{-t}\right) = e^{-t}$$

After expansion, we obtain

$$\left(t^2e^{-t}\right)'' + 2\left(t^2e^{-t}\right)' + t^2e^{-t} = 2e^{-t}$$

so that the equation becomes $2a = 1$ and $a = \frac{1}{2}$. Hence, a particular solution is

$$x(t) = \frac{1}{2}t^2e^{-t}.$$

Consider one more example.

$$x'' + 2x' + x = te^{-t}$$

with the same $\mu = -1$ and $R(t) = t$. Since $\deg R = 1$, the polynomial Q must have degree 1, that is, $Q(t) = at + b$. The coefficients a and b can be determined as follows. Substituting

$$x(t) = (at + b)\,t^2e^{-t} = \left(at^3 + bt^2\right)e^{-t}$$

into the equation, we obtain

$$\begin{aligned} x'' + 2x' + x &= \left(\left(at^3 + bt^2\right)e^{-t}\right)'' + 2\left(\left(at^3 + bt^2\right)e^{-t}\right)' + \left(at^3 + bt^2\right)e^{-t} \\ &= \left(2b + 6at\right)e^{-t}. \end{aligned}$$

Hence, comparing with the equation, we obtain

$$2b + 6at = t$$

whence $b = 0$ and $a = \frac{1}{6}$. Hence, the answer is

$$x\left(t\right) = \frac{t^3}{6}e^{-t}.$$

Proof of Theorem 3.9. Consider first the case $m = 0$, when $P\left(\mu\right) \neq 0$ (non-resonant case). Then we prove the claim by induction in $k = \deg R$. If $k = 0$ then this was shown above. Let us prove the inductive step from $k - 1$ to k. It suffices to consider the case $R\left(t\right) = t^k$, that is, the equation

$$P\left(\frac{d}{dt}\right)x = t^k e^{\mu t}, \tag{3.20}$$

because lower order terms are covered by the inductive hypothesis. We need to find a solution $x\left(t\right)$ of the form $Q\left(t\right)e^{\mu t}$ where $\deg Q = k$. Let us first check the function $t^k e^{\mu t}$ as a candidate for the solution.

Claim 3. *For an arbitrary polynomial $P\left(\lambda\right)$ and all $\mu \in \mathbb{C}$ and non-negative integer k, we have*

$$P\left(\frac{d}{dt}\right)\left(t^k e^{\mu t}\right) = t^k P\left(\mu\right)e^{\mu t} + \widetilde{R}\left(t\right)e^{\mu t}, \tag{3.21}$$

where $\widetilde{R}$ is a polynomial of degree $< k$ if $k > 0$, and $\widetilde{R} \equiv 0$ if $k = 0$.

We will use the Leibniz product formula

$$\left(fg\right)^{(n)} = \sum_{i=0}^{n}\binom{n}{i}f^{(i)}g^{(n-i)} = fg^{(n)} + nf'g^{(n-1)} + ... + f^{(n)}g, \tag{3.22}$$

where $f\left(t\right)$ and $g\left(t\right)$ are smooth enough functions of t. For example, if $n = 1$ then we have the product rule

$$\left(fg\right)' = f'g + fg'$$

if $n = 2$ then

$$\left(fg\right)'' = f''g + 2f'g' + fg'',$$

etc. The proof of (3.22) goes by induction in n (see Exercises).

It suffices to prove (3.21) for $P\left(\lambda\right) = \lambda^j$ since for an arbitrary polynomial $P\left(\lambda\right)$ identity (3.21) follows by combining the identities for $P\left(\lambda\right) = \lambda^j$. Using (3.22), we obtain

$$\begin{aligned} \left(t^k e^{\mu t}\right)^{(j)} &= t^k\left(e^{\mu t}\right)^{(j)} + \text{terms with smaller power of } t \text{ times } e^{\mu t} \\ &= t^k \mu^j e^{\mu t} + \widetilde{R}\left(t\right)e^{\mu t} \\ &= t^k P\left(\mu\right)e^{\mu t} + \widetilde{R}\left(t\right)e^{\mu t}, \end{aligned}$$

which proves (3.21).

Let us make change of unknown function as follows:

$$y = x - at^k e^{\mu t}$$

where $a = \frac{1}{P(\mu)}$. Then y satisfies the equation

$$P\left(\frac{d}{dt}\right) y = P\left(\frac{d}{dt}\right) x - aP\left(\frac{d}{dt}\right)\left(t^k e^{\mu t}\right) = t^k e^{\mu t} - t^k e^{\mu t} - a\widetilde{R}\left(t\right) e^{\mu t} = -a\widetilde{R}\left(t\right) e^{\mu t}.$$

Since $\deg \widetilde{R} < k$, by the inductive hypothesis this equation has a solution of the form $y = \widetilde{Q}\left(t\right) e^{\mu t}$ where $\deg \widetilde{Q} = \deg \widetilde{R} < k$. Therefore, we obtain a solution $x\left(t\right)$ of (3.20)

$$x = at^k e^{\mu t} + y = \left(at^k + \widetilde{Q}\right) e^{\mu t} = Q\left(t\right) e^{\mu t}$$

where $\deg Q = k$.

Consider now the resonant case $m > 0$. Again we can assume that $R\left(t\right) = t^k$ and argue by induction in k. Note that, for some polynomial $\widetilde{P}\left(\lambda\right)$, we have the identity

$$P\left(\lambda\right) = \left(\lambda - \mu\right)^m \widetilde{P}\left(\lambda\right), \tag{3.23}$$

and $\widetilde{P}\left(\mu\right) \neq 0$.

Claim 4. *For all $\mu \in \mathbb{C}$, $m \in \mathbb{N}$ and any function $f\left(t\right) \in C^m$, we have*

$$\left(\frac{d}{dt} - \mu\right)^m \left(f e^{\mu t}\right) = f^{(m)} e^{\mu t}. \tag{3.24}$$

It suffices to prove (3.24) for $m = 1$ and then apply induction in m. Indeed,

$$\left(\frac{d}{dt} - \mu\right)\left(f e^{\mu t}\right) = \left(f e^{\mu t}\right)' - \mu f e^{\mu t} = f' e^{\mu t} + f\mu e^{\mu t} - \mu f e^{\mu t} = f' e^{\mu t}.$$

Claim 5. *We have*

$$P\left(\frac{d}{dt}\right)\left(t^{k+m} e^{\mu t}\right) = ct^k \widetilde{P}\left(\mu\right) e^{\mu t} + \widetilde{R}\left(t\right) e^{\mu t}, \tag{3.25}$$

where $c = c\left(k, m\right) > 0$ and $\widetilde{R}$ is a polynomial of degree $< k$ if $k > 0$ and $\widetilde{R} \equiv 0$ if $k = 0$.

Indeed, we have by (3.21)

$$\widetilde{P}\left(\frac{d}{dt}\right)\left(t^{k+m}e^{\mu t}\right)=t^{k+m}\widetilde{P}\left(\mu\right)e^{\mu t}+S\left(t\right)e^{\mu t}$$

where $\deg S<k+m$. Applying (3.24), we obtain

$$\begin{aligned}P\left(\frac{d}{dt}\right)\left(t^{k+m}e^{\mu t}\right)&=\left(\frac{d}{dt}-\mu\right)^{m}\left(t^{k+m}\widetilde{P}\left(\mu\right)e^{\mu t}+S\left(t\right)e^{\mu t}\right)\\&=\widetilde{P}\left(\mu\right)\left(t^{k+m}\right)^{(m)}e^{\mu t}+S^{(m)}\left(t\right)e^{\mu t}\\&=c\widetilde{P}\left(\mu\right)t^{k}e^{\mu t}+S^{(m)}e^{\mu t},\end{aligned}$$

where $c=(k+m)(k+m-1)\ldots(k+1)>0$. Note also that if $k>0$ then

$$\deg S^{(m)}=\max(\deg S-m,0)<k.$$

In the case $k=0$ we have $\deg S<m$ whence $S^{(m)}\equiv 0$. Setting $\widetilde{R}=S^{(m)}$, we finish the proof of Claim 5.

If $k=0$ then Claim 5 implies that

$$P\left(\frac{d}{dt}\right)\left(t^{m}e^{\mu t}\right)=c\widetilde{P}\left(\mu\right)e^{\mu t}$$

whence $x\left(t\right)=at^{m}e^{\mu t}$ solves the equation $P\left(\frac{d}{dt}\right)x=e^{\mu t}$ where $a=\left(c\widetilde{P}\left(\mu\right)\right)^{-1}$. This proves the inductive basis for $k=0$.

For the inductive step from $k-1$ to k, consider a new unknown function

$$y\left(t\right)=x\left(t\right)-at^{k+m}e^{\mu t}$$

so that

$$P\left(\frac{d}{dt}\right)y=P\left(\frac{d}{dt}\right)x-act^{k}\widetilde{P}\left(\mu\right)e^{\mu t}-a\widetilde{R}\left(t\right)e^{\mu t}.$$

Choosing $a=\left(c\widetilde{P}\left(\mu\right)\right)^{-1}$ and using the equation $P\left(\frac{d}{dt}\right)x=t^{k}e^{\mu t}$, we obtain that the two terms on the right hand side cancel out and we obtain the following equation for y:

$$P\left(\frac{d}{dt}\right)y=-a\widetilde{R}\left(t\right)e^{\mu t}.$$

Since $\deg\widetilde{R}<k$, by the inductive hypothesis this equation has a solution of the form

$$y\left(t\right)=t^{m}\widetilde{Q}\left(t\right)e^{\mu t},$$

where $\deg\widetilde{Q}=\deg\widetilde{R}<k$. Hence, we obtain a solution x of the form

$$x\left(t\right)=at^{k+m}e^{\mu t}+t^{m}\widetilde{Q}\left(t\right)e^{\mu t}=t^{m}\left(at^{k}+\widetilde{Q}\left(t\right)\right)e^{\mu t}=t^{m}Q\left(t\right)e^{\mu t},$$

where $\deg Q=k$. ∎

Second proof of Theorem 3.9. This proof is based on the following two facts which we take without proof (see Exercise 43).

Claim 6. *If a complex number μ is a root of a polynomial $P(\lambda)$ of multiplicity m if and only if*

$$P^{(i)}(\mu) = 0 \text{ for all } i = 0, ..., m-1 \text{ and } P^{(m)}(\mu) \neq 0.$$

For example, μ is a simple root if $P(\mu) = 0$ but $P'(\mu) \neq 0$, and μ is a root of multiplicity 2 if $P(\mu) = P'(\mu) = 0$ while $P''(\mu) \neq 0$.

Claim 7. *For any polynomial $P(\lambda)$ any any two smooth enough functions $f(t), g(t)$,*

$$P\left(\frac{d}{dt}\right)(fg) = \sum_{i\geq 0} \frac{1}{i!} f^{(i)} P^{(i)}\left(\frac{d}{dt}\right) g \tag{3.26}$$

where the summation is taken over all non-negative integers i.

In fact, the sum is finite because $P^{(i)} \equiv 0$ for large enough i.

For example, if $P(\lambda) = \lambda^n$ then this becomes the Leibniz formula. Indeed, we have

$$P^{(i)}(\lambda) = n(n-1)...(n-i+1)\lambda^{n-i}$$

and

$$P^{(i)}\left(\frac{d}{dt}\right) g = n(n-1)...(n-i+1) g^{(n-i)}$$

and the formula (3.26) becomes

$$(fg)^{(n)} = \sum_{i=0}^{n} \binom{n}{i} f^{(i)} g^{(n-i)}$$

that is, the Leibniz formula.

Now let us prove that the equation

$$P\left(\frac{d}{dt}\right) x = R(t) e^{\mu t}$$

has a solution in the form

$$x(t) = t^m Q(t) e^{\mu t}$$

where m is the multiplicity of μ and $\deg Q = k = \deg R$. Using (3.26), we have for this function

$$\begin{aligned} P\left(\frac{d}{dt}\right) x &= P\left(\frac{d}{dt}\right)\left(t^m Q(t) e^{\mu t}\right) = \sum_{i\geq 0} \frac{1}{i!}\left(t^m Q(t)\right)^{(i)} P^{(i)}\left(\frac{d}{dt}\right) e^{\mu t} \\ &= \sum_{i\geq 0} \frac{1}{i!}\left(t^m Q(t)\right)^{(i)} P^{(i)}(\mu) e^{\mu t}. \end{aligned}$$

Since $P^{(i)}(\mu) = 0$ for all $i \leq m-1$, we can restrict the summation to $i \geq m$. Since $\left(t^m Q(t)\right)^{(i)} \equiv 0$ for $i > m+k$, we can assume $i \leq m+k$. Denoting

$$y(t) = \left(t^m Q(t)\right)^{(m)} \tag{3.27}$$

we obtain the ODE for y:

$$\frac{P^{(m)}(\mu)}{m!}y+\frac{P^{(m+1)}(\mu)}{(m+1)!}y'+...+\frac{P^{(m+k)}(\mu)}{(m+k)!}y^{(k)}=R(t),$$

which we rewrite in the form

$$b_0y+b_1y'+...+b_ky^{(k)}=R(t) \tag{3.28}$$

where $b_j=\frac{P^{(m+j)}(\mu)}{(m+j)!}$. Note that

$$b_0=\frac{P^{(m)}(\mu)}{m!}\neq 0.$$

Function y is sought as a polynomial of degree k. Indeed, if Q is a polynomial of degree k then it follows from (3.27) that so is y. Conversely, if y is a polynomial (3.29) of degree k then integrating (3.29) m times without adding constants, we obtain the same for $Q(t)$.

Hence, the problem amounts to the following: given a polynomial

$$R(t)=r_0t^k+r_1t^{k-1}+...+r_k$$

of degree k, prove that there exists a polynomial $y(t)$ of degree k that satisfies (3.28). Let us prove the existence of y by induction in k. The inductive basis for $k=0$. Then $R(t)\equiv r_0$, and $y(t)\equiv a$, so that (3.28) becomes $ab_0=r_0$ whence $a=r_0/b_0$ (where we use that $b_0\neq 0$).

The inductive step from $k-1$ to k. Represent y in the from

$$y=at^k+z(t), \tag{3.29}$$

where z is a polynomial of degree $<k$. Substituting (3.29) into (3.28), we obtain the equation for z

$$b_0z+b_1z'+...+b_kz^{(k)}=R(t)-\left(ab_0t^k+ab_1\left(t^k\right)'+...+ab_k\left(t^k\right)^{(k)}\right)=:\widetilde{R}(t).$$

Choosing a from the equation $ab_0=r_0$ we obtain that the term t^k in the right hand side of (3.29) cancels out, whence it follows that $\widetilde{R}(t)$ is a polynomial of degree $<k$. By the inductive hypothesis, the equation

$$b_0z+b_1z'+...+b_{k-1}z^{(k-1)}=\widetilde{R}(t)$$

has a solution $z(t)$ which is a polynomial of degree $\leq k-1$. Then $z^{(k)}=0$ so that we can add to this equation the term $b_kz^{(k)}$ without violating the equation. Hence, the function $y=at^k+z$ solves (3.28) and is a polynomial of degree k. ■

3.5 Some physical examples

Consider a second order ODE

$$x''+px'+qx=f(t). \tag{3.30}$$

It describes various physical phenomena as follows.

3.5.1 Elastic force

Let a point body of mass m moves along axis x under the elastic force whose value is governed by Hooke's law:

$$F_{el} = -ax$$

where a is a positive constant and $x = x(t)$ is the position of the body at time t. The friction force is always given by

$$F_{fr} = -bx'.$$

Finally, assume that there is one more external force $F_{ex} = F(t)$ depending only on t (for example, this may be an electromagnetic force assuming that the body is charged). Then the second Newton's law yields the equation

$$mx'' = F_{el} + F_{fr} + F_{ex} = -ax - bx' + F(t),$$

that is

$$x'' + \frac{b}{m}x' + \frac{a}{m}x = \frac{F(t)}{m}.$$

Clearly, this is an equation of the form (3.30).

3.5.2 Pendulum

A simple gravity pendulum is a small body on the end of a massless string, whose other end is fixed (say, at a point O). When given an initial push, the pendulum swings back and forth under the influence of gravity. Let $x(t)$ be the angular displacement of the pendulum from a downwards vertical axis. Assuming that the length of the string is l and the mass of the body is m, we obtain that the moment of the gravity with respect to the point O is $-mgl\sin x$. The moment of inertia with respect to O is ml^2. Assuming the presence of some additional moment $M(t)$ (for example, periodic pushes to the pendulum), we obtain from the Newton's second law for angular movement

$$ml^2x'' = -mgl\sin x + M(t)$$

whence

$$x'' + \frac{g}{l}\sin x = \frac{M(t)}{ml^2}.$$

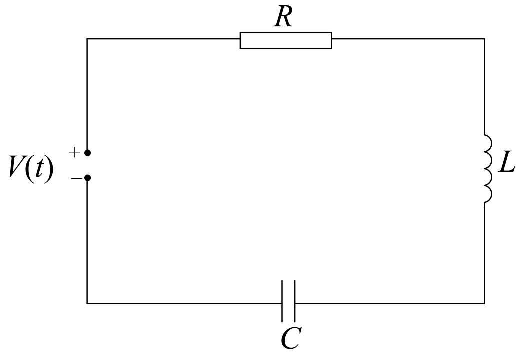
This is the equation of oscillations of the pendulum. If the values of x are small enough then we can replace $\sin x$ by x so that we get the equation of small oscillations

$$x'' + \frac{g}{l}x = \frac{M(t)}{ml^2}.$$

Obviously, it matches (3.30). In the presence of friction it may contains also the term x'.

3.5.3 Electrical circuit

We have considered already an RLC-circuit



As before, let R the resistance, L be the inductance, and C be the capacitance of the circuit. Let $V(t)$ be the voltage of the power source in the circuit and $x(t)$ be the current in the circuit at time t. Then we have see that the equation for $x(t)$ is

$$Lx'' + Rx' + \frac{x}{C} = V'.$$

If $L > 0$ then we can write it in the form

$$x'' + \frac{R}{L}x' + \frac{x}{LC} = \frac{V'}{L},$$

which matches (3.30).

3.5.4 A periodic right hand side

Come back to the equation (3.30) and set $f(t) = A\sin\omega t$, that is, consider the ODE

$$x'' + px' + qx = A\sin\omega t, \tag{3.31}$$

where A, ω are given positive reals. The function $A\sin\omega t$ is a model for a more general periodic force, which makes good physical sense in all the above examples. For example, in the case of electrical circuit the external force has the form $A\sin\omega t$ if the circuit is connected to an electrical socket with the alternating current (AC). In the case of elastic force or a pendulum, a periodic external force occurs, for example, when someone gives periodic pushes to the oscillating body. The number ω is called the *frequency* of the external force (the period $= \frac{2\pi}{\omega}$) or the external frequency, and the number A is called the *amplitude* (the maximum value) of the external force.

Note that in all three examples the coefficients p, q are non-negative, so this will be assumed in the sequel. Moreover, assume in addition that $q > 0$, which is physically most interesting case. To find a particular solution of (3.31), let us consider the ODE with complex right hand side:

$$x'' + px' + qx = Ae^{i\omega t}. \tag{3.32}$$

Consider first the non-resonant case when $i\omega$ is not a root of the characteristic polynomial $P(\lambda) = \lambda^2 + p\lambda + q$. Searching the solution in the from $ce^{i\omega t}$, we obtain

$$c = \frac{A}{P(i\omega)} = \frac{A}{-\omega^2 + pi\omega + q} =: a + ib$$

and the particular solution of (3.32) is

$$(a+ib)\,e^{i\omega t}=(a\cos\omega t-b\sin\omega t)+i\,(a\sin\omega t+b\cos\omega t)\,.$$

Taking its imaginary part, we obtain a particular solution to (3.31)

$$x\,(t)=a\sin\omega t+b\cos\omega t=B\sin\left(\omega t+\varphi\right) \tag{3.33}$$

where

$$B=\sqrt{a^2+b^2}=|c|=\frac{A}{\sqrt{\left(q-\omega^2\right)^2+\omega^2p^2}}$$

and $\varphi\in[0,2\pi)$ is determined from the identities

$$\cos\varphi=\frac{a}{B},\ \ \sin\varphi=\frac{b}{B}.$$

The number B is the amplitude of the solution and φ is the *phase*.

To obtain the general solution to (3.31), we need to add to (3.33) the general solution to the homogeneous equation

$$x''+px'+qx=0.$$

Let λ_1 and λ_2 are the roots of $P\,(\lambda)$, that is,

$$\lambda_{1,2}=-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}.$$

Consider first the case when λ_1 and λ_2 are real. Since $p\geq 0$ and $q>0$, we see that both λ_1 and λ_2 are strictly negative. The general solution of the homogeneous equation has the from

$$\begin{array}{rcl} C_1e^{\lambda_1t}+C_2e^{\lambda_2t}\ \text{ if }\lambda_1 & \neq & \lambda_2,\\ \left(C_1+C_2t\right)e^{\lambda_1t}\ \ \ \text{ if }\lambda_1 & = & \lambda_2.\end{array}$$

In the both cases, it decays exponentially in t as $t\to+\infty$. Hence, the general solution of (3.31) has the form

$$x\,(t)=B\sin\left(\omega t+\varphi\right)+\text{exponentially decaying terms.}$$

As we see, when $t\to\infty$ the leading term of $x\,(t)$ is the above particular solution $B\sin\left(\omega t+\varphi\right)$.

Assume now that λ_1 and λ_2 are complex, say $\lambda_{1,2} = \alpha \pm i\beta$ where

$$\alpha = -p/2 \leq 0 \quad \text{and} \quad \beta = \sqrt{q - \frac{p^2}{4}} > 0.$$

The general solution to the homogeneous equation is

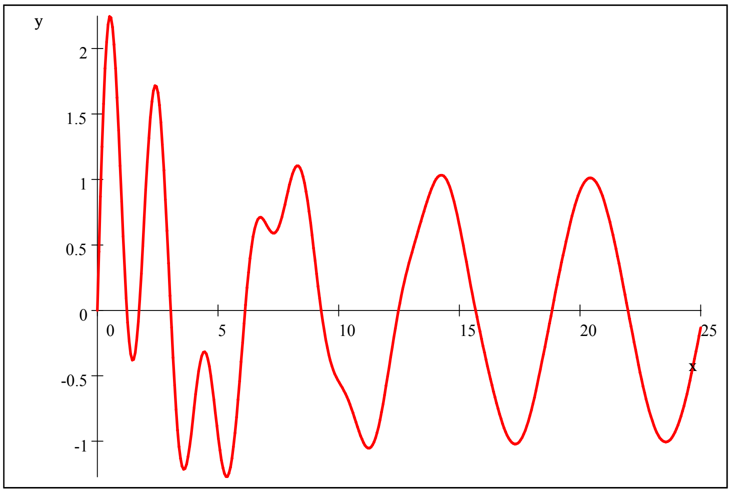
$$e^{\alpha t}\left(C_1 \cos \beta t + C_2 \sin \beta t\right) = Ce^{\alpha t} \sin\left(\beta t + \psi\right).$$

The number β is called the *natural frequency* of the physical system in question (pendulum, electrical circuit, spring) for the obvious reason - in absence of the external force, the system oscillate with the natural frequency β.

Hence, the general solution to (3.31) is

$$x\left(t\right) = B \sin\left(\omega t + \varphi\right) + Ce^{\alpha t} \sin\left(\beta t + \psi\right).$$

If $\alpha < 0$ then the leading term is again $B \sin\left(\omega t + \varphi\right)$. Here is a particular example of such a function: $\sin t + 2e^{-t/4} \sin \pi t$



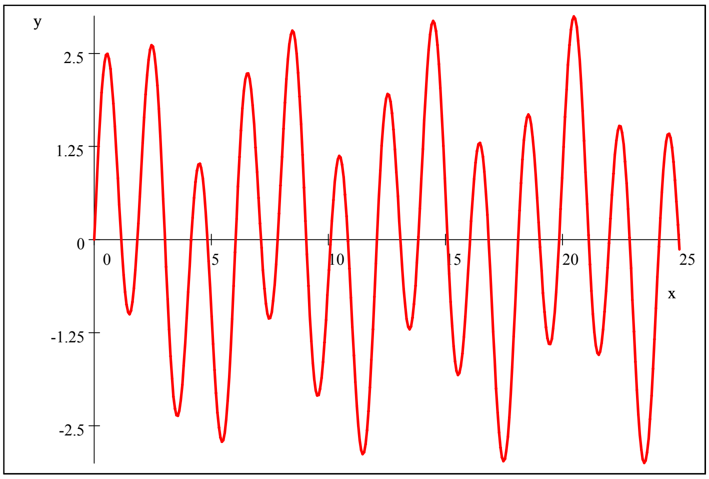
If $\alpha = 0$ that is, the equation has the form

$$x'' + \beta^2 x = A \sin \omega t.$$

The assumption that $i\omega$ is not a root implies $\omega \neq \beta$. The general solution is

$$x\left(t\right) = B \sin\left(\omega t + \varphi\right) + C \sin\left(\beta t + \psi\right),$$

which is the sum of two sin waves with different frequencies - the natural frequency and the external frequency. Here is a particular example of such a function: $\sin t + 2 \sin \pi t$



Consider the following question: what should be the external frequency ω to maximize the amplitude B? Assuming that A does not depend on ω and using the identity

$$B^2 = \frac{A^2}{\omega^4 + (p^2 - 2q)\,\omega^2 + q^2},$$

we see that the maximum B occurs when the denominators takes the minimum value. If $p^2 \geq 2q$ then the minimum value occurs at $\omega = 0$, which is not very interesting physically. Assume that $p^2 < 2q$ (in particular, this implies that $p^2 < 4q$, and, hence, λ_1 and λ_2 are complex). Then the maximum of B occurs when

$$\omega^2 = -\frac{1}{2}\left(p^2 - 2q\right) = q - \frac{p^2}{2}.$$

The value

$$\omega_0 := \sqrt{q - p^2/2}$$

is called the *resonant frequency* of the physical system in question. If the external force has this frequency then the system exhibits the highest response to this force. This phenomenon is called a *resonance*.

Note for comparison that the natural frequency is equal to $\beta = \sqrt{q - p^2/4}$, which is in general different from ω_0. In terms of ω_0 and β, we can write

$$\begin{aligned} B^2 &= \frac{A^2}{\omega^4 - 2\omega_0^2\omega^2 + q^2} = \frac{A^2}{\left(\omega^2 - \omega_0^2\right)^2 + q^2 - \omega_0^4} \\ &= \frac{A^2}{\left(\omega^2 - \omega_0^2\right) + p^2\beta^2}, \end{aligned}$$

where we have used that

$$q^2 - \omega_0^4 = q^2 - \left(q - \frac{p^2}{2}\right)^2 = qp^2 - \frac{p^4}{4} = p^2\beta^2.$$

In particular, the maximum amplitude that occurs when $\omega = \omega_0$ is $B_{\max} = \frac{A}{p\beta}$.

In conclusion, consider the case, when $i\omega$ is a root of $P(\lambda)$, that is

$$(i\omega)^2 + pi\omega + q = 0,$$

which implies $p = 0$ and $q = \omega^2$. In this case $\alpha = 0$ and $\omega = \omega_0 = \beta = \sqrt{q}$, and the equation has the form

$$x'' + \omega^2 x = A \sin \omega t.$$

Considering the ODE

$$x'' + \omega^2 x = Ae^{i\omega t},$$

and searching a particular solution in the form $x(t) = cte^{i\omega t}$, we obtain

$$\begin{aligned} \left(cte^{i\omega t}\right)'' + \omega^2 cte^{i\omega t} &= Ae^{i\omega t} \\ 2i\omega ce^{it\omega} &= Ae^{i\omega t} \end{aligned}$$

whence $c = \frac{A}{2i\omega}$. Alternatively, c can be found directly by

$$c = \frac{A}{P'(i\omega)} = \frac{A}{2i\omega}$$

(see Exercise 44). Hence, the complex particular solution is

$$x(t) = \frac{At}{2i\omega} e^{i\omega t} = -i\frac{At}{2\omega} \cos \omega t + \frac{At}{2\omega} \sin \omega t$$

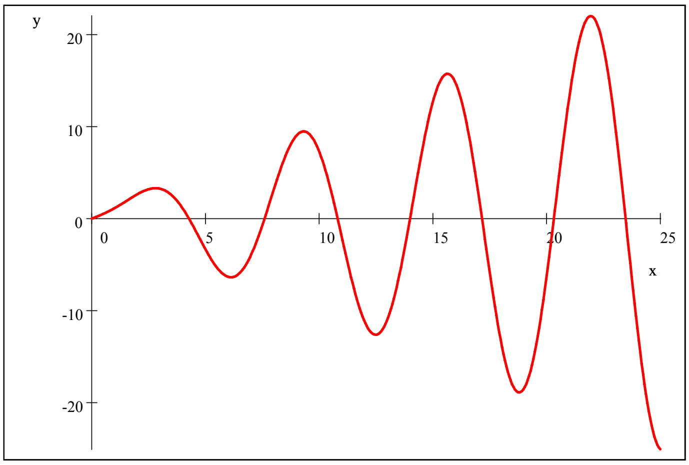
and its imaginary part is

$$x(t) = -\frac{At}{2\omega} \cos \omega t.$$

Hence, the general solution is

$$x(t) = -\frac{At}{2\omega} \cos \omega t + C \sin(\omega t + \psi).$$

Here is an example of such a function: $-t \cos t + 2 \sin t$



Hence, we have a *complete resonance*: the external frequency ω is simultaneously equal to the natural frequency and the resonant frequency. In the case of a complete resonance, the amplitude increases in time unbounded. Since unbounded oscillations are physically impossible, either the system breaks down over time or the mathematical model becomes unsuitable for describing the physical system.

3.6 The method of variation of parameters

We present here this method in a general context of a system. Consider a system $x' = A(t)x$ where $x(t)$ is a function with values in $\mathbb{R}^n$ and $A(t)$ is an $n \times n$ matrix continuously depending on $t \in I$. Let $x_1(t), ..., x_n(t)$ be n independent solutions. Consider now the system

$$x' = A(t)x + B(t) \tag{3.34}$$

where $B(t)$ is a vector in $\mathbb{R}^n$ continuously depending on t. Let us look for a solution to (3.34) in the form

$$x(t) = C_1(t)x_1(t) + ... + C_n(t)x_n(t) \tag{3.35}$$

where $C_1, C_2, .., C_n$ are now unknown real-valued functions to be determined. Originally the representation (3.35) was motivated by the formula $x = C_1x_1 + ... + C_nx_n$ for the general solution to the homogeneous equation $x' = Ax$ and, hence, the method in question is called the method of variation of parameters. However, another point of view on (3.35) is as follows. Since the functions $x_1, ..., x_n$ are linearly independent, by Lemma 3.3 the vectors $x_1(t), ..., x_n(t)$ are linearly independent in $\mathbb{R}^n$ for any $t \in I$. Hence, these vectors form a basis in $\mathbb{R}^n$ for any t, which implies that *any* function $x(t)$ can be represented in the form (3.35).

How to determine the coefficients $C_1(t), ..., C_n(t)$? It follows from (3.35) and $x_i' = Ax_i$, that

$$\begin{aligned} x' &= C_1x_1' + C_2x_2' + ... + C_nx_n' \\ &\quad + C_1'x_1 + C_2'x_2 + ... + C_n'x_n \\ &= C_1Ax_1 + C_2Ax_2 + ... + C_nAx_n \\ &\quad + C_1'x_1 + C_2'x_2 + ... + C_n'x_n \\ &= Ax + C_1'x_1 + C_2'x_2 + ... + C_n'x_n. \end{aligned}$$

Hence, the equation $x' = Ax + B$ becomes

$$C_1'x_1 + C_2'x_2 + ... + C_n'x_n = B. \tag{3.36}$$

Let us rewrite this equation in the matrix form. For that, consider the column-vector

$$C(t) = \begin{pmatrix} C_1(t) \\ ... \\ C_n(t) \end{pmatrix}$$

and the $n \times n$ matrix

$$X = (x_1 \mid x_2 \mid ... \mid x_n)$$

where each x_i is the column vector. The matrix X is called a *fundamental matrix* of the system $x' = Ax$. It follows from the matrix multiplication rule that, for any column vector $V = \begin{pmatrix} v_1 \\ ... \\ v_n \end{pmatrix}$,

$$XV = (x_1 \mid x_2 \mid ... \mid x_n) \begin{pmatrix} v_1 \\ ... \\ v_n \end{pmatrix} = v_1x_1 + ... + v_nx_n.$$

Alternatively, one can verify this identity first for the case when V is one of the vectors $e_1, ..., e_n$ from the canonical basis in $\mathbb{R}^n$ (for example, for $V = e_1$ we trivially get $Xe_1 = x_1$) and then conclude by the linearity that this identity is true for all V.

In particular, we have

$$C_1'x_1 + C_2'x_2 + ... + C_n'x_n = XC'$$

and the equation (3.36) becomes

$$XC' = B.$$

Note that the matrix X is invertible because $\det X$ is the Wronskian, which is non-zero for all t by Lemma 3.3. Therefore, we obtain.

$$C' = X^{-1}B.$$

Integrating in t, we find

$$C(t) = \int X^{-1}(t)\,B(t)\,dt$$

and

$$x(t) = XC = X(t)\int X^{-1}(t)\,B(t)\,dt.$$

Hence, we have proved the following theorem.

Theorem 3.10 *The general solution to the system $x' = A(t)\,x + B(t)$ is given by*

$$x(t) = X(t)\int X^{-1}(t)\,B(t)\,dt \tag{3.37}$$

where $X = (x_1 \mid x_2 \mid ... \mid x_n)$ is a fundamental matrix of the system.

Example. Consider the system

$$\begin{cases} x_1' = -x_2 \\ x_2' = x_1 \end{cases}$$

or, in the vector form,

$$x' = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x.$$

As we have seen before, this system has 2 independent solutions

$$x_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} \sin t \\ -\cos t \end{pmatrix}.$$

Hence, the corresponding fundamental matrix is

$$X = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}$$

and

$$X^{-1} = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}.$$

Consider now the ODE

$$x' = A(t)\,x + B(t)$$

where $B(t) = \begin{pmatrix} b_1(t) \\ b_2(t) \end{pmatrix}$. By (3.37), we obtain the general solution

$$\begin{aligned} x &= \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \int \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \begin{pmatrix} b_1(t) \\ b_2(t) \end{pmatrix} dt \\ &= \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \int \begin{pmatrix} b_1(t)\cos t + b_2(t)\sin t \\ b_1(t)\sin t - b_2(t)\cos t \end{pmatrix} dt. \end{aligned}$$

Consider a particular example of function $B(t)$, say, $B(t) = \begin{pmatrix} 1 \\ t \end{pmatrix}$. Then the integral is

$$\int \begin{pmatrix} \cos t + t\sin t \\ \sin t - t\cos t \end{pmatrix} dt = \begin{pmatrix} 2\sin t - t\cos t + C_1 \\ -2\cos t - t\sin t + C_2 \end{pmatrix}$$

whence

$$\begin{aligned} x &= \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} \begin{pmatrix} 2\sin t - t\cos t + C_1 \\ -2\cos t - t\sin t + C_2 \end{pmatrix} \\ &= \begin{pmatrix} C_1\cos t + C_2\sin t - t \\ C_1\sin t - C_2\cos t + 2 \end{pmatrix} \\ &= \begin{pmatrix} -t \\ 2 \end{pmatrix} + C_1 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} + C_2 \begin{pmatrix} \sin t \\ -\cos t \end{pmatrix}. \end{aligned}$$

Consider now a scalar ODE of order n

$$x^{(n)} + a_1(t)\,x^{(n-1)} + ... + a_n(t)\,x = b(t)$$

where $a_k(t)$ and $b(t)$ are continuous functions on some interval I. Recall that it can be reduced to the vector ODE

$$\mathbf{x}' = A(t)\,\mathbf{x} + B(t)$$

where

$$\mathbf{x}(t) = \begin{pmatrix} x(t) \\ x'(t) \\ ... \\ x^{(n-1)}(t) \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1 \end{pmatrix} \text{ and } \ B = \begin{pmatrix} 0 \\ 0 \\ ... \\ b \end{pmatrix}.$$

If $x_1, ..., x_n$ are n linearly independent solutions to the homogeneous ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n(t)\,x = 0$$

then denoting by $\mathbf{x}_1, ..., \mathbf{x}_n$ the corresponding vector solution, we obtain the fundamental matrix

$$X = (\mathbf{x}_1 \mid \mathbf{x}_2 \mid ...\mid \mathbf{x}_n) = \left(\begin{array}{cccc} x_1 & x_2 & ... & x_n \\ x_1' & x_2' & ... & x_n' \\ ... & ... & ... & ... \\ x_1^{(n-1)} & x_2^{(n-1)} & ... & x_n^{(n-1)} \end{array} \right).$$

We need to multiply X^{-1} by B. Since B has the only non-zero term at the position n, the product $X^{-1}B$ will be equal to b times the n-th column of X^{-1}.

Denote by y_{ik} the element of X^{-1} at position i, k where i is the row index and k is the column index. Denote also by y_k the k-th column of X^{-1}, that is, $y_k = \begin{pmatrix} y_{1k} \\ ... \\ y_{nk} \end{pmatrix}$. Then

$$X^{-1}B = by_n$$

and the general vector solution is

$$\mathbf{x} = X(t) \int b(t)\, y_n(t)\, dt.$$

We need the function $x(t)$ which is the first component of $\mathbf{x}$. Therefore, we need only to take the first row of X to multiply by the column vector $\int b(t)\, y_n(t)\, dt$. Hence,

$$x(t) = \sum_{i=1}^{n} x_i(t) \int b(t)\, y_{in}(t)\, dt.$$

Theorem 3.11 *Let $x_1, ..., x_n$ be n linearly independent solution to*

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = 0$$

and X be the corresponding fundamental matrix. Then, for any continuous function $b(t)$, the general solution to

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = b(t)$$

is given by

$$x(t) = \sum_{i=1}^{n} x_i(t) \int b(t)\, y_{in}(t)\, dt \tag{3.38}$$

where y_{ik} are the entries of the matrix X^{-1}.

Example. Consider the ODE

$$x'' + x = \sin t$$

The independent solutions are $x_1(t) = \cos t$ and $x_2(t) = \sin t$, so that

$$X = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

The inverse is

$$X^{-1} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$$

Hence, the solution is

$$\begin{aligned}
x(t) &= \cos t \int \sin t(-\sin t)\,dt + \sin t \int \sin t \cos t dt \\
&= -\cos t \int \sin^2 t dt + \frac{1}{2}\sin t \int \sin 2t dt \\
&= -\cos t\left(\frac{1}{2}t - \frac{1}{4}\sin 2t + C_1\right) + \frac{1}{4}\sin t\left(-\cos 2t + C_2\right) \\
&= -\frac{1}{2}t\cos t + \frac{1}{4}\left(\sin 2t\cos t - \sin t\cos 2t\right) + C_3\cos t + C_4\sin t \\
&= -\frac{1}{2}t\cos t + C_3\cos t + C_5\sin t.
\end{aligned}$$

Of course, the same result can be obtained by Theorem 3.9.

Consider one more example, when the right hand side is not a quasi-polynomial:

$$x'' + x = \tan t. \tag{3.39}$$

Then as above we obtain[4]

$$\begin{aligned}
x &= \cos t \int \tan t(-\sin t)\,dt + \sin t \int \tan t \cos t dt \\
&= \cos t\left(\frac{1}{2}\ln\left(\frac{1-\sin t}{1+\sin t}\right) + \sin t\right) - \sin t\cos t + C_1\cos t + C_2\sin t \\
&= \frac{1}{2}\cos t\ln\left(\frac{1-\sin t}{1+\sin t}\right) + C_1\cos t + C_2\sin t.
\end{aligned}$$

Let us show how one can use the method of variation of parameters directly, without using the formula (3.38). Again, knowing that the independent solutions of $x'' + x = 0$ are $x_1 = \cos t$ and $x_2 = \sin t$, let us look for the solution of (3.39) in the form

$$x(t) = C_1(t)\cos t + C_2(t)\sin t. \tag{3.40}$$

To obtain the equations for C_1, C_2, differentiate this formula:

$$\begin{aligned}
x'(t) &= -C_1(t)\sin t + C_2(t)\cos t \\
&\quad +C_1'(t)\cos t + C_2'(t)\sin t
\end{aligned} \tag{3.41}$$

The first equation for C_1, C_2 comes from the requirement that the second line here (that is, the sum of the terms with C_1' and C_2') must vanish, that is,

$$C_1'\cos t + C_2'\sin t = 0.$$

[4] The intergal $\int \tan x \sin t dt$ is taken as follows:

$$\int \tan x \sin t dt = \int \frac{\sin^2 t}{\cos t}dt = \int \frac{1-\cos^2 t}{\cos t}dt = \int \frac{dt}{\cos t} - \sin t.$$

Next, we have

$$\int \frac{dt}{\cos t} = \int \frac{d\sin t}{\cos^2 t} = \int \frac{d\sin t}{1-\sin^2 t} = \frac{1}{2}\ln\frac{1-\sin t}{1+\sin t}.$$

The motivation for this choice is as follows. Switching to the normal system, one must have the identity

$$\mathbf{x}(t) = C_1(t)\mathbf{x}_1(t) + C_2\mathbf{x}_2(t).$$

The first component of this vector identity gives the scalar identity (3.40). The second component of the vector identity implies

$$x'(t) = C_1(t)(\cos t)' + C_2(t)(\sin t)'$$

because the second components of the vectors $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2$ are the derivatives of the first components. Comparing with (3.41), we see that the sum of all terms containing C_1' and C_2' must be zero.

It follows from (3.41) that

$$\begin{aligned} x'' &= -C_1\cos t - C_2\sin t \\ &\quad -C_1'\sin t + C_2'\cos t, \end{aligned}$$

whence

$$x'' + x = -C_1'\sin t + C_2'\cos t$$

(note that the terms with C_1 and C_2 cancel out and that this will always be the case provided all computations are correct). Hence, the second equation for C_1' and C_2' is

$$-C_1'\sin t + C_2'\cos t = \tan t,$$

Solving the system of linear algebraic equations

$$\begin{cases} C_1'\cos t + C_2'\sin t = 0 \\ -C_1'\sin t + C_2'\cos t = \tan t \end{cases}$$

we obtain

$$C_1' = -\tan t\sin t, \qquad C_2' = \sin t$$

whence

$$x(t) = C_1\cos t + C_2\sin t = -\cos t\int \tan t\sin t dt + \sin t\int \sin t dt.$$

We are left to evaluate the integrals, which however was already done above.

3.7 The Liouville formula

Let $x_1(t), ..., x_n(t)$ be n functions from an interval $I \subset \mathbb{R}$ to $\mathbb{R}^n$. Consider the $n\times n$ matrix (x_{ij}) where x_{ij} is the i-th component of x_j, that is, the matrix that has $x_1, x_2, ..., x_n$ as columns. The *Wronskian* of the sequence $\{x_j\}_{j=1}^n$ is the determinant of this matrix, that is,

$$W(t) = \det(x_{ij}) = \det(x_1 \mid x_2 \mid ... \mid x_n).$$

Theorem 3.12 (The Liouville formula) *Let $\{x_i\}_{i=1}^n$ be a sequence of n solutions of the ODE $x' = A(t)x$, where $A: I \to \mathbb{R}^{n\times n}$ is continuous. Then the Wronskian $W(t)$ of this sequence satisfies the identity*

$$W(t) = W(t_0)\exp\int_{t_0}^t \operatorname{trace} A(\tau)\,d\tau, \tag{3.42}$$

for all $t, t_0 \in I$.

Recall that the trace (*Spur*) $\operatorname{trace} A$ of the matrix A is the sum of all the diagonal entries of the matrix.

Proof. Denote by r_i the i-th row of the Wronskian, that is $r_i = (x_{i1}, x_{i2}, ..., x_{in})$ and

$$W = \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n \end{pmatrix}$$

We use the following formula for differentiation of the determinant, which follows from the full expansion of the determinant and the product rule[5]:

$$W'(t) = \det \begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} + \det \begin{pmatrix} r_1 \\ r_2' \\ ... \\ r_n \end{pmatrix} + ... + \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n' \end{pmatrix} \tag{3.43}$$

The fact that each vector x_j satisfies the equation $x_j' = Ax_j$ can be written in the coordinate form as follows

$$x_{ij}' = \sum_{k=1}^{n} A_{ik} x_{kj}$$

whence we obtain the identity for the rows:

$$r_i' = \sum_{k=1}^{n} A_{ik} r_k.$$

That is, the derivative r_i' of the i-th row is a linear combination of all rows r_k. For example,

$$r_1' = A_{11} r_1 + A_{12} r_2 + ... + A_{1n} r_n$$

which implies that

$$\det \begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} = A_{11} \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n \end{pmatrix} + A_{12} \det \begin{pmatrix} r_2 \\ r_2 \\ ... \\ r_n \end{pmatrix} + ... + A_{1n} \det \begin{pmatrix} r_n \\ r_2 \\ ... \\ r_n \end{pmatrix}.$$

All the determinants except for the 1st one vanish since they have equal rows. Hence,

$$\det \begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} = A_{11} \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n \end{pmatrix} = A_{11} W(t).$$

[5] If $f_1(t), ..., f_n(t)$ are real-valued differentiable functions then the product rule implies

$$(f_1 ... f_n)' = f_1' f_2 ... f_n + f_1 f_2' ... f_n + ... + f_1 f_2 ... f_n'.$$

Hence, when differentiating the full expansion of the determinant, each term of the determinant gives rise to n terms where one of the multiples is replaced by its derivative. Combining properly all such terms, we obtain that the derivative of the determinant is the sum of n determinants where one of the rows is replaced by its derivative.

Evaluating similarly the other terms in (3.43), we obtain

$$W'(t) = (A_{11} + A_{22} + ... + A_{nn}) W(t) = (\operatorname{trace} A) W(t).$$

By Lemma 3.3, $W(t)$ is either identical 0 or never zero. In the first case there is nothing to prove. In the second case, solving the above ODE for $W(t)$ by the method of separation of variables, we obtain

$$\ln |W(t)| = \int \operatorname{trace} A(t)\, dt$$

whence

$$W(t) = C \exp\left(\int \operatorname{trace} A(t)\, dt\right).$$

Comparing the identities at times t and t_0, we obtain (3.42). ■

Let $x_1(t), ..., x_n(t)$ are n real-valued functions on an interval I of the class C^{n-1}. Recall that their Wronskian is defined by

$$W(t) = \det \begin{pmatrix} x_1 & x_2 & ... & x_n \\ x_1' & x_2' & ... & x_n' \\ ... & ... & ... & ... \\ x_1^{(n-1)} & x_2^{(n-1)} & ... & x_n^{(n-1)} \end{pmatrix}.$$

Corollary. *Consider an ODE*

$$x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = 0$$

where $a_k(t)$ are continuous functions on an interval $I \subset \mathbb{R}$. If $x_1(t), ..., x_n(t)$ are n solutions to this equation then their Wronskian $W(t)$ satisfies the identity

$$W(t) = W(t_0) \exp\left(-\int_{t_0}^{t} a_1(\tau)\, d\tau\right). \tag{3.44}$$

Proof. The scalar ODE is equivalent to the normal system $\mathbf{x}' = A\mathbf{x}$ where

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1 \end{pmatrix} \text{ and } \mathbf{x} = \begin{pmatrix} x \\ x' \\ ... \\ x^{(n-1)} \end{pmatrix}.$$

Since the Wronskian of the normal system coincides with $W(t)$, (3.44) follows from (3.42) because $\operatorname{trace} A = -a_1$. ■

In the case of the ODE of the 2nd order

$$x'' + a_1(t) x' + a_2(t) x = 0$$

the Liouville formula can help in finding the general solution if a particular solution is known. Indeed, if $x_0(t)$ is a particular non-zero solution and $x(t)$ is any other solution then we have by (3.44)

$$\det \begin{pmatrix} x_0 & x \\ x_0' & x' \end{pmatrix} = C \exp\left(-\int a_1(t)\, dt\right),$$

that is

$$x_0 x' - x x_0' = C \exp\left(-\int a_1(t)\, dt\right).$$

Using the identity

$$\frac{x_0 x' - x x_0'}{x_0^2} = \left(\frac{x}{x_0}\right)'$$

we obtain the ODE

$$\left(\frac{x}{x_0}\right)' = \frac{C \exp\left(-\int a_1(t)\, dt\right)}{x_0^2}, \tag{3.45}$$

and by integrating it we obtain $\frac{x}{x_0}$ and, hence, x (cf. Exercise 36).

Example. Consider the ODE

$$x'' - 2\left(1 + \tan^2 t\right) x = 0.$$

One solution can be guessed $x_0(t) = \tan t$ using the fact that

$$\frac{d}{dt} \tan t = \frac{1}{\cos^2 t} = \tan^2 t + 1$$

and

$$\frac{d^2}{dt^2} \tan t = 2 \tan t \left(\tan^2 t + 1\right).$$

Hence, for $x(t)$ we obtain from (3.45)

$$\left(\frac{x}{\tan t}\right)' = \frac{C}{\tan^2 t}$$

whence[6]

$$x = C \tan t \int \frac{dt}{\tan^2 t} = C \tan t \left(-t - \cot t + C_1\right),$$

that is, renaming the constants,

$$x(t) = C_1 (t \tan t + 1) + C_2 \tan t.$$

[6] To evaluate the integral $\int \frac{dt}{\tan^2 t} = \int \cot^2 t dt$ use the identity

$$(\cot t)' = -\cot^2 t - 1$$

that yields

$$\int \cot^2 t dt = -t - \cot t + C.$$

3.8 Linear homogeneous systems with constant coefficients

Here we will be concerned with finding the general solution to linear systems of the form $x' = Ax$ where A is a constant $n \times n$ matrix and $x(t)$ is a function on $\mathbb{R}$ with values in $\mathbb{R}^n$. As we know, it suffices to find n linearly independent solutions and then take their linear combination. We start with a simple observation. Let us try to find a solution in the form $x = e^{\lambda t} v$ where v is a non-zero vector in $\mathbb{R}^n$ that does not depend on t. Then the equation $x' = Ax$ becomes

$$\lambda e^{\lambda t} v = e^{\lambda t} A v$$

that is, $Av = \lambda v$. Hence, if v is an eigenvector of the matrix A with the eigenvalue λ then the function $x(t) = e^{\lambda t} v$ is a solution.

Claim 1. *If an $n \times n$ matrix A has n linearly independent eigenvectors $v_1, ..., v_n$ (that is, a basis of eigenvectors) with the eigenvalues $\lambda_1, ..., \lambda_n$ then the general solution of the ODE $x' = Ax$ is*

$$x(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t} v_k. \tag{3.46}$$

Proof. As we have seen already, each function $e^{\lambda_k t} v_k$ is a solution. Since vectors $\{v_k\}_{k=1}^n$ are linearly independent, the functions $\left\{e^{\lambda_k t} v_k\right\}_{k=1}^n$ are linearly independent, whence the claim follows. ■

In particular, if A has n distinct eigenvalues then their eigenvectors are automatically linearly independent, and Claim 1 applies.

Example. Consider a normal system

$$\begin{cases} x_1' = x_2 \\ x_2' = x_1 \end{cases}$$

The matrix A is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Recall that in order to find the eigenvalues one first writes the characteristic equation

$$\det(A - \lambda I) = 0$$

that is,

$$\det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1 = 0$$

whence $\lambda_{1,2} = \pm 1$. If λ is an eigenvalue then the eigenvectors satisfy the equation

$$(A - \lambda I) v = 0.$$

For $\lambda = 1$ we obtain

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = 0$$

which gives only one independent equation $v^1 - v^2 = 0$. Hence, an eigenvector for $\lambda_1 = 1$ is

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Similarly, for $\lambda = \lambda_2 = -1$ we have the equation for v

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = 0$$

which gives only one independent equation $v^1 + v^2 = 0$. Hence, an eigenvector for $\lambda_2 = -1$ is

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Since the vectors v_1 and v_2 are independent, we obtain the general solution in the form

$$C_1 e^t \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2 e^{-t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} C_1 e^t + C_2 e^{-t} \\ C_1 e^t - C_2 e^{-t} \end{pmatrix}.$$

In general, the eigenvalues and eigenvectors may be complex so that the formula (3.46) gives the general complex solution. If the matrix A is real then one can extract the real solution as follows. If λ is an imaginary eigenvalue then also $\overline{\lambda}$ is an eigenvalue because the characteristic equation has real coefficients. If v is an eigenvector of λ then $\overline{v}$ is an eigenvector of $\overline{\lambda}$ because $Av = \lambda v$ implies $A\overline{v} = \overline{\lambda}\overline{v}$.

Claim 2. *We have*

$$\operatorname{span}\left(e^{\lambda t} v, e^{\overline{\lambda} t}\overline{v}\right) = \operatorname{span}\left(\operatorname{Re}\left(e^{\lambda t} v\right), \operatorname{Im}\left(e^{\lambda t} v\right)\right).$$

In particular, in the sequence of independent solutions the functions $e^{\lambda t} v, e^{\overline{\lambda} t}\overline{v}$ *can be replaced by* $\operatorname{Re}\left(e^{\lambda t} v\right)$, $\operatorname{Im}\left(e^{\lambda t} v\right)$.

Proof. This is trivial because

$$\operatorname{Re} e^{\lambda t} v = \frac{e^{\lambda t} v + e^{\overline{\lambda} t}\overline{v}}{2} \quad \text{and} \quad \operatorname{Im} e^{\lambda t} v = \frac{e^{\lambda t} v - e^{\overline{\lambda} t}\overline{v}}{2i}.$$

■

Example. Consider a normal system

$$\begin{cases} x_1' = -x_2 \\ x_2' = x_1 \end{cases}$$

The matrix A is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and the the characteristic equation is

$$\det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1 = 0$$

whence $\lambda_{1,2} = \pm i$. For $\lambda = i$ we obtain the equation

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = 0$$

which amounts to the single equation $v^1 - iv^2 = 0$. An eigenvector is

$$v_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$

and the corresponding solution of the ODE is

$$x_1(t) = e^{it}\begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin t + i\cos t \\ \cos t + i\sin t \end{pmatrix}.$$

The general solution is

$$x(t) = C_1 \operatorname{Re} x_1 + C_2 \operatorname{Im} x_1 = C_1 \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} + C_2 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Example. Consider a normal system

$$\begin{cases} x_1' = x_2 \\ x_2' = 0. \end{cases}$$

This system is trivially solved to obtain $x_2 = C$ and $x_1 = Ct + C_1$. However, if we try to solve it using the above method, we fail. Indeed, the matrix is $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, the characteristic equation is

$$\det\begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix} = \lambda^2 = 0,$$

the only eigenvalue is $\lambda = 0$. The eigenvector satisfies the equation

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = 0$$

whence $v^2 = 0$. That is, the only eigenvector (up to a constant multiple) is $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and the only solution we obtain is $x(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The problem lies in the properties of this matrix - it does not have a basis of eigenvectors, which is needed for this method.

As it is known from Linear Algebra, any symmetric matrix has a basis of eigenvectors. However, as we have seen, it is not the case in general. In order to understand what to do, we try a different approach.

3.8.1 Functions of operators and matrices

Recall that an scalar ODE $x' = Ax$ has a solution $x(t) = Ce^{At}t$. Now if A is a $n \times n$ matrix, we may be able to use this formula if we define what is e^{At}. It suffices to define what is e^A for any matrix A. It is convenient to do this for linear operators acting in $\mathbb{R}^n$. Denote the family of all linear operators in $\mathbb{R}^n$ by $\mathcal{L}(\mathbb{R}^n)$. This is obviously a linear space over $\mathbb{R}$ (or $\mathbb{C}$). Besides, there is the operation of (noncommutative) multiplication in this space, simply given by composition of operators.

Any $n \times n$ matrix defines a linear operator in $\mathbb{R}^n$ using multiplication of column-vectors by this matrix. Moreover, any linear operator can be represented in this form so that there is an one-to-one correspondence[7] between linear operators and matrices.

[7] This correspondence depends on the choice of a basis in $\mathbb{R}^n$ – see the next Section.

If we fix a norm in $\mathbb{R}^n$ then we can define the *operator norm* in $\mathcal{L}(\mathbb{R}^n)$ by

$$\|A\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{3.47}$$

It is known that $\|A\|$ is finite and satisfies all properties of a norm (that is, the positivity, the scaling property, and the triangle inequality). In addition, the operator norm satisfies the property

$$\|AB\| \leq \|A\| \, \|B\| . \tag{3.48}$$

Indeed, it follows from (3.47) that $\|Ax\| \leq \|A\| \, \|x\|$ whence

$$\|(AB)\, x\| = \|A\,(Bx)\| \leq \|A\| \, \|Bx\| \leq \|A\| \, \|B\| \, \|x\|$$

whence (3.48) follows.

Hence, $\mathcal{L}(\mathbb{R}^n)$ is a normed linear space. Since this space is finite dimensional (its dimension is n^2), it is complete as a normed space. This allows to consider limits and series of operators, and the latter can be used to define e^A as follows.

Definition. If $A \in \mathcal{L}(\mathbb{R}^n)$ then define $e^A \in \mathcal{L}(\mathbb{R}^n)$ by means of the identity

$$e^A = \mathrm{id} + A + \frac{A^2}{2!} + \ldots + \frac{A^k}{k!} + \ldots = \sum_{k=0}^{\infty} \frac{A^k}{k!},$$

where the convergence is understood in the sense of the operator norm in $\mathcal{L}(\mathbb{R}^n)$.

Claim 3. *The exponential series converges for any* $A \in \mathcal{L}(\mathbb{R}^n)$.

Proof. It suffices to show that the series converges absolutely, that is,

$$\sum_{k=0}^{\infty} \left\| \frac{A^k}{k!} \right\| < \infty.$$

It follows from (3.48) that $\left\|A^k\right\| \leq \|A\|^k$ whence

$$\sum_{k=0}^{\infty} \left\| \frac{A^k}{k!} \right\| \leq \sum_{k=0}^{\infty} \frac{\|A\|^k}{k!} = e^{\|A\|} < \infty,$$

and the claim follows. ■

Lemma 3.13 *For any $A \in \mathcal{L}(\mathbb{R}^n)$ the function $F(t) = e^{At}$ satisfies the equation $F' = AF$. Consequently, the general solution of the ODE $x' = Ax$ is given by $x = e^{At}v$ where $v \in \mathbb{R}^n$ is an arbitrary vector.*

Proof. We have by the definition

$$F(t) = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!}.$$

Consider the series of the derivatives:

$$G(t) = \sum_{k=0}^{\infty} \left(\frac{A^k t^k}{k!} \right)' = \sum_{k=1}^{\infty} \frac{A^k t^{k-1}}{(k-1)!}.$$

It is easy to see (in the same way as Claim 3) that this series converges absolutely and locally uniformly in t. Hence, $G = F'$, whence we obtain

$$F' = A \sum_{k=1}^{\infty} \frac{A^{k-1} t^{k-1}}{(k-1)!} v = AF.$$

Obviously,

$$x' = \lim_{h\to 0} \frac{e^{A(t+h)}v - e^{At}v}{h} = \left(\lim_{h\to 0} \frac{e^{A(t+h)} - e^{At}}{h} \right) v = \left(e^{At}\right)' v = \left(Ae^{At}\right) v = Ax$$

so that $x(t)$ solves the ODE for all v. Having chosen n linearly independent vectors $v_1, ..., v_n$, we obtain n solutions $x_k = e^{At} v_k$ that are also linearly independent (which follows from Lemma 3.3). Hence, the general solution is

$$C_1 e^{At} v_1 + ... + C_n e^{At} v_n = e^{At} \left(C_1 v_1 + ... + C_n v_n\right)$$

which can be simply written as $e^{At}v$ for any $v \in \mathbb{R}^n$. ■

Remark. Note that the function $x(t) = e^{At}v$ solves the IVP

$$\begin{cases} x' = Ax \\ x(0) = v. \end{cases}$$

Choosing $v_1, ..., v_n$ to be the canonical basis in $\mathbb{R}^n$, we obtain that the columns of the matrix e^{At} form a basis in the space of solutions, that is, e^{At} is a fundamental matrix of the system $x' = Ax$.

Example. Let A be the diagonal matrix

$$A = \operatorname{diag}(\lambda_1, ..., \lambda_n).$$

Then

$$A^k = \operatorname{diag}\left(\lambda_1^k, ..., \lambda_n^k\right)$$

and

$$e^{At} = \operatorname{diag}\left(e^{\lambda_1 t}, ..., e^{t\lambda_k}\right).$$

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then $A^2 = 0$ and all higher power of A are also 0 and we obtain

$$e^{At} = \operatorname{id} + At = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Hence, for the ODE $x' = Ax$, we obtain two independent solutions

$$x_1(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} t \\ 1 \end{pmatrix}$$

and the general solution

$$x(t) = \begin{pmatrix} C_1 + C_2 t \\ C_2 \end{pmatrix}.$$

Definition. Operators $A, B \in \mathcal{L}(\mathbb{R}^n)$ are said *to commute* if $AB = BA$.

In general, the operators do not have to commute. If A and B commute then various nice formulas take places, for example,

$$(A + B)^2 = A^2 + 2AB + B^2. \tag{3.49}$$

Indeed, in general we have

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2,$$

which yields (3.49) if $AB = BA$.

Lemma 3.14 *If A and B commute then*

$$e^{A+B} = e^A e^B.$$

Proof. Let us prove a sequence of claims.

Claim 1. *If A, B, C commute pairwise then so do AC and B.*

Indeed,

$$(AC)B = A(CB) = A(BC) = (AB)C = (BA)C = B(AC).$$

Claim 2. *If A and B commute then so do e^A and B.*

Indeed, it follows from Claim 1 that A^k and B commute for any natural k, whence

$$e^A B = \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right) B = B \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right) = Be^A.$$

Claim 3. *If $A(t)$ and $B(t)$ are differentiable functions from $I \to \mathcal{L}(\mathbb{R}^n)$ then*

$$(A(t)B(t))' = A'(t)B(t) + A(t)B'(t).$$

Warning: watch the correct order of the multiples.

Indeed, we have for any component

$$(AB)'_{ij} = \left(\sum_k A_{ik}B_{kj}\right)' = \sum_k A'_{ik}B_{kj} + \sum_k A_{ik}B'_{kj} = (A'B)_{ij} + (AB')_{ij} = (A'B + AB')_{ij}.$$

Now we can prove the lemma. Consider the function $F : \mathbb{R} \to \mathcal{L}(\mathbb{R}^n)$ defined by

$$F(t) = e^{tA}e^{tB}.$$

Differentiating it using Lemma 3.13, Claims 2 and 3, we obtain

$$F'(t) = \left(e^{tA}\right)' e^{tB} + e^{tA}\left(e^{tB}\right)' = Ae^{tA}e^{tB} + e^{tA}Be^{tB} = Ae^{tA}e^{tB} + Be^{tA}e^{tB} = (A + B)F(t).$$

On the other hand, Lemma 3.13 the function $G(t) = e^{t(A+B)}$ satisfies the same equation

$$G' = (A + B)G.$$

Since $G(0) = F(0) = \operatorname{id}$ (because $e^0 = \operatorname{id}$) we obtain that the vector functions $F(t)$ and $G(t)$ solve the same IVP, whence by the uniqueness theorem they are identically equal. In particular, $F(1) = G(1)$, which means $e^A e^B = e^{A+B}$. ■

Alternative proof. Let us briefly discuss a direct proof of $e^{A+B} = e^A e^B$. One first proves the binomial formula

$$(A+B)^n = \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k}$$

using the fact that A and B commute (this can be done by induction in the same way as for numbers). Then we have

$$e^{A+B} = \sum_{n=0}^{\infty} \frac{(A+B)^n}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{A^k B^{n-k}}{k!\,(n-k)!}$$

and, using the Cauchy product formula,

$$e^A e^B = \sum_{m=0}^{\infty} \frac{A^m}{m!} \sum_{l=0}^{\infty} \frac{B^l}{l!} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{A^k B^{n-k}}{k!\,(n-k)!}.$$

Of course, one need to justify the Cauchy product formula for absolutely convergent series of operators. ■

Definition. An $n \times n$ matrix is called a *Jordan cell* if it has the form

$$A = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}, \tag{3.50}$$

where λ is any complex number.

Here all the entries on the main diagonal are λ and all the entries just above the main diagonal are 1 (and all other values are 0). Let us use Lemma 3.14 in order to evaluate e^{tA} where A is a Jordan cell. Clearly, we have $A = \lambda \operatorname{id} + N$ where

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}. \tag{3.51}$$

A matrix (3.51) is called a *nilpotent Jordan cell*. Since the matrices $\lambda \operatorname{id}$ and N commute (because id commutes with anything), Lemma 3.14 yields

$$e^{tA} = e^{t\lambda \operatorname{id}} e^{tN} = e^{t\lambda} e^{tN}. \tag{3.52}$$

Hence, we need to evaluate e^{tN}, and for that we first evaluate the powers N^2, N^3, etc. Observe that the components of matrix N are as follows

$$N_{ij} = \begin{cases} 1, & \text{if } j = i+1 \\ 0, & \text{otherwise} \end{cases}$$

where i is the row index and j is the column index. It follows that

$$\left(N^2\right)_{ij} = \sum_{k=1}^{n} N_{ik} N_{kj} = \begin{cases} 1, & \text{if } j = i+2 \\ 0, & \text{otherwise} \end{cases}$$

that is,

$$N^2 = \begin{pmatrix} 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

Here the entries with value 1 are located on the diagonal that is two positions above the main diagonal. Similarly, we obtain

$$N^k = \begin{pmatrix} 0 & \ddots & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

where the entries with value 1 are located on the diagonal that is k positions above the main diagonal, provided $k < n$, and $N^k = 0$ if $k \geq n$.

Any matrix A with the property that $A^k = 0$ for some natural k is called *nilpotent*. Hence, N is a nilpotent matrix, which explains the term "a nilpotent Jordan cell". It follows that

$$e^{tN} = \mathrm{id} + \frac{t}{1!}N + \frac{t^2}{2!}N^2 + ... + \frac{t^{n-1}}{(n-1)!}N^{n-1} = \begin{pmatrix} 1 & \frac{t}{1!} & \frac{t^2}{2!} & \ddots & \frac{t^{n-1}}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!} \\ \vdots & & \ddots & \ddots & \frac{t}{1!} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}. \tag{3.53}$$

Combining with (3.52), we obtain the following statement:

Lemma 3.15 *If A is a Jordan cell* (3.50) *then, for any $t \in \mathbb{R}$,*

$$e^{tA} = \begin{pmatrix} e^{\lambda t} & \frac{t}{1!}e^{t\lambda} & \frac{t^2}{2!}e^{t\lambda} & \ddots & \frac{t^{n-1}}{(n-1)!}e^{t\lambda} \\ 0 & e^{t\lambda} & \frac{t}{1!}e^{t\lambda} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!}e^{t\lambda} \\ \vdots & & \ddots & \ddots & \frac{t}{1!}e^{t\lambda} \\ 0 & \cdots & \cdots & 0 & e^{t\lambda} \end{pmatrix}. \tag{3.54}$$

By Lemma 3.13, the columns of the matrix e^{tA} form linearly independent solutions to the system $x' = Ax$. Hence, we obtain the following basis of solutions:

$$\begin{aligned} x_1(t) &= e^{\lambda t}(1, 0, ..., 0) \\ x_2(t) &= e^{\lambda t}\left(\frac{t}{1!}, 1, 0, ..., 0\right) \\ x_3(t) &= e^{\lambda t}\left(\frac{t^2}{2!}, \frac{t}{1!}, 1, 0, ..., 0\right) \\ &\ldots \\ x_n(t) &= e^{\lambda t}\left(\frac{t^{n-1}}{(n-1)!}, ..., \frac{t}{1!}, 1\right), \end{aligned}$$

and the general solution is $C_1x_1 + ... + C_nx_n$ where $C_1, ..., C_n$ are arbitrary constants.

Definition. If A is a $m \times m$ matrix and B is a $l \times l$ matrix then their *tensor product* is an $n \times n$ matrix C where $n = m + l$ and

$$C = \left(\begin{array}{|c|c|} \hline A & 0 \\ \hline 0 & B \\ \hline \end{array}\right)$$

That is, matrix C consists of two blocks A and B located on the main diagonal, and all other terms are 0.

Notation for the tensor product: $C = A \otimes B$.

Lemma 3.16 *We have*

$$e^{A \otimes B} = e^A \otimes e^B,$$

that is, in the above notation,

$$e^C = \left(\begin{array}{|c|c|} \hline e^A & 0 \\ \hline 0 & e^B \\ \hline \end{array}\right).$$

Proof. We claim that if A_1, A_2 are $m \times m$ matrices and B_1, B_2 are $l \times l$ matrices then

$$(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1A_2) \otimes (B_1B_2). \tag{3.55}$$

Indeed, in the extended form this identity means

$$\left(\begin{array}{|c|c|} \hline A_1 & 0 \\ \hline 0 & B_1 \\ \hline \end{array}\right)\left(\begin{array}{|c|c|} \hline A_2 & 0 \\ \hline 0 & B_2 \\ \hline \end{array}\right) = \left(\begin{array}{|c|c|} \hline A_1A_2 & 0 \\ \hline 0 & B_1B_2 \\ \hline \end{array}\right)$$

which follows easily from the rule of multiplication of matrices. Hence, the tensor product commutes with the matrix multiplication. It is also obvious that the tensor product commutes with addition of matrices and taking limits. Therefore, we obtain

$$e^{A\otimes B}=\sum_{k=0}^{\infty}\frac{(A\otimes B)^k}{k!}=\sum_{k=0}^{\infty}\frac{A^k\otimes B^k}{k!}=\left(\sum_{k=0}^{\infty}\frac{A^k}{k!}\right)\otimes\left(\sum_{k=0}^{\infty}\frac{B^k}{k!}\right)=e^A\otimes e^B.$$

■

Definition. A tensor product of a finite number of Jordan cells is called a *Jordan normal form.*

Lemmas 3.15 and 3.16 allow to evaluate e^{tA} when A is a Jordan normal form.

Example. Solve the system $x'=Ax$ where

$$A=\begin{pmatrix}1&1&0&0\\0&1&0&0\\0&0&2&1\\0&0&0&2\end{pmatrix}.$$

Clearly, the matrix A is the tensor product of two Jordan cells:

$$J_1=\begin{pmatrix}1&1\\0&1\end{pmatrix}\text{ and }J_2=\begin{pmatrix}2&1\\0&2\end{pmatrix}.$$

By Lemma 3.15, we obtain

$$e^{tJ_1}=\begin{pmatrix}e^t&te^t\\0&e^t\end{pmatrix}\text{ and }e^{tJ_2}=\begin{pmatrix}e^{2t}&te^{2t}\\0&e^{2t}\end{pmatrix}$$

whence by Lemma 3.16,

$$e^{tA}=\begin{pmatrix}e^t&te^t&0&0\\0&e^t&0&0\\0&0&e^{2t}&te^{2t}\\0&0&0&e^{2t}\end{pmatrix}.$$

The columns of this matrix form 4 linearly independent solutions

$$\begin{aligned}x_1&=\left(e^t,0,0,0\right)\\x_2&=\left(te^t,e^t,0,0\right)\\x_3&=\left(0,0,e^{2t},0\right)\\x_4&=\left(0,0,te^{2t},e^{2t}\right)\end{aligned}$$

and the general solution is

$$\begin{aligned}x(t)&=C_1x_1+C_2x_2+C_3x_3+C_4x_4\\&=\left(C_1e^t+C_2te^t,C_2e^t,C_3e^{2t}+C_4te^{2t},C_4e^{2t}\right).\end{aligned}$$

3.8.2 Transformation of an operator to a Jordan normal form

Given a basis $b = \{b_1, b_2, ..., b_n\}$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$) and a vector $x \in \mathbb{R}^n$ (or $\mathbb{C}^n$), denote by x_b the column vector that represents x in this basis. That is, if x_b^i is the i-th component of x_b then

$$x = x_b^1 b_1 + x_b^2 b_2 + ... + x_b^n b_n = \sum_{i=1}^{n} x_b^i b_i.$$

Similarly, if A is a linear operator in $\mathbb{R}^n$ (or $\mathbb{C}^n$) then denote by A_b the matrix that represents A in the basis b, that is, for all vectors x,

$$(Ax)_b = A_b x_b,$$

where in the right hand side we have the product of the $n \times n$ matrix A_b and the $n \times 1$ column x_b.

If $x = b_i$ then $x_b = (0, ...1, ...0)$ where 1 is at position i, and $A_b x$ is the i-th column of A_b. In other words, we have the identity

$$A_b = \left((Ab_1)_b \mid (Ab_2)_b \mid \cdots \mid (Ab_n)_b \right)$$

that can be stated as the following rule:

the i-th column of A_b is the column vector Ab_i written in the basis $b_1, ..., b_n$.

Example. Consider the operator A in $\mathbb{R}^2$ that is given in the canonical basis $e = \{e_1, e_2\}$ by the matrix

$$A_e = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Consider another basis $b = \{b_1, b_2\}$ defined by

$$b_1 = e_1 - e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } b_2 = e_1 + e_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then

$$(Ab_1)_e = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$(Ab_2)_e = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

It follows that $Ab_1 = b_2$ and $Ab_2 = b_1$ whence

$$A_b = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The following theorem is proved in Linear Algebra courses.

Theorem. *For any operator $A \in \mathcal{L}(\mathbb{C}^n)$ there is a basis b in $\mathbb{C}^n$ such that the matrix A_b is in the Jordan normal form.*

Let J be a Jordan cell of A_b with λ on the diagonal and suppose that the rows (and columns) of J in A_b are indexed by $j, j+1, ..., j+p-1$ so that J is a $p \times p$ matrix. Then the sequence of vectors $b_j, ..., b_{j+p-1}$ is referred to as the *Jordan chain* of the given Jordan cell. In particular, the basis b splits to a number of Jordan chains.

Since

$$A_b - \lambda\,\mathrm{id} = \begin{array}{cc} \begin{array}{cccccccc} & & j & \cdots & \cdots & j+p-1 & & \\ & & \downarrow & & & \downarrow & & \end{array} & \\ \left(\begin{array}{cccccccc} \ddots & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & 1 & \cdots & 0 & & \\ & & 0 & \ddots & \ddots & \vdots & & \\ & & \vdots & \ddots & \ddots & 1 & & \\ & & 0 & \cdots & 0 & 0 & & \\ & & & & & & \ddots & \\ & & & & & & & \ddots \end{array}\right) & \begin{array}{c} \\ \\ \leftarrow j \\ \cdots \\ \cdots \\ \leftarrow j+p-1 \\ \\ \\ \end{array} \end{array}$$

and the k-th column of $A_b - \lambda \operatorname{id}$ is the vector $(A - \lambda \operatorname{id})\, b_k$ written in the basis b, we conclude that

$$\begin{aligned} (A - \lambda \operatorname{id})\, b_j &= 0 \\ (A - \lambda \operatorname{id})\, b_{j+1} &= b_j \\ &\dots \\ (A - \lambda \operatorname{id})\, b_{j+p-1} &= b_{j+p-2}. \end{aligned}$$

In particular, b_j is an eigenvector of A with the eigenvalue λ. The vectors $b_{j+1}, ..., b_{j+p-1}$ are called the *generalized eigenvectors* of A (more precisely, b_{j+1} is the $1st$ generalized eigenvector, b_{j+2} is the second generalized eigenvector, etc.). Hence, any Jordan chain contains exactly one eigenvector and the rest vectors are the generalized eigenvectors.

Theorem 3.17 *Consider the system $x' = Ax$ with a constant linear operator A and let A_b be the Jordan normal form of A. Then each Jordan cell J of A_b of dimension p with λ on the diagonal gives rise to p linearly independent solutions as follows:*

$$\begin{aligned} x_1(t) &= e^{\lambda t} v_1 \\ x_2(t) &= e^{\lambda t}\left(\frac{t}{1!} v_1 + v_2\right) \\ x_3(t) &= e^{\lambda t}\left(\frac{t^2}{2!} v_1 + \frac{t}{1!} v_2 + v_3\right) \\ &\dots \\ x_p(t) &= e^{\lambda t}\left(\frac{t^{p-1}}{(p-1)!} v_1 + ... + \frac{t}{1!} v_{p-1} + v_p\right), \end{aligned}$$

where $\{v_1, ..., v_p\}$ is the Jordan chain of J. The set of all n solutions obtained across all Jordan cells is linearly independent.

Proof. In the basis b, we have by Lemmas 3.15 and 3.16

$$e^{tA_b} = \begin{pmatrix} \ddots & & & & \\ & \boxed{\ddots} & & & \\ & & \boxed{\begin{matrix} e^{\lambda t} & \frac{t}{1!}e^{t\lambda} & \cdots & \frac{t^{p-1}}{(p-1)!}e^{t\lambda} \\ 0 & e^{t\lambda} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{t}{1!}e^{t\lambda} \\ 0 & \cdots & 0 & e^{t\lambda} \end{matrix}} & & \\ & & & \boxed{\ddots} & \\ & & & & \ddots \end{pmatrix},$$

where the block in the middle is e^{tJ}. By Lemma 3.13, the columns of this matrix give n linearly independent solutions to the ODE $x' = A_b x$. Therefore, the vectors that are represented by these columns in the basis b, form n linearly independent solutions to the ODE $x' = Ax$. Out of these solutions, select p solutions that correspond to p columns of the cell e^{tJ}, that is,

$$
\begin{aligned}
x_1(t) &= (\dots \underbrace{e^{\lambda t}, 0, \dots, 0}_{p} \dots) \\
x_2(t) &= (\dots \underbrace{\tfrac{t}{1!}e^{\lambda t}, e^{\lambda t}, 0, \dots, 0}_{p} \dots) \\
&\dots \\
x_p(t) &= (\dots \underbrace{\tfrac{t^{p-1}}{(p-1)!}e^{\lambda t}, \dots, \tfrac{t}{1!}e^{\lambda t}, e^{t\lambda}}_{p} \dots),
\end{aligned}
$$

where all the vectors are written in the basis b, the horizontal braces mark the columns of the cell J, and all the terms outside the horizontal braces are zeros. Representing these vectors in the coordinateless form via the Jordan chain $v_1, ..., v_p$, we obtain the solutions as in the statement of Theorem 3.17. ■

Let λ be an eigenvalue of A. Denote by m the *algebraic multiplicity* of λ, that is, its multiplicity as a root of characteristic polynomial[8] $P(\lambda) = \det(A - \lambda \,\mathrm{id})$. Denote by g the *geometric multiplicity* of λ, that is the dimension of the eigenspace of λ:

$$g = \dim\ker(A - \lambda \,\mathrm{id}).$$

In other words, g is the maximal number of linearly independent eigenvectors of λ. The numbers m and g can be characterized in terms of the Jordan normal form A_b of A as follows: m is the total number of occurrences of λ on the diagonal[9] of A_b, whereas g is equal to the number of the Jordan cells with λ on the diagonal[10]. It follows that $g \leq m$ and the equality occurs if and only if all the Jordan cells with the eigenvalue λ have dimension 1.

Despite this relation to the Jordan normal form, m and g can be determined without a priori finding the Jordan normal form, as it is clear from the definitions of m and g.

Theorem 3.17′ *Let $\lambda \in \mathbb{C}$ be an eigenvalue of an operator A with the algebraic multiplicity m and the geometric multiplicity g. Then λ gives rise to m linearly independent solutions of the system $x' = Ax$ that can be found in the form*

$$x(t) = e^{\lambda t}\left(u_1 + u_2 t + \dots + u_s t^{s-1}\right) \tag{3.56}$$

where $s = m - g + 1$ and u_j are vectors that can be determined by substituting the above function to the equation $x' = Ax$.

The set of all n solutions obtained in this way using all the eigenvalues of A is linearly independent.

[8] To compute $P(\lambda)$, one needs to write the operator A in some basis b as a matrix A_b and then evaluate $\det(A_b - \lambda \,\mathrm{id})$. The characteristic polynomial does not depend on the choice of basis b. Indeed, if b' is another basis then the relation between the matrices A_b and $A_{b'}$ is given by $A_b = CA_{b'}C^{-1}$ where C is the matrix of transformation of basis. It follows that $A_b - \lambda \,\mathrm{id} = C(A_{b'} - \lambda \,\mathrm{id})C^{-1}$ whence $\det(A_b - \lambda \,\mathrm{id}) = \det C \det(A_{b'} - \lambda \,\mathrm{id}) \det C^{-1} = \det(A_{b'} - \lambda \,\mathrm{id})$.

[9] If λ occurs k times on the diagonal of A_b then λ is a root of multiplicity k of the characteristic polynomial of A_b that coincides with that of A. Hence, $k = m$.

[10] Note that each Jordan cell correponds to exactly one eigenvector.

Remark. For practical use, one should substitute (3.56) into the system $x' = Ax$ considering u_{ij} as unknowns (where u_{ij} is the i-th component of the vector u_j) and solve the resulting linear algebraic system with respect to u_{ij}. The result will contain m arbitrary constants, and the solution in the form (3.56) will appear as a linear combination of m independent solutions.

Proof. Let $p_1, .., p_g$ be the dimensions of all the Jordan cells with the eigenvalue λ (as we know, the number of such cells is g). Then λ occurs $p_1 + ... + p_j$ times on the diagonal of the Jordan normal form, which implies

$$\sum_{j=1}^{g} p_j = m.$$

Hence, the total number of linearly independent solutions that are given by Theorem 3.17 for the eigenvalue λ is equal to m. Let us show that each of the solutions of Theorem 3.17 has the form (3.56). Indeed, each solution of Theorem 3.17 is already in the form

$$e^{\lambda t} \text{ times a polynomial of } t \text{ of degree } \leq p_j - 1.$$

To ensure that these solutions can be represented in the form (3.56), we only need to verify that $p_j - 1 \leq s - 1$. Indeed, we have

$$\sum_{j=1}^{g} (p_j - 1) = \left(\sum_{j=1}^{g} p_j \right) - g = m - g = s - 1,$$

whence the inequality $p_j - 1 \leq s - 1$ follows. ■

In particular, if $m = g$, that is, $s = 1$, then m independent solutions can be found in the form $x(t) = e^{\lambda t} v$, where v is one of m independent eigenvectors of λ. This case has been already discussed above. Consider some examples, where $g < m$.

Example. Solve the system

$$x' = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix} x.$$

The characteristic polynomial is

$$P(\lambda) = \det(A - \lambda \operatorname{id}) = \det \begin{pmatrix} 2-\lambda & 1 \\ -1 & 4-\lambda \end{pmatrix} = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2,$$

and the only eigenvalue is $\lambda_1 = 3$ with the algebraic multiplicity $m_1 = 2$. The equation for an eigenvector v is

$$(A - \lambda \operatorname{id}) v = 0$$

that is, for $v = (a, b)$,

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

which is equivalent to $-a + b = 0$. Setting $a = 1$ and $b = 1$, we obtain the unique (up to a constant multiple) eigenvector

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Hence, the geometric multiplicity is $g_1 = 1$. Hence, there is only one Jordan cell with the eigenvalue λ_1, which allows to immediately determine the Jordan normal form of the given matrix:

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}.$$

By Theorem 3.17, we obtain the solutions

$$\begin{aligned} x_1(t) &= e^{3t} v_1 \\ x_2(t) &= e^{3t}(tv_1 + v_2) \end{aligned}$$

where v_2 is the 1st generalized eigenvector that can be determined from the equation

$$(A - \lambda\,\mathrm{id})\, v_2 = v_1.$$

Setting $v_2 = (a, b)$, we obtain the equation

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

which is equivalent to $-a + b = 1$. Hence, setting $a = 0$ and $b = 1$, we obtain

$$v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

whence

$$x_2(t) = e^{3t} \begin{pmatrix} t \\ t+1 \end{pmatrix}.$$

Finally, the general solution is

$$x(t) = C_1 x_1 + C_2 x_2 = e^{3t} \begin{pmatrix} C_1 + C_2 t \\ C_1 + C_2(t+1) \end{pmatrix}.$$

Example. Solve the system

$$x' = \begin{pmatrix} 2 & 1 & 1 \\ -2 & 0 & -1 \\ 2 & 1 & 2 \end{pmatrix} x.$$

The characteristic polynomial is

$$\begin{aligned} P(\lambda) &= \det(A - \lambda\,\mathrm{id}) - \det \begin{pmatrix} 2-\lambda & 1 & 1 \\ -2 & -\lambda & -1 \\ 2 & 1 & 2-\lambda \end{pmatrix} \\ &= -\lambda^3 + 4\lambda^2 - 5\lambda + 2 = (2-\lambda)(\lambda-1)^2. \end{aligned}$$

The roots are $\lambda_1 = 2$ with $m_1 = 1$ and $\lambda_2 = 1$ with $m_2 = 2$. The eigenvectors v for λ_1 are determined from the equation

$$(A - \lambda_1\,\mathrm{id})\, v = 0,$$

whence, for $v = (a, b, c)$

$$\begin{pmatrix} 0 & 1 & 1 \\ -2 & -2 & -1 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0,$$

that is,

$$\begin{cases} b + c = 0 \\ -2a - 2b - c = 0 \\ 2a + b = 0. \end{cases}$$

The second equation is a linear combination of the first and the last ones. Setting $a = 1$ we find $b = -2$ and $c = 2$ so that the unique (up to a constant multiple) eigenvector is

$$v = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix},$$

which gives the first solution

$$x_1(t) = e^{2t} \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}.$$

The eigenvectors for $\lambda_2 = 1$ satisfy the equation

$$(A - \lambda_2 \operatorname{id}) v = 0,$$

whence, for $v = (a, b, c)$,

$$\begin{pmatrix} 1 & 1 & 1 \\ -2 & -1 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0,$$

whence

$$\begin{cases} a + b + c = 0 \\ -2a - b - c = 0 \\ 2a + b + c = 0. \end{cases}$$

Solving the system, we obtain a unique (up to a constant multiple) solution $a = 0$, $b = 1$, $c = -1$. Hence, we obtain only one eigenvector

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Therefore, $g_2 = 1$, that is, there is only one Jordan cell with the eigenvalue λ_2, which implies that the Jordan normal form of the given matrix is as follows:

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

By Theorem 3.17, the cell with $\lambda_2 = 1$ gives rise to two more solutions

$$x_2(t) = e^t v_1 = e^t \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

and

$$x_3\left(t\right)=e^t\left(tv_1+v_2\right),$$

where v_2 is the first generalized eigenvector to be determined from the equation

$$\left(A-\lambda_2\,\mathrm{id}\right)v_2=v_1.$$

Setting $v_2=(a,b,c)$ we obtain

$$\begin{pmatrix} 1 & 1 & 1 \\ -2 & -1 & -1 \\ 2 & 1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}=\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

that is

$$\begin{cases} a+b+c=0 \\ -2a-b-c=1 \\ 2a+b+c=-1. \end{cases}$$

This system has a solution $a=-1$, $b=0$ and $c=1$. Hence,

$$v_2=\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$

and the third solution is

$$x_3\left(t\right)=e^t\left(tv_1+v_2\right)=e^t\begin{pmatrix} -1 \\ t \\ 1-t \end{pmatrix}.$$

Finally, the general solution is

$$x\left(t\right)=C_1x_1+C_2x_2+C_3x_3=\begin{pmatrix} C_1e^{2t}-C_3e^t \\ -2C_1e^{2t}+\left(C_2+C_3t\right)e^t \\ 2C_1e^{2t}+\left(C_3-C_2-C_3t\right)e^t \end{pmatrix}.$$

4 Qualitative analysis of ODEs

4.1 Autonomous systems

Consider a vector ODE

$$x' = f(x)$$

where the right hand side does not depend on t. Such equations are called *autonomous.* Here f is a C^1 function defined on an open set $\Omega \subset \mathbb{R}^n$ so that the domain of the ODE is $\mathbb{R} \times \Omega$.

Definition. The set Ω is called the *phase space* of the ODE and any path $x : (a, b) \to \Omega$ where $x(t)$ is a solution of the ODE, is called a *phase trajectory.* A plot of all phase trajectories is called a *phase diagram* or a *phase portrait.*

Recall that the graph of a solution (or the integral curve) is the set of points $(t, x(t))$ in $\mathbb{R} \times \Omega$. Hence, the phase trajectory can be regarded as the projection of the integral curve onto Ω.

For any $y \in \Omega$, denote by $\varphi(t, y)$ the maximal solution to the IVP

$$\begin{cases} x' = f(x) \\ x(0) = y. \end{cases}$$

Recall that, by Theorem 2.16, the domain of function $\varphi(t, y)$ is an open subset of $\mathbb{R}^{n+1}$ and φ belongs to C^1 in its domain. Since f does not depend on t, it follows that the solution to

$$\begin{cases} x' = f(x) \\ x(t_0) = y \end{cases}$$

is given by $x(t) = \varphi(t - t_0, y)$.

Observe that if $f(x_0) = 0$ for some $x_0 \in \Omega$ then constant function $x(t) \equiv x_0$ is a solution of $x' = f(x)$. Conversely, if $x(t) \equiv x_0$ is a solution then $f(x_0) = 0$. The constant solutions play important role in the qualitative analysis of the ODE.

Definition. If $f(x_0) = 0$ at some point $x_0 \in \Omega$ then x_0 is called a *stationary point* of the ODE $x' = f(x)$ (other terms: rest point, singular point, equilibrium point, fixed point, etc).

Observe that if x_0 is a stationary point then $\varphi(t, x_0) \equiv x_0$.

Definition. A stationary point x_0 is called *Lyapunov stable* if for any $\varepsilon > 0$ there exists $\delta > 0$ with the following property: for all $x \in \Omega$ such that $\|x - x_0\| < \delta$, the solution $\varphi(t, x)$ is defined for all $t > 0$ and

$$\sup_{t \in (0,+\infty)} \|\varphi(t, x) - x_0\| < \varepsilon. \tag{4.1}$$

In other words,

$$\sup_{t \in (0,+\infty)} \|\varphi(t, x) - x_0\| \to 0 \text{ as } x \to x_0.$$

If we replace here the interval $(0,+\infty)$ by any bounded interval $[a,b]$ containing 0 then by the continuity of $\varphi(t,x)$,

$$\sup_{t\in[a,b]}\|\varphi(t,x)-x_0\|=\sup_{t\in[a,b]}\|\varphi(t,x)-\varphi(t,x_0)\|\to 0 \text{ as } x\to x_0.$$

Hence, the main issue for the stability is the behavior of solutions as $t\to+\infty$.

Definition. A stationary point x_0 is called *asymptotically stable* if it is Lyapunov stable and

$$\|\varphi(t,x)-x_0\|\to 0 \text{ as } t\to+\infty$$

for all $x\in\Omega$ such that $\|x-x_0\|$ is small enough.

Observe, the stability and asymptotic stability do not depend on the choice of the norm in $\mathbb{R}^n$ because all norms in $\mathbb{R}^n$ are equivalent.

4.2 Stability for a linear system

Consider a linear system $x'=Ax$ in $\mathbb{R}^n$ where A is a constant operator. Clearly, $x=0$ is a stationary point.

Theorem 4.1 *If for any eigenvalue λ of A, we have $\operatorname{Re}\lambda<0$ then 0 is asymptotically stable. If for some eigenvalue λ of A, $\operatorname{Re}\lambda>0$ then 0 is unstable.*

Proof. By Theorem 3.17′, n independent solutions can be found in the form

$$x_i(t)=e^{\lambda_i t}P_i(t)$$

where λ_i are the eigenvalues, $P_i(t)$ is a vector valued polynomial of t, that is, $P_i(t)=u_1+u_2t+...+u_st^{s-1}$ for some vectors $u_1,...,u_s$. Hence, the general solution has the form

$$x(t)=\sum_{i=1}^n C_ie^{\lambda t}P_i(t).$$

Since $x(0)=\sum_{i=1}^n C_iP_i(0)$, we see that the coefficients C_i are the components of $x(0)$ in the basis $\{P_i(0)\}$.

Let now x denote the initial vector (rather than a solution) and $x_1,...,x_n$ be the components of x in this basis. Then the the solution $\varphi(t,x)$ is given by

$$\varphi(t,x)=\sum_{i=1}^n x_ie^{\lambda_i t}P_i(t).$$

It follows that

$$\begin{aligned}\|\varphi(t,x)\| &\le \sum_{i=1}^n|x_i|\left|e^{\lambda_i t}\right|\|P_i(t)\|\\ &\le \max_i\left|e^{\lambda_i t}\right|\|P_i(t)\|\sum_{i=1}^n|x_i|\\ &= \max_i e^{t\operatorname{Re}\lambda_i}\|P_i(t)\|\,\|x\|_1.\end{aligned}$$

Observe that

$$\|P_i(t)\| \leq C\left(t^N + 1\right)$$

for all $t \geq 0$ and for some C and N. Chose N and C the same for all i.

If all $\operatorname{Re}\lambda_i$ are negative then, setting

$$\alpha = \min|\operatorname{Re}\lambda_i| > 0,$$

we obtain $e^{t\operatorname{Re}\lambda_i} \leq e^{-\alpha t}$ whence

$$\|\varphi(t,x)\| \leq Ce^{-\alpha t}\left(t^N + 1\right)\|x\|$$

(where we have replaced $\|x\|_1$ by $\|x\|$ which can be done by adjusting the constant C).

Since the function $\left(t^N + 1\right)e^{-\alpha t}$ is bounded on $(0, +\infty)$ and we obtain that there is a constant C_1 such that for all $t \geq 0$

$$\|\varphi(t,x)\| \leq C_1\|x\|,$$

whence it follows that 0 is stable. Moreover, since $\left(t^N + 1\right)e^{-\alpha t} \to 0$ as $t \to +\infty$, we conclude that 0 is asymptotically stable.

Let $\operatorname{Re}\lambda > 0$ for some eigenvalue λ. To prove that 0 is unstable is suffices to show that there exists an unbounded real solution $x(t)$, that is, a solution for which $\|x(t)\|$ is not bounded on $(0, +\infty)$ as a function of t. Indeed, setting $x_0 = x(0)$ we obtain that also $\varphi(t, \varepsilon x_0) = \varepsilon x(t)$ is unbounded, for any non-zero ε. If 0 were stable this would imply that $\varphi(t,x)$ is bounded provided $\|x\|$ is small enough, which is not the case if $x = \varepsilon x_0$.

To construct an unbounded solution, consider an eigenvector v of the eigenvalue λ. It gives rise to the solution

$$x(t) = e^{\lambda t}v$$

for which

$$\|x(t)\| = \left|e^{\lambda t}\right|\|v\| = e^{t\operatorname{Re}\lambda}\|v\|.$$

Hence, $\|x(t)\|$ is unbounded. If $x(t)$ is a real solution then this finishes the proof. In general, if $x(t)$ is a complex solution then then either $\operatorname{Re}x(t)$ or $\operatorname{Im}x(t)$ is unbounded (in fact, both are), whence the instability of 0 follows. ■

This theorem does not answer the question what happens when $\operatorname{Re}\lambda = 0$. We will investigate this for $n = 2$ where we also give a more detailed description of the phase diagrams.

Consider now a linear system $x' = Ax$ in $\mathbb{R}^2$ where A is a constant operator in $\mathbb{R}^2$. Let $b = \{b_1, b_2\}$ be the Jordan basis of A so that A_b has the Jordan normal form. Consider first the case when the Jordan normal form of A has two Jordan cells, that is,

$$A_b = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Then b_1 and b_2 are the eigenvectors of the eigenvalues λ_1 and λ_2, respectively, and the general solution is

$$x(t) = C_1e^{\lambda_1 t}b_1 + C_2e^{\lambda_2 t}b_2.$$

In other words, in the basis b,

$$\varphi(t,x) = \left(e^{\lambda_1 t}x_1, e^{\lambda_2 t}x_2\right)$$

where now $x = (x_1, x_2) \in \mathbb{R}^2$ denotes the initial point rather than the solution. It follows that

$$\|\varphi(t,x)\|_1 = \left|e^{\lambda_1 t}\right| |x_1| + \left|e^{\lambda_2 t}\right| |x_2| = e^{t \operatorname{Re} \lambda_1} |x_1| + e^{t \operatorname{Re} \lambda_2} |x_2| \, .$$

The following cases take place:

1. If $\operatorname{Re} \lambda_1$ or $\operatorname{Re} \lambda_2$ is positive then $\|\varphi(t,x)\|$ goes $+\infty$ as $t \to +\infty$ for $x = b_1$ or $x = b_2$ so that 0 is unstable.

2. If both $\operatorname{Re} \lambda_1$ and $\operatorname{Re} \lambda_2$ are negative, then 0 is asymptotically stability as in Theorem 4.1 (and by Theorem 4.1).

3. If both $\operatorname{Re} \lambda_1$ and $\operatorname{Re} \lambda_2$ are non-negative then

$$\|\varphi(t,x)\|_1 \le \|x\|_1 \, ,$$

which implies that the stationary point 0 is stable (but the asymptotic stability cannot be claimed).

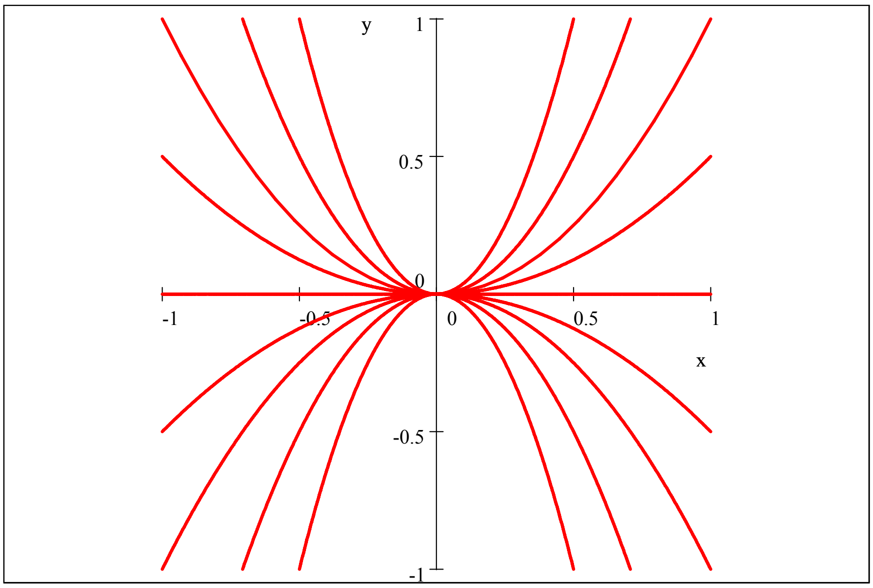
Note that the case 3 is not covered by Theorem 4.1.

Let us consider the phase diagrams of the system in various cases.

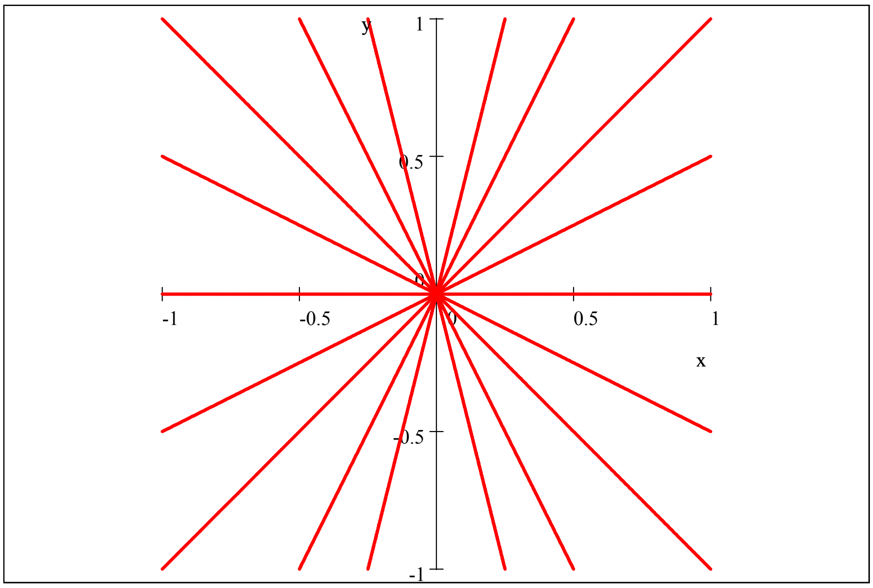
Case λ_1, λ_2 are real.

Renaming $e^{\lambda_1 t} x_1$ to x an $e^{\lambda_2 t} x_2$ to y, we obtain that the phase trajectory in the plane (x, y) satisfies the equation $y = C |x|^\gamma$ where $\gamma = \lambda_2 / \lambda_1$ (assuming that $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$). Hence, the phase diagram consists of all curves of this type as well as of the half-axis $x > 0, x < 0, y > 0, y < 0$.

If $\gamma > 0$ (that is, λ_1 and λ_2 are of the same sign) then the phase diagram (or a stationary point) is called a *node.* One distinguishes a *stable node* when $\lambda_1, \lambda_2 < 0$ and *unstable node* when $\lambda_1, \lambda_2 > 0$. Here is a node with $\gamma > 1$:

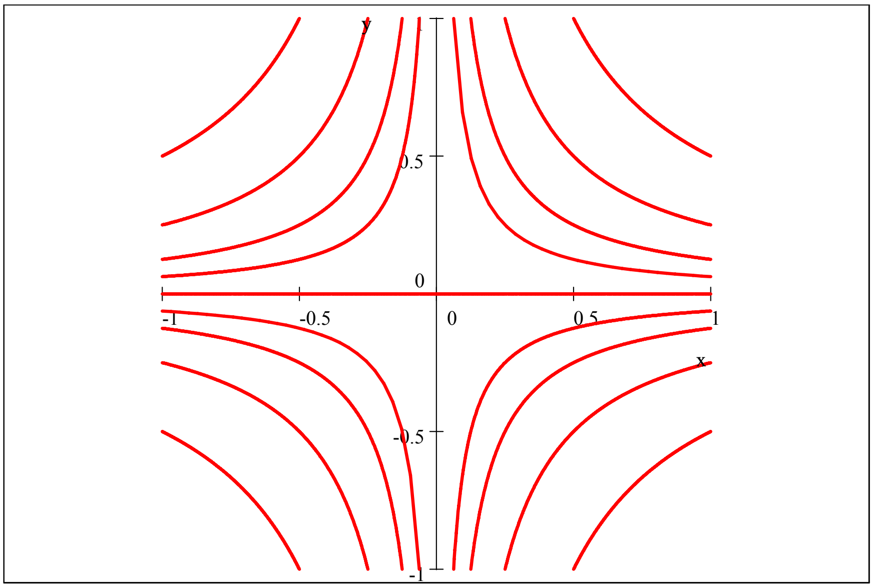


and here is a node with $\gamma = 1$:



If one or both of λ_1, λ_2 is 0 then we have a *degenerate phase diagram* (horizontal or vertical straight lines or just dots).

If $\gamma < 0$ (that is, λ_1 and λ_2 are of different signs) then the phase diagram is called a *saddle*:



Of course, the saddle is always unstable.

Case λ_1 and λ_2 are complex, say $\lambda_1 = \alpha - i\beta$ and $\lambda_2 = \alpha + i\beta$ with $\beta \neq 0$.

Then we rewrite the general solution in the real form

$$x(t) = C_1 \operatorname{Re} e^{(\alpha - i\beta)t} b_1 + C_2 \operatorname{Im} e^{(\alpha - i\beta)t} b_1.$$

Note that b_1 is an eigenvector of λ_1 and, hence, must have a non-trivial imaginary part in any real basis. We claim that in some real basis b_1 has the form $(1, i)$. Indeed, if $b_1 = (p, q)$ in the canonical basis e_1, e_2 then by rotating the basis we can assume $p, q \neq 0$. Since b_1 is an eigenvector, it is defined up to a constant multiple, so that we can take $p = 1$. Then, setting $q = q_1 + iq_2$ we obtain

$$b_1 = e_1 + (q_1 + iq_2) e_2 = (e_1 + q_1 e_2) + iq_2 e_2 = e_1' + ie_2'$$

where $e_1' = e_1 + q_1 e_2$ and $e_2' = q_2 e_2$ is a new basis (the latter follows from the fact that q is imaginary and, hence, $q_2 \neq 0$). Hence, in the basis $e' = \{e_1', e_2'\}$ we have $b_1 = (1, i)$.

It follows that in the basis e'

$$e^{(\alpha+\beta i)t} b_1 = e^{\alpha t} (\cos \beta t + i \sin \beta t) \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} e^{\alpha t} \cos \beta t - ie^{\alpha t} \sin \beta t \\ e^{\alpha t} \sin \beta t + ie^{\alpha t} \cos \beta t \end{pmatrix}$$

and

$$x(t) = C_1 \begin{pmatrix} e^{\alpha t} \cos \beta t \\ e^{\alpha t} \sin \beta t \end{pmatrix} + C_2 \begin{pmatrix} -e^{\alpha t} \sin \beta t \\ e^{\alpha t} \cos \beta t \end{pmatrix} = C \begin{pmatrix} e^{\alpha t} \cos(\beta t + \psi) \\ e^{\alpha t} \sin(\beta t + \psi) \end{pmatrix},$$

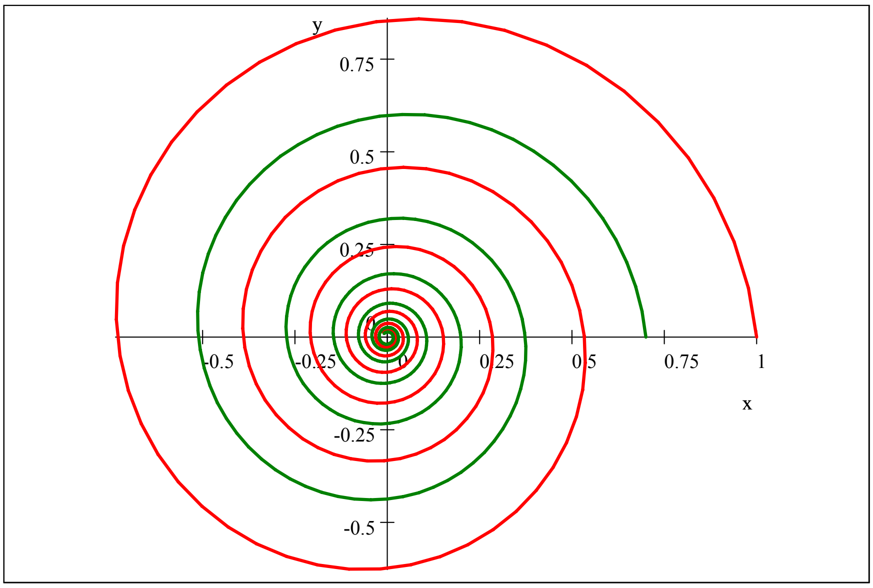
where $C = \sqrt{C_1^2 + C_2^2}$ and

$$\cos \psi = \frac{C_1}{C}, \ \sin \psi = \frac{C_2}{C}.$$

If (r, θ) are the polar coordinates on the plane in the basis e', then the polar coordinates for the solution $x(t)$ are

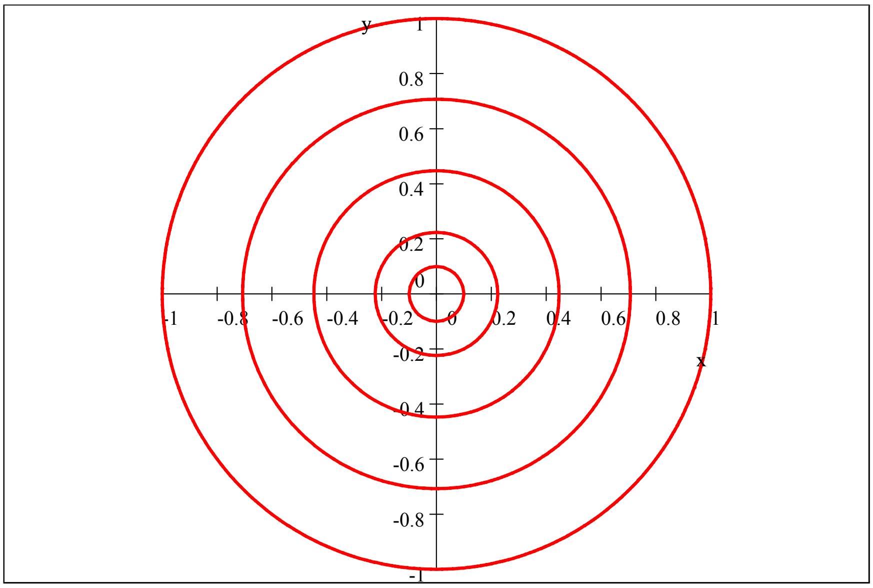
$$r(t) = Ce^{\alpha t} \text{ and } \theta(t) = \beta t + \psi.$$

If $\alpha \neq 0$ then these equations define a *logarithmic spiral,* and the phase diagram is called a *focus* or a *spiral*:



The focus is stable is $\alpha < 0$ and unstable if $\alpha > 0$.

If $\alpha = 0$ (that is, the both eigenvalues λ_1 and λ_2 are purely imaginary), then $r(t) = C$, that is, we get a family of concentric circles around 0, and this phase diagram is called a *center:*



In this case, the stationary point is stable but not asymptotically stable.

Consider now the case when the Jordan normal form of A has only one Jordan cell, that is,

$$A_b = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

In this case, λ must be real because if λ is an imaginary root of a characteristic polynomial then $\overline{\lambda}$ must also be a root, which is not possible since $\overline{\lambda}$ does not occur on the diagonal of A_b. Then the general solution is

$$x(t) = C_1 e^{\lambda t} b_1 + C_2 e^{\lambda t} (b_1 t + b_2) = (C_1 + C_2 t) e^{\lambda t} b_1 + C_2 e^{\lambda t} b_2$$

whence $x(0) = C_1 b_1 + C_2 b_2$. Renaming by $x = (x_1, x_2)$ the initial point, we obtain in the basis b

$$\varphi(t, x) = \left(e^{\lambda t}(x_1 + x_2 t), e^{\lambda t} x_2\right)$$

whence

$$\|\varphi(t, x)\|_1 = e^{\lambda t} |x_1 + x_2 t| + e^{\lambda t} |x_2| .$$

Hence, we obtain the following cases of stability:

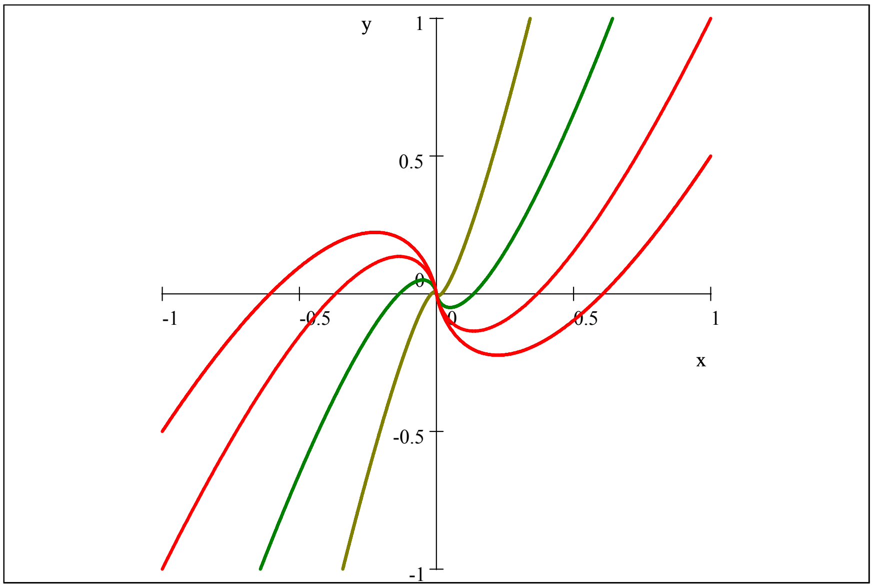
1. If $\lambda < 0$ then the stationary point 0 is asymptotically stable (which follows also from Theorem 4.1).

2. If $\lambda \geq 0$ then the stationary point 0 is unstable (indeed, if $x_2 \neq 0$ then the solution is unbounded).

Renaming $e^{\lambda t}(x_1 + x_2 t)$ by y and $e^{\lambda t} x_2$ by x, we obtain the following relation between x and y:

$$y = \frac{x \ln |x|}{\lambda} + Cx$$

(this follows from $\frac{y}{x} = \frac{x_1}{x_2} + t$ and $t = \frac{1}{\lambda} \ln \frac{x}{x_2}$). Here is the phase diagram in this case:



This phase diagram is also called a node. It is stable if $\lambda < 0$ and unstable if $\lambda > 0$. If $\lambda = 0$ then we obtain a degenerate phase diagram - parallel straight lines.

Hence, the main types of the phases diagrams are the *node* (λ_1, λ_2 are real, non-zero and of the same sign), the *saddle* (λ_1, λ_2 are real, non-zero and of opposite signs), *focus/spiral* (λ_1, λ_2 are imaginary and $\operatorname{Re} \lambda \neq 0$) and *center* ($\lambda_1, \lambda_2$ are purely imaginary). Otherwise, the phase diagram consists of parallel straight lines or just dots, and is referred to as degenerate.

To summarize the stability investigation, let us emphasize that in the case $\operatorname{Re}\lambda = 0$ both stability and instability can happen, depending on the structure of the Jordan normal form.

4.3 Lyapunov's theorem

Theorem 4.2 *Let x_0 be a stationary point of the system $x' = f(x)$ where $f \in C^2(\Omega)$. Let $A = f'(x_0)$, that is, A is the Jacobian matrix of f at x_0. If $\operatorname{Re}\lambda < 0$ for any eigenvalue λ of A then the stationary point x_0 is asymptotically stable for $x' = f(x)$.*

Remark. This theorem has the second part that says the following: if $\operatorname{Re}\lambda > 0$ for some eigenvalue λ of A then x_0 is unstable for $x' = f(x)$. The proof is somewhat lengthy and will not be presented here.

Comparing with Theorem 4.1, we see that the conditions for the stability of the stationary point x_0 for the system $x' = f(x)$ coincide with those for the *linearized system* $y' = Ay$ (provided either $\operatorname{Re}\lambda < 0$ for all eigenvalues λ or $\operatorname{Re}\lambda > 0$ for some eigenvalue λ). Setting $y = x - x_0$, we obtain that the system $x' = f(x)$ transforms to

$$y' = f(x) = f(x_0 + y) = f(x_0) + f'(x_0)\, y + o(\|y\|)$$

that is,

$$y' = Ay + o(\|y\|).$$

Hence, the linearized system $y' = Ax$ is obtained by neglecting the term $o(\|y\|)$ which is small provided $\|y\|$ is small. The message is that by throwing away this term we do not change the type of stability of the stationary point (under the above conditions for the eigenvalues). Note also that the equation $y' = Ay$ is the variational equation for $x' = f(x)$ at the solution $x \equiv x_0$.

Example. Consider the system

$$\begin{cases} x' = \sqrt{4+4y} - 2e^{x+y} \\ y' = \sin 3x + \ln(1-4y). \end{cases}$$

It is easy to see that the right hand side vanishes at $(0,0)$ so that $(0,0)$ is a stationary point. Setting

$$f(x,y) = \begin{pmatrix} \sqrt{4+4y} - 2e^{x+y} \\ \sin 3x + \ln(1-4y) \end{pmatrix},$$

we obtain

$$A = f'(0,0) = \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ \partial_x f_2 & \partial_y f_2 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ 3 & -4 \end{pmatrix}.$$

Another way to obtain this matrix is to expand each component of $f(x,y)$ by the Taylor formula:

$$\begin{aligned} f_1(x,y) &= 2\sqrt{1+y} - 2e^{x+y} = 2\left(1 + \frac{y}{2} + o(x)\right) - 2\left(1 + (x+y) + o(|x| + |y|)\right) \\ &= -2x - y + o(|x| + |y|) \end{aligned}$$

and

$$\begin{aligned} f_2(x,y) &= \sin 3x + \ln(1-4y) = 3x + o(x) - 4y + o(y) \\ &= 3x - 4y + o(|x|+|y|). \end{aligned}$$

Hence,

$$f(x,y) = \begin{pmatrix} -2 & -1 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + o(|x|+|y|),$$

whence we obtain the same matrix A.

The characteristic polynomial of A is

$$\det \begin{pmatrix} -2-\lambda & -1 \\ 3 & -4-\lambda \end{pmatrix} = \lambda^2 + 6\lambda + 11,$$

and the eigenvalues are

$$\lambda_{1,2} = -3 \pm i\sqrt{2}.$$

Hence, $\operatorname{Re}\lambda < 0$ for all λ, whence we conclude that 0 is asymptotically stable.

The main tool for the proof of theorem 4.2 is the following lemma, that is of its own interest. Recall that given a vector $v \in \mathbb{R}^n$ and a differentiable function F in a domain in $\mathbb{R}^n$, the directional derivative $\partial_v F$ can be determined by

$$\partial_v F(x) = F'(x)\,v = \sum_{i=1}^n \partial_i F(x)\, v_i.$$

Lemma 4.3 *Consider the system $x' = f(x)$ where $f \in C^1(\Omega)$ and let x_0 be a stationary point of it. Let $V(x)$ be a C^1 scalar function in an open set U such that $x_0 \in U \subset \Omega$ and the following conditions hold:*

1. *$V(x) > 0$ for any $x \in U \setminus \{x_0\}$ and $V(x_0) = 0$.*
2. *For all $x \in U$,*
$$\partial_{f(x)} V(x) \leq 0. \tag{4.2}$$

Then the stationary point 0 is stable.

Furthermore, if all $x \in U$

$$\partial_{f(x)} V(x) \leq -W(x), \tag{4.3}$$

where $W(x)$ is a continuous function on U such that $W(x) > 0$ for $x \in U \setminus \{x_0\}$ and $W(x_0) = 0$ then the stationary point 0 is asymptotically stable.

Function V with the properties 1-2 is called the *Lyapunov function.* Note that in the expression $\partial_{f(x)} V(x)$ the vector field $f(x)$ which is used for the directional derivative of V, depends on x. By definition, we have

$$\partial_{f(x)} V(x) = \sum_{i=1}^n \partial_i V(x)\, f_i(x).$$

In this context, $\partial_f V$ is also called the *orbital derivative* of V with respect to the ODE $x' = f(x)$.

Before the proof, let us show examples of the Lyapunov functions.

Example. Consider the system $x' = Ax$ where matrix A has the diagonal form $A = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ where λ_i are all real. Obviously, 0 is a stationary point. Consider the function

$$V(x) = \sum_{i=1}^{n} x_i^2 = \|x\|_2^2,$$

which is positive in $\mathbb{R}^n \setminus \{0\}$ and vanishes at 0. Then $\partial_i V = 2x_i$, $f_i(x) = \lambda_i x_i$ whence

$$\partial_f V = \sum_{i=1}^{n} 2\lambda_i x_i^2.$$

If all λ_i are non-positive then $\partial_f V \leq 0$ so that V satisfies (4.2). If all λ_i are negative then set

$$\eta = 2 \min_i |\lambda_i| > 0.$$

It follows that

$$\partial_f V \leq -\eta \sum_{i=1}^{n} x_i^2 = -\eta V,$$

so that the condition (4.3) is satisfied. Therefore, if all $\lambda_i \leq 0$ then 0 is stable and if all $\lambda_i < 0$ then 0 is asymptotically stable. Of course, in this example this can be seen directly from the formula for the general solution.

Example. Consider the second order scalar ODE $x'' + kx' = F(x)$ which describes the movement of a body under the external potential force $F(x)$ and friction with the coefficient k. This can be written as a system

$$\begin{cases} x' = y \\ y' = -ky + F(x). \end{cases}$$

Note that the phase space is $\mathbb{R}^2$ (assuming that F is defined on $\mathbb{R}$) and a point (x, y) in the phase space is a couple position-velocity.

Assume $F(0) = 0$ so that $(0, 0)$ is a stationary point. We would like to answer the question if $(0, 0)$ is stable or not. The Lyapunov function can be constructed in this case as the full energy

$$V(x, y) = \frac{y^2}{2} + U(x),$$

where $U(x) = -\int F(x)\,dx$ is the potential energy and $\frac{y^2}{2}$ is the kinetic energy. More precisely, assume that $k \geq 0$, $F(x) < 0$ for $x > 0$, $F(x) > 0$ for $x < 0$ and set

$$U(x) = -\int_0^x F(s)\,ds,$$

so that $U(0) = 0$ and $U(x) > 0$ for $x \neq 0$. Then the function $V(x, y)$ is positive away from $(0, 0)$ and vanishes at $(0, 0)$. Let us compute the orbital derivative of V setting $f(x, y) = (y, F(x))$:

$$\begin{aligned} \partial_f V &= y\partial_x V + (-ky + F(x))\,\partial_y V = yU'(x) + (-ky + F(x))\,y \\ &= -yF(x) - ky^2 + F(x)\,y = -ky^2 \leq 0. \end{aligned}$$

Hence, V is indeed the Lyapunov function, and by Lemma 4.3 the stationary point $(0,0)$ is stable.

Physically this has a simple meaning. The fact that $F(x) < 0$ for $x > 0$ and $F(x) > 0$ for $x < 0$ means that the force always acts in the direction of the origin thus trying to return the displaced body to the stationary point, which causes the stability.

Proof of Lemma 4.3. For any solution $x(t)$ in U, we have by the chain rule

$$\frac{d}{dt}V(x(t)) = V'(x)x'(t) = V'(x)f(x) = \partial_{f(x)}V(x) \leq 0. \tag{4.4}$$

Therefore, the function V is decreasing along any solution $x(t)$ as long as $x(t)$ remains inside U.

By shrinking U, we can assume that U is bounded and that V is defined on $\overline{U}$. Also, without loss of generality, assume that x_0 is the origin of $\mathbb{R}^n$. Set

$$B_r = B(0,r) = \{x \in \mathbb{R}^n : \|x\| < r\}.$$

Since U is open and contains 0 there is $\varepsilon_0 > 0$ such that $B_{\varepsilon_0} \subset U$. For any $\varepsilon \in (0, \varepsilon_0)$, set

$$m(\varepsilon) = \inf_{x \in \overline{U} \setminus B_\varepsilon} V(x).$$

Since V is continuous and $\overline{U} \setminus B_\varepsilon$ is a compact set (bounded and closed), by the minimal value theorem, the infimum of V is taken at some point. Since V is positive away from 0, we obtain $m(\varepsilon) > 0$. It follows from the definition of $m(\varepsilon)$ that $V(x) \geq m(\varepsilon)$ outside B_ε. In particular, if $x \in U$ and $V(x) < m(\varepsilon)$ then $x \in B_\varepsilon$.

Now given $\varepsilon > 0$ we need to find $\delta > 0$ such that $x \in B_\delta$ implies $\varphi(t,x) \in B_\varepsilon$ for all $t \geq 0$ (where $\varphi(t,x)$ is the maximal solution to the given ODE with the initial value x at $t = 0$). First of all, we can assume that $\varepsilon < \varepsilon_0$. By the continuity of V, δ can be chosen so small that $V(x) < m(\varepsilon)$ for all $x \in B_\delta$. Then the solution $\varphi(t,x)$ for $t > 0$ must also satisfy the condition $V(\varphi(t,x)) < m(\varepsilon)$ and hence, $\varphi(t,x) \in B_\varepsilon$, as long as $\varphi(t,x)$ is defined. Shortly, we have shown the following implications:

$$x \in B_\delta \;\Rightarrow\; V(x) < m(\varepsilon) \;\Rightarrow\; V(\varphi(t,x)) < m(\varepsilon) \;\Rightarrow\; \varphi(t,x) \in B_\varepsilon.$$

We are left to verify that $\varphi(t,x)$ is defined for all $t > 0$ and $\varphi(t,x) \in U$. Indeed, assume that $\varphi(t,x)$ is defined only for $t < T$ where T is finite. Then the graph of the solution $(t, \varphi(t,x))$ is located in the set $[0,T] \times \overline{B(x_0,\varepsilon)}$, which is compact, whereas $\varphi(t,x)$ is a maximal solution that must leave any compact in $\mathbb{R} \times U$ when $t \to T$ (see Theorem 2.8). Hence, T must be $+\infty$, which finishes the proof of the first part.

For the second part, we obtain by (4.3) and (4.4)

$$\frac{d}{dt}V(x(t)) \leq -W(x(t)).$$

It suffices to show that

$$V(x(t)) \to 0 \text{ as } t \to \infty$$

since this will imply that $x(t) \to 0$ (recall that 0 is the only point where V vanishes). Since $V(x(t))$ is decreasing in t, the limit

$$L = \lim_{t \to +\infty} V(x(t))$$

exists. Assume that $L > 0$. Then, for all $t > 0$, $V(x(t)) \geq L$. By the continuity of V, there is $r > 0$ such that

$$V(y) < L \text{ for all } y \in B_r.$$

Hence, $x(t) \notin B_r$ for all $t > 0$. Set

$$m = \inf_{y \in \overline{U} \setminus B_r} W(y) > 0.$$

It follows that

$$\frac{d}{dt} V(x(t)) \leq -W(x(t)) \leq -m$$

for all $t > 0$. However, this implies that

$$V(x(t)) \leq V(x(0)) - mt$$

which becomes negative for large enough t. This contradiction proves that $L = 0$ and, hence, $x(t) \to 0$ as $t \to +\infty$. ■

Proof of Theorem 4.2. Without loss of generality, set $x_0 = 0$. Using that $f \in C^2$, we obtain by the Taylor formula, for any component f_k of f,

$$f_k(x) = f_k(0) + \sum_{i=1}^{n} \partial_i f_k(0) x_i + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} f_k(0) x_i x_j + o\left(\|x\|^2\right) \text{ as } x \to 0.$$

Noticing that $\partial_i f_k(0) = A_{ki}$ write

$$f(x) = Ax + h(x)$$

where $h(x)$ is defined by

$$h_k(x) = \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} f_k(0) x_i x_j + o\left(\|x\|^2\right).$$

Setting $B = \max_{i,j,k} |\partial_{ij} f_k(0)|$, we obtain

$$\|h(x)\|_\infty = \max_{1 \leq k \leq n} |h_k(x)| \leq B \sum_{i,j=1}^{n} |x_i x_j| + o\left(\|x\|^2\right) = B \|x\|_1^2 + o\left(\|x\|^2\right).$$

Hence, for any choice of the norms, there is a constant C such that

$$\|h(x)\| \leq C \|x\|^2$$

provided $\|x\|$ is small enough.

Assuming that $\operatorname{Re} \lambda < 0$ for all eigenvalues of A, consider the following function

$$V(x) = \int_0^\infty \left\|e^{sA} x\right\|_2^2 ds$$

and prove that $V(x)$ is the Lyapunov function.

Let us first verify that $V(x)$ is finite. Indeed, in the proof of Theorem 4.1 we have established the inequality

$$\left\|e^{tA} x\right\| \leq C e^{-\alpha t} \left(t^N + 1\right) \|x\|,$$

where N is some natural number (depending on the dimensions of the cells in the Jordan normal form of A) and $\alpha > 0$ is the minimum of $-\operatorname{Re}\lambda$ over all eigenvalues λ of A. This inequality clearly implies that the integral in the definition of V is finite.

Next, let us show that $V(x)$ is of the class C^1 (in fact, C^∞). For that, represent x in the canonical basis $e_1, ..., e_n$ as $x = \sum x_i e_i$ and notice that

$$\|x\|_2^2 = \sum_{i=1}^{n} |x_i|^2 = x \cdot x.$$

Therefore,

$$\begin{aligned}\left\|e^{sA}x\right\|_2^2 &= e^{sA}x \cdot e^{sA}x = \left(\sum_i x_i \left(e^{sA}e_i\right)\right) \cdot \left(\sum_j x_j \left(e^{sA}e_j\right)\right) \\ &= \sum_{i,j} x_i x_j \left(e^{sA}e_i \cdot e^{sA}e_j\right).\end{aligned}$$

Integrating in s, we obtain

$$V(x) = \sum_{i,j} b_{ij} x_i x_j$$

with some constants b_{ij}, which clearly implies $V(x)$ is of the class C^∞.

Remark. Usually we work with any norm in $\mathbb{R}^n$. In this case we have selected the 2-norm to ensure the smoothness of $V(x)$.

Function $V(x)$ is obviously non-negative and $V(x) = 0$ if and only if $x = 0$. In order to complete the proof of the fact that $V(x)$ is the Lyapunov function, we need to estimate $\partial_{f(x)}V(x)$. Let us first evaluate $\partial_{Ax}V(x)$. Recall that by (4.4)

$$\frac{d}{dt}V\left(e^{tA}x\right) = \partial_{Ax}V\left(e^{tA}x\right)$$

whence

$$\partial_{Ax}V(x) = \left.\frac{d}{dt}V\left(e^{tA}x\right)\right|_{t=0}.$$

On the other hand,

$$V\left(e^{tA}x\right) = \int_0^\infty \left\|e^{(s+t)A}x\right\|_2^2 ds = \int_t^\infty \left\|e^{sA}x\right\|_2^2 ds$$

whence

$$\frac{d}{dt}V\left(e^{tA}x\right) = -\left\|e^{tA}x\right\|_2^2.$$

It follows that

$$\partial_{Ax}V(x) = \left.\frac{d}{dt}V\left(e^{tA}x\right)\right|_{t=0} = -\|x\|_2^2.$$

Now we can estimate $\partial_{f(x)}V(x)$ as follows:

$$\begin{aligned}\partial_{f(x)}V(x) &= \partial_{Ax}V(x) + \partial_{h(x)}V(x) = -\|x\|_2^2 + \sum_{i=1}^{n} \partial_i V(x)\, h_i(x) \\ &\leq -\|x\|_2^2 + \|V'(x)\|_2 \|h(x)\|_2 \\ &\leq -\|x\|_2^2 + C\|V'(x)\|_2 \|x\|_2^2,\end{aligned}$$

where in the second line we have used the Cauchy-Schwarz inequality

$$x \cdot y \leq \|x\|_2 \|y\|_2$$

and in the third line - the estimate $\|h(x)\|_2 \leq C\|x\|_2$ which is true provided $\|x\|$ is small enough. Since the function $V(x)$ has minimum at 0, we have $V'(0) = 0$. Hence, if $\|x\|$ is small enough then the above estimate of $\|h(x)\|$ holds and $\|V'(x)\|_2 < \frac{1}{2}C^{-1}$. It follows that, for such x,

$$\partial_{f(x)} V(x) \leq -\frac{1}{2} \|x\|_2^2 ,$$

and we conclude by Lemma 4.3, that the stationary point 0 is asymptotically stable. ■

4.4 Zeros of solutions

In this section, we consider a scalar linear second order ODE

$$x'' + p(t) x' + q(t) x = 0, \tag{4.5}$$

where $p(t)$ and $q(t)$ are continuous functions on some interval $I \subset \mathbb{R}$. We will be concerned with the structure of *zeros* of a solution $x(t)$, that is, with the points t where $x(t) = 0$.

For example, the ODE $x'' + x = 0$ has solutions $\sin t$ and $\cos t$ that have infinitely many zeros, while a similar ODE $x'' + x = 0$ has solutions $\sinh t$ and $\cosh t$ with finitely many zeros (in fact, any solution to the latter equation may have at most 1 zero). An interesting question is how to determine or to estimate the number of roots of (4.5) in general.

Let us start with the following simple observation.

Lemma 4.4 *If $x(t)$ is a solution to* (4.5) *on I that is not identical zero then, on any bounded closed interval $J \subset I$, the function $x(t)$ has at most finitely many distinct zeros. Moreover, every zero of $x(t)$ is simple.*

A zero t_0 of $x(t)$ is called *simple* if $x'(t_0) \neq 0$ and *multiple* if $x'(t_0) = 0$. This definition matches the notion of simple and multiple roots of polynomials. Note that if t_0 is a simple zero then $x(t)$ changes signed at t_0.

Proof. If t_0 is a multiple zero then then $x(t)$ solves the IVP

$$\begin{cases} x'' + px' + qx = 0 \\ x(t_0) = 0 \\ x'(t_0) = 0 \end{cases},$$

whence, by the uniqueness theorem, we conclude that $x(t) \equiv 0$.

Let $x(t)$ have infinitely many distinct zeros on J, say $x(t_k) = 0$ where $\{t_k\}_{k=1}^{\infty}$ is a sequence of distinct reals in J. Then, by the Weierstrass theorem, the sequence $\{t_k\}$ contains a convergent subsequence. Without loss of generality, we can assume that $t_k \to t_0 \in J$. Then $x(t_0) = 0$ but also $x'(t_0) = 0$, which follows from

$$x'(t_0) = \lim_{k\to\infty} \frac{x(t_k) - x(t_0)}{t_k - t_0} = 0.$$

Hence, the zero t_0 is multiple, whence $x(t) \equiv 0$. ■

Theorem 4.5 (Theorem of Sturm). *Consider two ODEs on an interval $I \subset \mathbb{R}$*

$$x'' + p(t)\,x' + q_1(t)\,x = 0 \text{ and } y'' + p(t)\,y' + q_2(t)\,y = 0,$$

where $p \in C^1(I)$, $q_1, q_2 \in C(I)$, and, for all $t \in I$,

$$q_1(t) \le q_2(t).$$

If $x(t)$ is a non-zero solution of the first ODE and $y(t)$ is a solution of the second ODE then between any two distinct zeros of $x(t)$ there is a zero of $y(t)$ (that is, if $a < b$ are zeros of $x(t)$ then there is a zero of $y(t)$ in $[a,b]$).

A mnemonic rule: the larger $q(t)$ the more likely a solution has zeros.

Example. Let q_1 and q_2 be positive constants and $p = 0$. Then the solutions are

$$x(t) = C_1 \sin\left(\sqrt{q_1}t + \varphi_1\right) \text{ and } y(t) = C_2 \sin\left(\sqrt{q_2}t + \varphi_2\right).$$

Zeros of function $x(t)$ form an arithmetic sequence with the difference $\frac{\pi}{\sqrt{q_1}}$, and zeros of $y(t)$ for an arithmetic sequence with the difference $\frac{\pi}{\sqrt{q_2}} \le \frac{\pi}{\sqrt{q_1}}$. Clearly, between any two terms of the first sequence there is a term of the second sequence.

Example. Let $q_1(t) = q_2(t) = q(t)$ and let x and y be linearly independent solution to the same ODE $x'' + px' + qx = 0$. Then we claim that if $a < b$ are consecutive zeros of $x(t)$ then there is exactly one zero of y in $[a,b]$ and this zero belongs to (a,b). Indeed, by Theorem 4.5, y has zero in $[a,b]$, say $y(c) = 0$. Let us verify that $c \neq a, b$. Assuming that $c = a$ and, hence, $y(a) = 0$, we obtain that y solves the IVP

$$\begin{cases} y'' + py' + qy = 0 \\ y(a) = 0 \\ y'(a) = Cx'(a) \end{cases}$$

where $C = \frac{y'(a)}{x'(a)}$ (note that $x'(a) \neq 0$ by Lemma 4.4). Since $Cx(t)$ solves the same IVP, we conclude by the uniqueness theorem that $y(t) \equiv Cx(t)$. However, this contradicts to the hypothesis that x and y are linearly independent. Finally, let us show that $y(t)$ has a unique root in $[a, b]$. Indeed, if $c < d$ are two zeros of y in $[a, b]$ then switching x and y in the previous argument, we conclude that x has a zero in $(c, d) \subset (a, b)$, which is not possible.

It follows that if $\{a_k\}_{k=1}^N$ is an increasing sequence of consecutive zeros of $x(t)$ then in any interval (a_k, a_{k+1}) there is exactly one root c_k of y so that the roots of x and y intertwine. An obvious example for this is the case when $x(t) = \sin t$ and $y(t) = \cos t$.

Proof of Theorem 4.5. By Exercise 37, the ODE

$$x'' + p(t)x' + q(t)x = 0$$

transforms to

$$u'' + Q(t)u = 0.$$

by the change

$$u(t) = x(t)\exp\left(\frac{1}{2}\int p(t)\,dt\right)$$

where

$$Q(t) = q - \frac{p^2}{4} - \frac{p'}{2}$$

(here we use the hypothesis that $p \in C^1$). Obviously, the zeros of $x(t)$ and $u(t)$ are the same. Also, if $q_1 \leq q_2$ then also $Q_1 \leq Q_2$. Therefore, it suffices to consider the case $p \equiv 0$.

Assume in the sequel that $p \equiv 0$. Since the set of zeros of $x(t)$ on any bounded closed interval is finite, it suffices to show that function $y(t)$ has a zero between any two consecutive zeros of $x(t)$. Let $a < b$ be two consecutive zeros of $x(t)$ so that $x(t) \neq 0$ in (a, b). Without loss of generality, we can assume that $x(t) > 0$ in (a, b). This implies that $x'(a) > 0$ and $x'(b) < 0$. Indeed, $x(t) > 0$ in (a, b) implies

$$x'(a) = \lim_{t\to a, t>a} \frac{x(t) - x(a)}{t - a} \geq 0.$$

It follows that $x'(a) > 0$ because if $x'(a) = 0$ then a is a multiple root, which is prohibited by Lemma 4.4. In the same way, $x'(b) < 0$. If $y(t)$ does not vanish in $[a, b]$ then we can assume that $y(t) > 0$ on $[a, b]$. Let us show that these assumptions lead to a contradiction.

Multiplying the equation $x'' + q_1 x = 0$ by y, the equation $y'' + q_2 y = 0$ by x, and subtracting one from the other, we obtain

$$(x'' + q_1(t)x)y - (y'' + q_2(t)y)x = 0,$$

$$x''y - y''x = (q_2 - q_1)xy,$$

whence

$$(x'y - y'x)' = (q_2 - q_1)xy.$$

Integrating the above identity from a to b and using $x(a) = x(b) = 0$, we obtain

$$x'(b)y(b) - x'(a)y(a) = [x'y - y'x]_a^b = \int_a^b (q_2(t) - q_1(t))\,x(t)\,y(t)\,dt. \tag{4.6}$$

Since $q_2 \geq q_1$ on $[a, b]$ and $x(t)$ and $y(t)$ are non-negative on $[a, b]$, the integral in (4.6) is non-negative. On the other hand, the left hand side of (4.6) is negative because $y(a)$ and $y(b)$ are positive whereas $x'(b)$ and $-x'(a)$ are negative. This contradiction finishes the proof. ■

Consider the differential operator

$$L = \frac{d^2}{dt^2} + p(t) \frac{d}{dt} + q(t) \tag{4.7}$$

so that the ODE (4.5) can be shortly written as $Lx = 0$. Assume in the sequel that $p \in C^1(I)$ and $q \in C(I)$ for some interval I.

Definition. Any C^2 function y satisfying $Ly \leq 0$ is called a *supersolution* of the operator L (or of the ODE $Lx = 0$).

Corollary. *If L has a positive supersolution $y(t)$ on an interval I then any non-zero solution $x(t)$ of $Lx = 0$ has at most one zero on I.*

Proof. Indeed, define function $\widetilde{q}(t)$ by the equation

$$y'' + p(t) y' + \widetilde{q}(t) y = 0.$$

Comparing with

$$Ly = y'' + p(t) y' + q(t) y \leq 0,$$

we conclude that $\widetilde{q}(t) \geq q(t)$. Since $x'' + px' + qx = 0$, we obtain by Theorem 4.5 that between any two distinct zeros of $x(t)$ there must be a zero of $y(t)$. Since $y(t)$ has no zeros, $x(t)$ cannot have two distinct zeros. ■

Example. If $q(t) \leq 0$ on some interval I then function $y(t) \equiv 1$ is obviously a positive supersolution. Hence, any non-zero solution of $x'' + q(t) x = 0$ has at most one zero on I. It follows that, for any solution of the IVP,

$$\begin{cases} x'' + q(t) x = 0 \\ x(t_0) = 0 \\ x'(t_0) = a \end{cases}$$

with $q(t) \leq 0$ and $a \neq 0$, we have $x(t) \neq 0$ for all $t \neq t_0$. In particular, if $a > 0$ then $x(t) > 0$ for all $t > t_0$.

Corollary. *(The comparison principle) Assume that the operator L has a positive supersolution y on an interval $[a, b]$. If $x_1(t)$ and $x_2(t)$ are two C^2 functions on $[a, b]$ such that $Lx_1 = Lx_2$ and $x_1(t) \leq x_2(t)$ for $t = a$ and $t = b$ then $x_1(t) \leq x_2(t)$ holds for all $t \in [a, b]$.*

Proof. Setting $x = x_2 - x_1$, we obtain that $Lx = 0$ and $x(t) \geq 0$ at $t = a$ and $t = b$. That is, $x(t)$ is a solution that has non-negative values at the endpoints a and b. We need to prove that $x(t) \geq 0$ inside $[a, b]$ as well. Indeed, assume that $x(c) < 0$ at some point $c \in (a, b)$. Then, by the intermediate value theorem, $x(t)$ has zeros on each interval $[a, c)$ and $(c, b]$. However, since L has a positive supersolution on $[a, b]$, $x(t)$ cannot have two zeros on $[a, b]$ by the previous corollary. ■

Consider the following *boundary value problem* (BVP) for the operator (4.7):

$$\begin{cases} Lx = f(t) \\ x(a) = \alpha \\ x(b) = \beta \end{cases}$$

where $f(t)$ is a given function on I, a, b are two given distinct points in I and α, β are given reals. It follows from the comparison principle that if L has a positive supersolution on $[a, b]$ then solution to the BVP is unique. Indeed, if x_1 and x_2 are two solutions then the comparison principle yields $x_1 \leq x_2$ and $x_2 \leq x_1$ whence $x_1 \equiv x_2$.

The hypothesis that L has a positive supersolution is essential since in general there is no uniqueness: the BVP $x'' + x = 0$ with $x(0) = x(\pi) = 0$ has a whole family of solutions $x(t) = C \sin t$ for any real C.

Let us return to the study of the cases with "many" zeros.

Theorem 4.6 *Consider ODE $x'' + q(t)x = 0$ where $q(t) \geq a > 0$ on $[t_0, +\infty)$. Then zeros of any non-zero solution $x(t)$ on $[t_0, +\infty)$ form a sequence $\{t_k\}_{k=1}^{\infty}$ that can be numbered so that $t_{k+1} > t_k$, and $t_k \to +\infty$. Furthermore, if*

$$\lim_{t \to +\infty} q(t) = b$$

then

$$\lim_{k \to \infty} (t_{k+1} - t_k) = \frac{\pi}{\sqrt{b}}. \tag{4.8}$$

Proof. By Lemma 4.4, the number of zeros of $x(t)$ on any bounded interval $[t_0, T]$ is finite, which implies that the set of zeros in $[t_0, +\infty)$ is at most countable and that all zeros can be numbered in the increasing order.

Consider the ODE $y'' + ay = 0$ that has solution $y(t) = \sin \sqrt{a}t$. By Theorem 4.5, $x(t)$ has a zero between any two zeros of $y(t)$, that is, in any interval $\left[\frac{\pi k}{\sqrt{a}}, \frac{\pi(k+1)}{\sqrt{a}}\right] \subset [t_0, +\infty)$. This implies that $x(t)$ has in $[t_0, +\infty)$ infinitely many zeros. Hence, the set of zeros of $x(t)$ is countable and forms an increasing sequence $\{t_k\}_{k=1}^{\infty}$. The fact that any bounded interval contains finitely many terms of this sequence implies that $t_k \to +\infty$.

To prove the second claim, fix some $T > t_0$ and set

$$m = m(T) = \inf_{t \in [T, +\infty)} q(t).$$

Consider the ODE $y'' + my = 0$. Since $m \leq q(t)$ in $[T, +\infty)$, between any two zeros of $y(t)$ in $[T, +\infty)$ there is a zero of $x(t)$. Consider a zero t_k of $x(t)$ that is contained in $[T, +\infty)$ and prove that

$$t_{k+1} - t_k \leq \frac{\pi}{\sqrt{m}}. \tag{4.9}$$

Assume from the contrary that that $t_{k+1} - t_k > \frac{\pi}{\sqrt{m}}$. Consider a solution

$$y(t) = \sin\left(\frac{\pi t}{\sqrt{m}} + \varphi\right),$$

whose zeros form an arithmetic sequence $\{s_j\}$ with difference $\frac{\pi}{\sqrt{m}}$, that is, for all j,

$$s_{j+1} - s_j = \frac{\pi}{\sqrt{m}} < t_{k+1} - t_k.$$

Choosing the phase φ appropriately, we can achieve so that, for some j,

$$[s_j, s_{j+1}] \subset (t_k, t_{k+1}) .$$

However, this means that between zeros s_j, s_{j+1} of y there is no zero of x. This contradiction proves (4.9).

If $b = +\infty$ then by letting $T \to \infty$ we obtain $m \to \infty$ and, hence, $t_{k+1} - t_k \to 0$ as $k \to \infty$, which proves (4.8) in this case.

Consider the case when b is finite. Then setting

$$M = M(T) = \sup_{t \in [T,+\infty)} q(t) ,$$

we obtain in the same way that

$$t_{k+1} - t_k \geq \frac{\pi}{\sqrt{M}}.$$

When $T \to \infty$, both $m(T)$ and $M(T)$ tend to b, which implies that

$$t_{k+1} - t_k \to \frac{\pi}{\sqrt{b}}.$$

■

4.5 The Bessel equation

The *Bessel equation* is the ODE

$$t^2 x'' + t x' + \left(t^2 - \alpha^2\right) x = 0 \tag{4.10}$$

where $t > 0$ is an independent variable, $x = x(t)$ is the unknown function, $\alpha \in \mathbb{R}$ is a given parameter[11]. The *Bessel functions*[12] are certain particular solutions of this equation. The value of α is called the order of the Bessel equation.

Theorem 4.7 *Let $x(t)$ be a non-zero solution to the Bessel equation on $(0, +\infty)$. Then the zeros of $x(t)$ form an infinite sequence $\{t_k\}_{k=1}^{\infty}$ such that $t_k < t_{k+1}$ for all $k \in \mathbb{N}$ and $t_{k+1} - t_k \to \pi$ as $k \to \infty$.*

Proof. Write the Bessel equation in the form

$$x'' + \frac{1}{t} x' + \left(1 - \frac{\alpha^2}{t^2}\right) x = 0, \tag{4.11}$$

set $p(t) = \frac{1}{t}$ and $q(t) = \left(1 - \frac{\alpha^2}{t^2}\right)$. Then the change

$$\begin{aligned} u(t) &= x(t) \exp\left(\frac{1}{2}\int p(t)\, dt\right) \\ &= x(t) \exp\left(\frac{1}{2}\ln t\right) = x(t)\sqrt{t} \end{aligned}$$

brings the ODE to the form

$$u'' + Q(t) u = 0$$

where

$$Q(t) = q - \frac{p^2}{4} - \frac{p'}{2} = 1 - \frac{\alpha^2}{t^2} + \frac{1}{4t^2}. \tag{4.12}$$

Note the roots of $x(t)$ are the same as those of $u(t)$. Observe also that $Q(t) \to 1$ as $t \to \infty$ and, in particular, $Q(t) \geq \frac{1}{2}$ for $t \geq T$ for large enough T. Theorem 4.6 yields that the

[11] In general, one can let α to be a complex number as well but here we restrict ourselves to the real case.

[12] The Bessel function of the first kind is defined by

$$J_\alpha(t) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\,\Gamma(m+\alpha+1)} \left(\frac{t}{2}\right)^{2m+\alpha}.$$

It is possible to prove that $J_\alpha(t)$ solves (4.10). If α is non-integer then J_α and $J_{-\alpha}$ are linearly independent solutions to (4.10). If $\alpha = n$ is an integer then the independent solutions are J_n and Y_n where

$$Y_n(t) = \lim_{\alpha \to n} \frac{J_\alpha(t)\cos\alpha\pi - J_{-\alpha}(t)}{\sin\alpha\pi}$$

is the Bessel function of the second kind.

roots of $x(t)$ in $[T,+\infty)$ form an increasing sequence $\{t_k\}_{k=1}^{\infty}$ such that $t_{k+1}-t_k\to\pi$ as $k\to\infty$.

Now we need to prove that the number of zeros of $x(t)$ in $(0,T]$ is finite. Lemma 4.4 says that the number of zeros is finite in any interval $[\tau,T]$ where $\tau>0$, but cannot be applied to the interval $(0,T]$ because the ODE in question is not defined at 0. Let us show that, for small enough $\tau>0$, the interval $(0,\tau)$ contains no zeros of $x(t)$. Consider the following function on $(0,\tau)$

$$z(t)=\ln\frac{1}{t}-\sin t$$

which is positive in $(0,\tau)$ provided τ is small enough (in fact, $z(t)\to+\infty$ as $t\to0$). For this function we have

$$z'=-\frac{1}{t}-\cos t \ \text{ and } \ z''=\frac{1}{t^2}+\sin t$$

whence

$$z''+\frac{1}{t}z'+z=\ln\frac{1}{t}-\frac{\cos t}{t}.$$

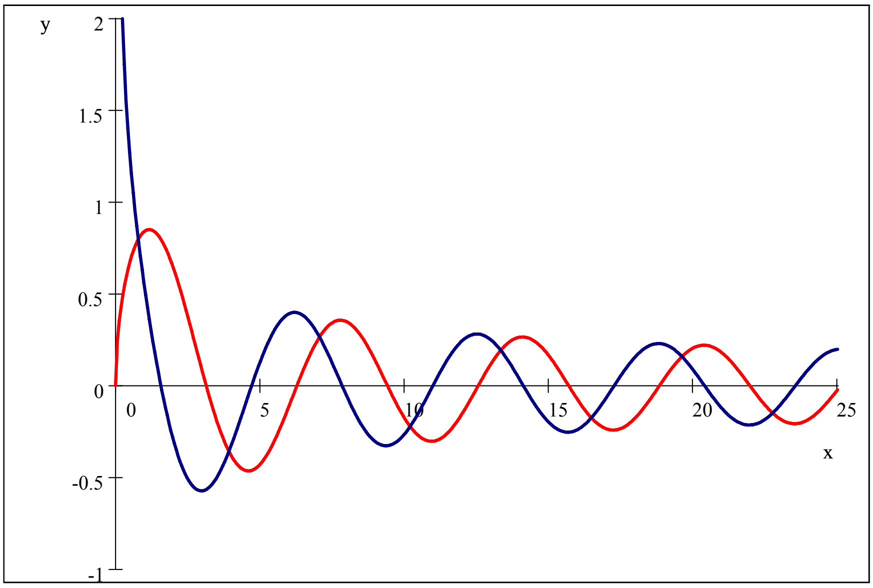
Since $\frac{\cos t}{t}\sim\frac{1}{t}$ and $\ln\frac{1}{t}=o\left(\frac{1}{t}\right)$ as $t\to0$, we see that the right hand side here is negative in $(0,\tau)$ provided τ is small enough. It follows that

$$z''+\frac{1}{t}z'+\left(1-\frac{\alpha^2}{t^2}\right)z<0, \tag{4.13}$$

so that $z(t)$ is a positive supersolution of the Bessel equation in $(0,\tau)$. By Corollary of Theorem 4.5, $x(t)$ has at most one zero in $(0,\tau)$. By further reducing τ, we obtain that $x(t)$ has no zeros on $(0,\tau)$, which finishes the proof. ■

Example. In the case $\alpha=\frac{1}{2}$ we obtain from (4.12) $Q(t)\equiv1$ and the ODE for $u(t)$ becomes $u''+u=0$. Using the solutions $u(t)=\cos t$ and $u(t)=\sin t$ and the relation $x(t)=t^{-1/2}u(t)$, we obtain the independent solutions of the Bessel equation: $x(t)=t^{-1/2}\sin t$ and $x(t)=t^{-1/2}\cos t$. Clearly, in this case we have exactly $t_{k+1}-t_k=\pi$.

The functions $t^{-1/2}\sin t$ and $t^{-1/2}\cos t$ show the typical behavior of solutions to the Bessel equation: oscillations with decaying amplitude as $t\to\infty$:



Remark. In (4.13) we have used that $\alpha^2 \geq 0$ which is the case for real α. For imaginary α one may have $\alpha^2 < 0$ and the above argument does not work. In this case a solution to the Bessel equation can actually have a sequence of zeros that accumulate at 0[13].

4.6 Sturm-Liouville problem

Fix an interval $[a, b] \subset \mathbb{R}$ and functions $p \in C^1[a, b]$, $q, w \in C[a, b]$, such that $p, w > 0$ on $[a, b]$ and consider the following problem on $[a, b]$:

$$\begin{cases} (p(t)\,x')' + q(t)\,x + \lambda w x = 0, \\ x(a) = x(b) = 0. \end{cases} \tag{4.14}$$

Here $x(t)$ is an unknown function on $[a, b]$ and λ is an unknown parameter. Clearly, $x(t) \equiv 0$ always solves (4.14).

Definition. The *Sturm-Liouville problem* is the task to find all *non-zero* functions $x(t)$ on $[a, b]$ and constants λ that satisfy (4.14).

As we will see later on, such solutions may exist only for specific values of λ. Hence, a part of the problem is to find those λ for which non-zero solutions exist.

Definition. The variable λ is called the *spectral parameter* of the problem (4.14). The values of λ for which a non-zero solution of (4.14) exists are called the *eigenvalues* of (4.14). A non-zero solution $x(t)$ is called the *eigenfunction* of (4.14). The condition $x(a) = x(b) = 0$ is called the *Dirichlet boundary condition.*

Similar problems can be considered with other boundary conditions, for example, with $x'(a) = x'(b) = 0$ (the Neumann boundary condition) but we will restrict ourselves to the problem (4.14).

Note that the ODE in (4.14) can be rewritten in the form

$$\begin{aligned} px'' + p'x' + qx + \lambda wx &= 0, \\ x'' + \frac{p'}{p}x' + \frac{q}{p}x + \lambda\frac{w}{p}x &= 0, \end{aligned}$$

[13] Consider the ODE

$$v'' + \frac{c^2}{t^2}v = 0$$

where $c > \frac{1}{2}$. It is the Euler equation and its solution can be found in the form $v(t) = t^b$, where b is found from the equation

$$b(b-1) + c^2 = 0$$

that is, $b = \frac{1}{2} \pm i\beta$ where $\beta = \sqrt{c^2 - \frac{1}{4}}$. Hence, the solutions are

$$\sqrt{t}\cos(\beta \ln t) \quad \text{and} \quad \sqrt{t}\sin(\beta \ln t),$$

and both have sequences of zeros converging to 0. By Theorem 4.5, a solution u to the ODE

$$u'' + \left(1 + \frac{c^2}{t^2}\right)u = 0$$

will also have a sequence of zeros accumulating to 0, which implies the same property for the solutions to the Bessel equation with negative α^2.

that is,

$$x'' + Px' + Qx + \lambda Wx = 0. \tag{4.15}$$

where

$$P = \frac{p'}{p}, \quad Q = \frac{q}{p} \text{ and } Q = \frac{w}{p}.$$

In the form (4.15), functions P and Q can be any continuous functions on $[a, b]$ and W must be a positive continuous function on $[a, b]$. Under these conditions, the ODE (4.15) can be converted back to (4.14) by finding p from the equation $\frac{p'}{p} = P$, which always has a positive continuously differentiable solution

$$p(t) = \exp\left(\int P(t)\,dt\right).$$

Hence, the two forms (4.14) and (4.15) of the ODE are equivalent but (4.14) has certain advantages that will be seen in Theorem 4.9 below.

Observe that if $x(t)$ is the eigenfunction then $Cx(t)$ is also the eigenfunction, where C is a non-zero constant. It turns our that the converse is true as well.

Lemma 4.8 *If $x(t)$ and $y(t)$ are two eigenfunctions with the same eigenvalue then $y(t) = Cx(t)$ for some constant C (that is, every eigenvalue has the geometric multiplicity 1).*

Proof. Observe that $x'(a) \neq 0$ (otherwise, a is multiple zero of x) so that we can set $C = \frac{y'(a)}{x'(a)}$. Then the function

$$z(t) = y(t) - Cx(t)$$

vanishes at $t = a$ and the derivative $z'(t)$ also vanishes at $t = a$ by the choice of C. Hence, $z(t) \equiv 0$ on $[a, b]$ whence the result follows. ■

Hence, when solving the Sturm-Liouville problem, one needs to find all eigenvalues and one eigenfunction for each eigenvalue.

Example. Consider the simplest instance of the Sturm-Liouville problem

$$\begin{cases} x'' + \lambda x = 0 \\ x(0) = x(a) = 0 \end{cases}$$

Let us first observe that if $\lambda \leq 0$ then there is no solution. Indeed, in this case the function $y(t) \equiv 1$ is a positive supersolution: $y'' + \lambda y \leq 0$, whence it follows that a non-zero solution $x(t)$ cannot have two distinct zeros. Hence, we can restrict to the case $\lambda > 0$. The general solution to the ODE $x'' + \lambda x = 0$ is then

$$x = C_1 \cos\left(\sqrt{\lambda} t\right) + C_2 \sin\left(\sqrt{\lambda} t\right).$$

The boundary condition amount to

$$\begin{aligned} C_1 &= 0 \\ C_2 \sin\left(a\sqrt{\lambda}\right) &= 0. \end{aligned}$$

Hence, possible values for λ are

$$\lambda = \frac{\pi^2 k^2}{a^2}, \quad k \in \mathbb{N}$$

and the corresponding eigenfunctions are

$$x\left(t\right)=\sin\left(\sqrt{\lambda}t\right)=\sin\frac{\pi kt}{a}.$$

Especially simple form it takes when $a=\pi$: in this case, the eigenvalues are given by

$$\lambda=k^2$$

and the eigenfunctions are

$$x\left(t\right)=\sin kt.$$

Consider an example showing how the Sturm-Liouville problem occurs in applications.

Example. *(The heat equation)* The heat equation is a *partial differential equation* (PDE) for a function $u=u\left(t,x\right)$ of the form

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}.$$

One of the problems associated with this PDE is a so called *initial-boundary problem*

$$\begin{cases} \dfrac{\partial u}{\partial t}=\dfrac{\partial^2 u}{\partial x^2}, & t\geq 0,\ x\in\left[a,b\right], & \text{(the heat equation)} \\ u\left(0,x\right)=f\left(x\right), & x\in\left[a,b\right], & \text{(the initial condition)} \\ u\left(t,a\right)=u\left(t,b\right)=0, & t\geq 0, & \text{(the boundary condition)} \end{cases} \tag{4.16}$$

where $f\left(x\right)$ is a given function on $\left[a,b\right]$. Physically this corresponds to finding the temperature $u\left(t,x\right)$ at time t at point x provided it is known that the temperature at the boundary points $x=a$ and $x=b$ remains constant 0 for all $t\geq 0$ while the temperature at the initial time $t=0$ was $f\left(x\right)$.

This problem can be solved by the *method of separation of variables* as follows. Let us first try and find solutions to the heat equation in the form $u\left(t\right)=y\left(x\right)z\left(t\right)$ The heat equation becomes

$$z'y=zy''$$

that is

$$\frac{z'}{z}\left(t\right)=\frac{y''}{y}\left(x\right).$$

Hence, we have the identity of two functions one of them depending on t and the other – on x. Of course, this can happen only if both functions are constants. Denote this constant by $-\lambda$ so that we obtain two separate equations

$$\begin{aligned} z'+\lambda x &= 0 \\ y''+\lambda y &= 0. \end{aligned}$$

To ensure the boundary conditions for u, it suffices to require that

$$y\left(a\right)=y\left(b\right)=0.$$

Hence, the function y must solve the Sturm-Liouville problem

$$\begin{cases} y'' + \lambda y = 0 \\ y(a) = y(b) = 0 \end{cases}$$

(of course, we are interested only in non-zero solutions y). Setting for simplicity $a = 0$ and $b = \pi$, we obtain as above the sequence of the eigenvalues $\lambda_k = k^2$ and the eigenfunctions

$$y_k(x) = \sin kx,$$

where $k \in \mathbb{N}$. For $\lambda = k^2$, the ODE $z' + \lambda z = 0$ has the general solution

$$z_k(t) = C_k e^{-k^2 t}.$$

Hence, we obtain a sequence $u_k(t,x) = C_k e^{-k^2 t} \sin kx$ of solutions to the heat equation that satisfy the boundary condition.

Note that $u_k(0,x) = C_k \sin kx$. Hence, if the initial function $f(x)$ has the form $C_k \sin kx$ then the solution to the problem (4.16) is the function $u_k(t,x)$. In a more general situation, if

$$f(x) = \sum_{k=1}^{N} C_k \sin kx \tag{4.17}$$

then the solution to (4.16) is

$$u(t,x) = \sum_{k=1}^{N} C_k e^{-k^2 t} \sin kx. \tag{4.18}$$

This is trivial for a finite N but in certain sense is true also when $N = \infty$. This is the most useful case because for $N = \infty$ the right hand side of (4.17) is a sin-Fourier series. Given a function f on $[0,\pi]$ such that $f(0) = 0 = f(\pi)$ (which are necessary condition for the consistency of (4.16), extend $f(x)$ oddly to $[-\pi, 0)$ so that the Fourier series of f on $[-\pi,\pi]$ contains only the sin-terms. Then one obtains the solution $u(t,x)$ also in the form of the Fourier series (4.18). Of course, some justifications are needed here in order to be able to differentiate (4.18) term-by-term, and some additional restrictions should be imposed on f. However, we do no go into further details of this subject.

This example shows how the Sturm-Liouville problem occurs naturally in PDEs and motivates the further study of the Sturm-Liouville problem.

Theorem 4.9 *Consider the Sturm-Liouville problem* (4.14).

(*a*) *If λ is the eigenvalue of* (4.14) *with the eigenfunction $x(t)$ then*

$$\lambda = \frac{\int_a^b \left(p(x')^2 - qx^2\right) dt}{\int_a^b wx^2 dt}. \tag{4.19}$$

(*b*) (The orthogonality relations) *If $x_1(t)$ and $x_2(t)$ are the eigenfunctions of* (4.14) *with the distinct eigenvalues then*

$$\int_a^b x_1(t)\, x_2(t)\, w(t)\, dt = 0. \tag{4.20}$$

Remark. Given a continuous positive function w on $[a,b]$, the expression

$$(f,g) := \int_a^b f(t)\, g(t)\, w(t)\, dt$$

can be interpreted as an *inner product* in the linear space $C[a,b]$. Indeed, the functional $f,g \mapsto (f,g)$ is obviously symmetric, bilinear and positive definite, that is, $(f,f) \geq 0$ and $f(,f) = 0$ if and only if $f = 0$. Hence, (f,g) satisfies the definition of an inner product. Using the inner product, one can introduce the 2-norm of a function $f \in C[a,b]$ by

$$\|f\|_2 = \sqrt{(f,f)}$$

and the angle α between two non-zero functions $f,g \in C[a,b]$ by

$$\cos\alpha = \frac{(f,g)}{\|f\|_2\,\|g\|_2}.$$

In particular, f and g are orthogonal (that is, $\alpha = \pi/2$) if and only if $(f,g) = 0$.

Hence, part (b) of Theorem 4.9 means that the eigenfunctions of different eigenvalues are orthogonal with respect to the chosen inner product[14].

Proof of Theorem 4.9. Let λ_i be the eigenvalue of the eigenfunction x_i, $i = 1,2$. Multiplying the ODE

$$(px_1')' + qx_1 + \lambda_1 w x_1 = 0$$

by x_2 and integrating over $[a,b]$, we obtain

$$\int_a^b (px_1')'\, x_2 dt + \int_a^b qx_1x_2 dt + \lambda_1 \int wx_1x_2 dt = 0.$$

Integrating by parts in the first integral, we obtain that it is equal to

$$[p_1x_1'x_2]_a^b - \int_a^b px_1'x_2' dt.$$

By the boundary condition $x_2(a) = x_2(b) = 0$, we see that the first term vanishes, and we obtain the identity

$$\int_a^b px_1'x_2' = \int_a^b qx_1x_2 dt + \lambda_1 \int_a^b wx_1x_2 dt. \tag{4.21}$$

(a) If $x_1 = x_2 = x$ and $\lambda_1 = \lambda$ then (4.21) implies

$$\int_a^b p\,(x')^2\, dt = \int_a^b qx^2 dt + \lambda \int wx^2 dt$$

[14] This is similar to the fact that the eigenvectors with different eigenvalues of any real symmetric $n \times n$ matrix A are automatically orthogonal with respect to the canonical inner product in $\mathbb{R}^n$. Indeed, if x_1 and x_2 are the eigenvectors with the eigenvalues $\lambda_1 \neq \lambda_2$ then $Ax_1 = \lambda_1 x_1$ implies $(Ax_1, x_2) = \lambda_1(x_1,x_2)$ and $Ax_2 = \lambda_2 x_2$ implies $(x_1, Ax_2) = \lambda_2(x_1,x_2)$. By the symmetry of A, we have $(Ax_1,x_2) = (x_1,Ax_2)$ whence $\lambda_1(x_1,x_2) = \lambda_2(x_1,x_2)$ and $(x_1,x_2) = 0$.

Part (a) of Theorem 4.9 is analogous to the identity $\lambda = \frac{(Ax,x)}{\|x\|^2}$ for an eigenvector x with the eigenvalue λ, which trivially follows from $Ax = \lambda x$ by taking the inner product with x.

whence (4.19) follows.

(*b*) Switching the indices 1 and 2 in (4.21) and noticing that all the integrals are symmetric with respect to the indices 1 and 2, we obtain

$$\int_a^b px_1'x_2' = \int_a^b qx_1x_2dt + \lambda_2 \int_a^b wx_1x_2dt. \tag{4.22}$$

Since $\lambda_1 \neq \lambda_2$, the two identities (4.21) and (4.22) can be simultaneously satisfied only if

$$\int_a^b wx_1x_2dt = 0$$

which was to be proved. ■

Example. Recall that the Sturm-Liouville problem

$$\begin{cases} x'' + \lambda x = 0 \\ x(0) = x(\pi) = 0 \end{cases}$$

has the eigenfunctions $\sin kt$, $k \in \mathbb{N}$. Hence, the orthogonality relation (4.20) becomes

$$\int_0^\pi \sin k_1 t \sin k_2 t \; dt = 0 \text{ for all } k_1 \neq k_2,$$

which is, of course, obvious without Theorem 4.9. A version of this relation on the interval $[-\pi, \pi]$ is used in the theory of Fourier series.

Let us briefly discuss some more interesting examples. It follows from the proof of Theorem 4.9(b) that the orthogonality relation remains true in a more general situation when the given ODE is defined in an *open* interval (a, b) and the following conditions are satisfied:

(i) the integral $\int_a^b x_1 x_2 w dt$ converges as improper;

(ii) $[p x_1' x_2]_a^b = [p x_1 x_2']_a^b = 0$ where the values at a and b are understood in the sense of limit.

Example. *The Legendre polynomials* are the eigenfunctions of the following problem on $(-1, 1)$:

$$\begin{cases} (1 - t^2) x'' - 2t x' + \lambda x = 0 \\ x(\pm 1) \text{ finite.} \end{cases}$$

The ODE can be written in the Sturm-Liouville form as

$$\left((1 - t^2) x' \right)' + \lambda x = 0.$$

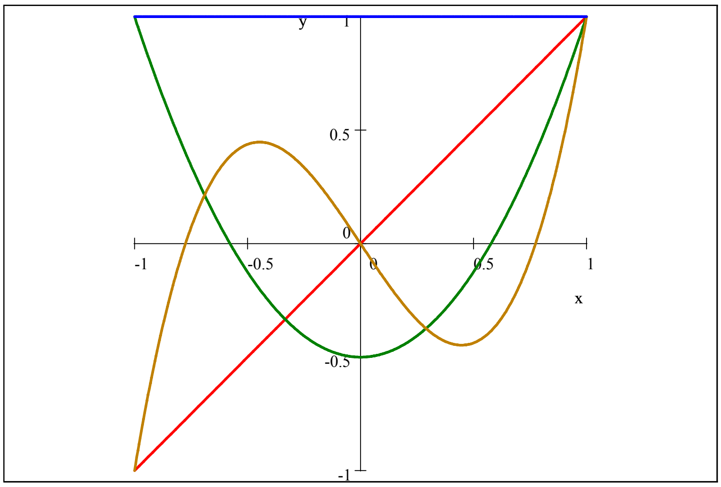
The eigenvalues are $\lambda_n = n(n+1)$, where n is non-negative integer. The eigenfunction of λ_n is

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} \left[(t^2 - 1)^n \right]$$

which is obviously a polynomial of degree n (the coefficient $\frac{1}{2^n n!}$ is chosen for normalization purposes). Since $p(t) = 1 - t^2$ vanishes at ± 1, the above conditions (i) and (ii) are satisfied, and we obtain that the sequence $\{P_n\}$ is orthogonal in $[-1, 1]$ with the weight function $w = 1$.

Here are the first few Legendre polynomial and their graphs:

$$P_0(t) = 1, \quad P_1(t) = t, \quad P_2(t) = \frac{1}{2}\left(3t^2 - 1\right), \; P_3(t) = \frac{1}{2}\left(5t^3 - 3t\right), \; ...$$



Example. *The Chebyshev polynomials* are the eigenfunctions of the following problem on $(-1,1)$:

$$\begin{cases} (1-t^2)\,x''-tx'+\lambda x=0\\ x\,(\pm1)\ \text{finite.}\end{cases}$$

The ODE can be rewritten in the Sturm-Liouville form

$$\left(\sqrt{1-t^2}x'\right)'+\frac{\lambda x}{\sqrt{1-t^2}}=0$$

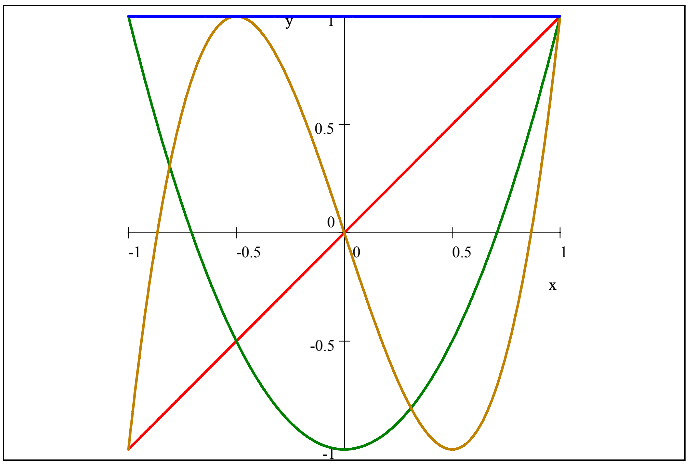
so that $p=\sqrt{1-t^2}$ and $w=\frac{1}{\sqrt{1-t^2}}$. The eigenvalues are $\lambda=n^2$ where n is a non-negative integer, and the eigenfunction of λ_n is

$$T_n\,(t)=\cos\,(n\arccos t)\,,$$

which is a polynomial of the degree n. Since $p\,(\pm1)=0$ and $\int_{-1}^{1}w\,(t)\,dt<\infty$, the conditions (i) and (ii) are satisfied so that $\{T_n\}$ are orthogonal with the weight $\frac{1}{\sqrt{1-t^2}}$.

Here are the first few Chebyshev polynomials and their graphs:

$$T_0\,(t)=1,\quad T_1\,(t)=t,\ \ T_2\,(t)=2t^2-1,\ T_3\,(t)=4t^3-3t,\ \ldots$$



Example. *The Hermite polynomials* are the eigenfunctions of the problem on $(-\infty, +\infty)$

$$\begin{cases} x'' - tx' + \lambda x = 0 \\ x(t) = o\left(t^N\right) \text{ as } t \to \pm\infty. \end{cases}$$

The ODE can be rewritten in the Sturm-Liouville form

$$\left(x' e^{-t^2/2}\right)' + \lambda e^{-t^2/2} x = 0,$$

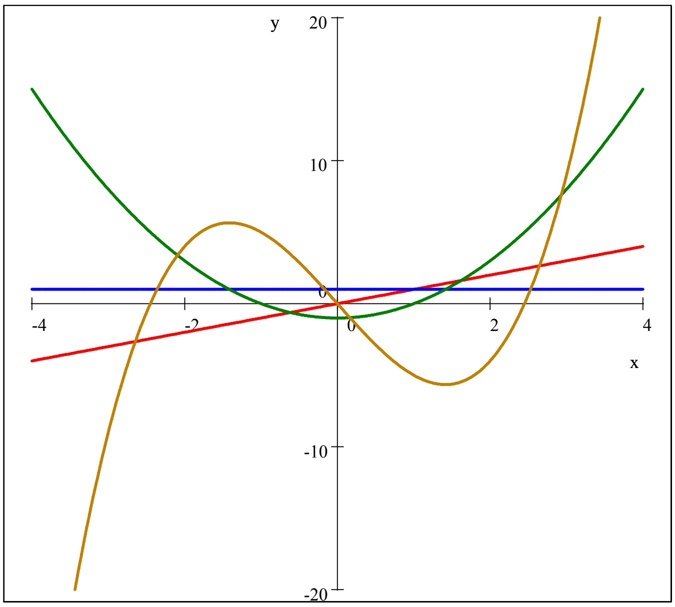
so that $p = w = e^{-t^2/2}$. The eigenvalues are $\lambda_n = n$, n is a non-negative integer, and the eigenfunction of λ_n is

$$H_n(t) = (-1)^n e^{t^2/2} \frac{d^n}{dt^n} e^{-t^2/2},$$

which is a polynomial of degree n. Since $p(t)$ decays fast enough as $t \to \infty$, the conditions (i) and (ii) are satisfied and we obtain that $\{H_n\}$ is orthogonal on $(-\infty, +\infty)$ with the weight $e^{-t^2/2}$.

Here are the first few Hermite polynomials and their graphs:

$$H_0(t) = 1, \quad H_1(t) = t, \quad H_2(t) = t^2 - 1, \; H_3(t) = t^3 - 6t, \; ...$$



Theorem 4.10 (The Sturm-Liouville theorem) *Assume that* $p \in C^2[a,b]$, $q, w \in C[a,b]$ *and* $p, w > 0$ *on* $[a,b]$. *Then the Sturm-Liouville problem*

$$\begin{cases} (px')' + qx + \lambda wx = 0 \\ x(a) = x(b) = 0 \end{cases}$$

has a sequence $\{\lambda_k\}_{k=1}^{\infty}$ *of eigenvalues and the corresponding eigenfunctions* $x_k(t)$ *such that*

(a) $\lambda_k < \lambda_{k+1}$ *and* $\lambda_k \to +\infty$ *as* $k \to 0$.

(b) *The eigenfunction* $x_k(t)$ *has exactly* $k-1$ *zeros in* (a,b).

Proof. Write the Sturm-Liouville equation in the form

$$x'' + Px' + Qx + \lambda W x = 0, \tag{4.23}$$

where

$$P = \frac{p'}{p}, \quad Q = \frac{q}{p} \text{ and } Q = \frac{w}{p}.$$

Note that $P \in C^1[a,b]$. As in the proof of Theorem 4.5, we can get rid of P by the change

$$u(t) = x(t)\exp\left(\frac{1}{2}\int P dt\right) = x(t)\exp\left(\frac{1}{2}\int \frac{p'}{p}dt\right) = x(t)\sqrt{p},$$

which leads to the ODE

$$u'' + \widetilde{Q}(t)u + \lambda W(t)u = 0.$$

where

$$\widetilde{Q} = Q - \frac{1}{4}P^2 - \frac{P'}{2}.$$

This means that we can assume from the very beginning that $p = 1$ and write the Sturm-Liouville problem in the form

$$\begin{cases} x'' + qx + \lambda wx = 0 \\ x(a) = x(b) = 0. \end{cases} \tag{4.24}$$

Also, q can be assumed non-positive because replacing q by $q - Cw$ we just replace λ by $\lambda + C$ without changing the eigenfunctions. Hence, assume in the sequel that $q < 0$ on $[a,b]$. It follows (for example, from the Example after Theorem 4.5 or from Theorem 4.9(a)) that all the eigenvalues are positive. Thus, we can assume in the sequel that the spectral parameter λ is positive.

Extend the functions $q(t)$, $w(t)$ continuously to all $t \in \mathbb{R}$ so that, for large enough t, $q(t) = 0$ and $w(t) = 1$; hence, the ODE for large t becomes $x'' + \lambda x = 0$. For a fixed $\lambda > 0$, consider the following IVP on $\mathbb{R}$

$$\begin{cases} x'' + (q + \lambda w)x = 0 \\ x(a) = 0 \\ x'(a) = 1 \end{cases}$$

and denote the solution by $x(t,\lambda)$. We are interested in those $\lambda > 0$, for which

$$x(b,\lambda) = 0,$$

because these λ will be exactly the eigenvalues, and the corresponding solutions $x(t,\lambda)$ (restricted to $[a,b]$) – the eigenfunctions. In other words, we look for those λ for which b is a zero of the function $x(t,\lambda)$ (as a function of t).

For any $\lambda > 0$, consider all zeros of $x(t,\lambda)$ in $t \in [a,+\infty)$. For large enough t, the equation becomes $x'' + \lambda x = 0$ and its zeros form an increasing sequence going to $+\infty$. For a bounded range of t, there is only finitely many zeros of $x(t,\lambda)$ (Lemma 4.4). Hence, for any $\lambda > 0$, all zeros of $x(t,\lambda)$ in $[a,+\infty)$ can be enumerated in the increasing order. Denote them by $\{z_k(\lambda)\}_{k=0}^{\infty}$ where $z_0(\lambda) = a$ and $z_k(\lambda) > a$ for $k \in \mathbb{N}$. The condition $x(b,\lambda) = 0$ means that b is one of zeros of $x(t,\lambda)$, that is

$$z_k(\lambda) = b \text{ for some } k \in \mathbb{N}.$$

In order to be able to solve this equation for λ, consider some properties of the functions $z_k(\lambda)$, $k \in \mathbb{N}$.

Claim 1 *For any fixed $k \in \mathbb{N}$, the function $z_k(\lambda)$ is continuous function of λ.*

Now let us prove by induction in $k \geq 0$ that $z_k(\lambda)$ is continuous in λ. The case $k = 0$ is trivial because $z_k(\lambda) \equiv a$. Assuming that the function $z_{k-1}(\lambda)$ is continuous, let us prove that $z_k(\lambda)$ is also continuous. Fix some $\lambda_0 > 0$ and write for simplicity of notation $z_k = z_k(\lambda_0)$. We need to prove that for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$|\lambda - \lambda_0| < \delta \implies |z_k(\lambda) - z_k| < \varepsilon.$$

It suffices to prove this for sufficiently small ε.

Choose $\varepsilon > 0$ so small that the function $x(t, \lambda_0)$ has in the interval $(z_k - \varepsilon, z_k + \varepsilon)$ only one zero (which is possible by Lemma 4.4). By the continuity of $x(t, \lambda)$ in λ, there is $\delta > 0$ such that if $|\lambda - \lambda_0| < \delta$ then the function $x(t, \lambda)$ has the same sign at $t = z_k \pm \varepsilon$ as $x(t, \lambda_0)$. Hence, by the intermediate value theorem, the function $x(t, \lambda)$ must have a zero in $(z_k - \varepsilon, z_k + \varepsilon)$.

How to ensure that $x(t, \lambda)$ has exactly one zero in this interval? Let $M = \sup w$ and consider the ODE

$$y'' + \lambda M y = 0,$$

that has solution $y = \sin\left(\sqrt{\lambda M} t + \varphi\right)$. Since $x(t, \lambda)$ solves the ODE

$$x'' + (q + \lambda w) x = 0$$

and $q + \lambda w \leq \lambda M$, Theorem 4.5 implies that between any two zeros of $x(t, \lambda)$ is a zero of $y(t)$. It follows that the distance between two consecutive zeros of $x(t, \lambda)$ is at least

$$\frac{\pi}{\sqrt{\lambda M}} \geq \frac{\pi}{\sqrt{2\lambda_0 M}},$$

where we have assumed that $\lambda < 2\lambda_0$ which is true if λ is close enough to λ_0. Assuming further that

$$2\varepsilon < \frac{\pi}{\sqrt{2\lambda_0 M}},$$

we obtain that $x(t, \lambda)$ cannot have two zeros in $(z_k - \varepsilon, z_k + \varepsilon)$.

Now we are left to ensure that the unique zero of $x(t, \lambda)$ in $(z_k - \varepsilon, z_k + \varepsilon)$ is exactly the k-th zero, that is $z_k(\lambda)$. Write for simplicity $z_{k-1} = z_{k-1}(\lambda_0)$. Then, by the choice of ε, the interval $(z_{k-1} - \varepsilon, z_{k-1} + \varepsilon)$ contains exactly one zero of $x(t, \lambda)$. By the inductive hypothesis, the function $z_{k-1}(\lambda)$ is continuous. If λ is close enough to λ_0 then $|z_{k-1}(\lambda) - z_{k-1}| < \varepsilon$ so that the unique zero of $x(t, \lambda)$ in the interval $(z_{k-1} - \varepsilon, z_{k-1} + \varepsilon)$ is $z_{k-1}(\lambda)$. Between the intervals $(z_{k-1} - \varepsilon, z_{k-1} + \varepsilon)$ and $(z_k - \varepsilon, z_k + \varepsilon)$, that is, in $[z_{k-1} + \varepsilon, z_k - \varepsilon]$, function $x(t, \lambda_0)$ has no zeros and, hence, keep the sign, say, positive. Hence, by the continuity of $x(t, \lambda)$ in (t, λ), the function $x(t, \lambda)$ is positive in this interval as well, provided λ is closed enough to λ_0. Hence, the zero of $x(t, \lambda)$ in $(z_k - \varepsilon, z_k + \varepsilon)$ is the next zero after $z_{k-1}(\lambda)$, that is, $z_k(\lambda)$. This proves that $|z_k(\lambda) - z_k| < \varepsilon$ and, hence, the continuity of $z_k(\lambda)$.

The next claim is a slight modification of the Sturm theorem (Theorem 4.5).

Claim 2 *Let $x'' + q_1(t) x = 0$ and $y'' + q_2(t) y = 0$ on some interval I where $q_1(t) < q_2(t)$ on I. If $x \not\equiv 0$ and α, β are distinct zeros of x then there is a zero of y in (α, β).*

Indeed, as in the proof of Theorem 4.5, we can assume that α, β are consecutive zeros of x and $x(t) > 0$ in (a, β). Also, if y has no zeros in (a, β) then we can assume that $y(t) > 0$ on (α, β) whence $y(\alpha) \geq 0$ and $y(\beta) \geq 0$. Then as in the proof of Theorem 4.5,

$$[x'y - xy']_{\alpha}^{\beta} = \int_{\alpha}^{\beta} (q_2 - q_1)\, xydt.$$

The integral in the right hand side is positive because $q_2 > q_1$ and x, y are positive on (α, β), while the left hand side is

$$x'(\beta)\, y(\beta) - x'(\alpha)\, y(\beta) \leq 0$$

because $x'(\beta) \leq 0$ and $x'(\alpha) \geq 0$. This contradiction finishes the proof.

Claim 3 *For any $k \in \mathbb{N}$, $z_k(\lambda)$ strictly monotone decreases in λ.*

We need to prove that if $\lambda < \lambda'$ then

$$z_k(\lambda') < z_k(\lambda). \tag{4.25}$$

By Claim 2, strictly between any two zeros of $x(t, \lambda)$ there is a zero of $x(t, \lambda')$. In particular, the interval $(z_{k-1}(\lambda), z_k(\lambda))$ contains a zero of $x(t, \lambda')$, that is,

$$z_j(\lambda') \in (z_{k-1}(\lambda), z_k(\lambda)) \text{ for some } j \in \mathbb{N}. \tag{4.26}$$

Now let us prove (4.25) by induction in k. Inductive basis for $k = 1$: since the interval $(z_0(\lambda), z_1(\lambda))$ contains $z_j(\lambda')$, we obtain

$$z_1(\lambda') \leq z_j(\lambda') < z_1(\lambda).$$

Inductive step from $k-1$ to k. By the inductive hypothesis, we have

$$z_{k-1}(\lambda') < z_{k-1}(\lambda).$$

Therefore, (4.26) can be true only if $j > k - 1$ that is, $j \geq k$. It follows that

$$z_k(\lambda') \leq z_j(\lambda') < z_k(\lambda),$$

which finishes the proof.

Claim 4 *For any $k \in \mathbb{N}$, we have* $\sup_{\lambda>0} z_k(\lambda) = +\infty$ *and* $\inf_{\lambda>0} z_k(\lambda) = a$.

We have

$$q + \lambda w \leq \sup q + \lambda \sup w.$$

Since $\sup q \leq 0$ and $\sup w < +\infty$, for any $\varepsilon > 0$ there is $\lambda > 0$ such that

$$\sup_{t\in\mathbb{R}} (q + \lambda w) < \varepsilon.$$

Comparing with the ODE $y'' + \varepsilon y = 0$, we obtain that the distance between any two zeros of $x(t, \lambda)$ is at least $\frac{\pi}{\sqrt{\varepsilon}}$, whence

$$z_1(\lambda) - z_0(\lambda) \geq \frac{\pi}{\sqrt{\varepsilon}},$$

which implies that

$$\sup_{\lambda>0} z_k(\lambda) \geq \sup_{\lambda>0} z_1(\lambda) = \infty.$$

Similarly, we have

$$q + \lambda w \geq \inf q + \lambda \inf w.$$

Since $\inf w > 0$ and $\inf q > -\infty$, for any $E > 0$ there is $\lambda > 0$ such that

$$\inf_{t\in\mathbb{R}} (q + \lambda w) > E.$$

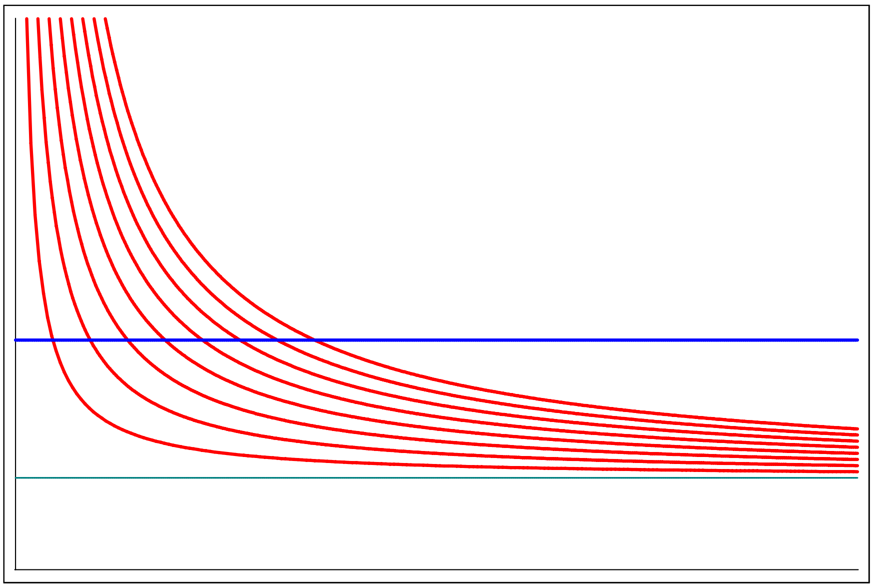
Comparing with the ODE $y'' + Ey = 0$, we obtain that the distance between any two zeros of $x(t, \lambda)$ is at most $\frac{\pi}{\sqrt{E}}$, whence it follows that

$$z_k(\lambda) \leq a + k\frac{\pi}{\sqrt{E}}.$$

Since E can be arbitrarily big, we obtain

$$\inf_{\lambda>0} z_k(\lambda) = a.$$

Hence, the function $z_k(\lambda)$ on $(0, +\infty)$ is continuous, strictly decreasing, its inf is a and sup is $+\infty$. By the intermediate value theorem, $z_k(\lambda)$ takes exactly once all the values in $(a, +\infty)$. Therefore, there is a unique value of λ such that $z_k(\lambda) = b$. Denote this value by λ_k so that $z_k(\lambda_k) = b$. On a plot below, one can see the graphs of $z_k(\lambda)$ with two horizontal lines at the levels a and b, respectively. The intersections with the latter one give the sequence $\{\lambda_k\}$.



Since $x(a, \lambda_k) = x(b, \lambda_k) = 0$, we obtain that the function $x_k(t) := x(t, \lambda_k)$ is the eigenfunction with the eigenvalue λ_k.

Let us show that $\lambda_k < \lambda_{k+1}$. For any $k \geq 0$, we have

$$z_{k+1}(\lambda_k) > z_k(\lambda_k) = b = z_{k+1}(\lambda_{k+1}),$$

which implies by Claim 3 that $\lambda_k < \lambda_{k+1}$.

Let us show that $\lambda_k \to \infty$ as $k \to \infty$. Indeed, if $\{\lambda_k\}$ is bounded, say, $\lambda_k \leq \lambda$ for all k, then

$$b = z_k (\lambda_k) \geq z_k (\lambda) ,$$

which contradicts the fact that $z_k (\lambda) \to \infty$ as $k \to \infty$.

By construction, all zeros of $x_k (t)$ on $[a, +\infty)$ are $z_j (\lambda_k)$. Since $z_k (\lambda_k) = b$, all zeros of $x_k (t)$ in (a, b) are $z_j (\lambda_k)$ with $j = 1, ..., k - 1$. Hence, $x_k (t)$ has exactly $k - 1$ zeros on (a, b), which finishes the proof. ■

Remark. The Sturm-Liouville theorem has the second half (which was actually proved by Steklov) that states the following: the sequence of eigenfunctions $\{x_k\}_{k=1}^{\infty}$ constructed above is *complete*. This means, in particular, that any function $f \in C [a, b]$ (and more generally, any square integrable function f on $[a, b]$) can be split into the series

$$f (t) = \sum_{k=1}^{\infty} c_k x_k (t) , \tag{4.27}$$

where the convergence is understood in the quadratic mean, that is, if f_n is the n-th partial sum then

$$\int_a^b |f (t) - f_n (t)|^2 \, dt \to 0 \text{ as } n \to \infty.$$

The completeness of the sequence of the eigenfunctions has important consequences for applications. As we have seen in the example of the heat equation, the representation of the form (4.27) with $x_k (t) = \sin kt$ was used for the initial function f. The existence of such a representation leads to the solvability of the initial-boundary value problem for a wide enough class of the initial functions f. Similar results for other PDEs require the completeness of the sequence of the eigenfunctions of the general Sturm-Liouville problem.

The proof of the completeness requires additional tools (that will be introduced in Functional Analysis) and is beyond the scope of this course.