Ordinary Differential Equation

Alexander Grigorian
University of Bielefeld

Lecture Notes, April - July 2009

Contents

1 Introduction: the notion of ODEs and examples

A general *ordinary differential equation*[1] (shortly, ODE) has a form

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

where $x \in \mathbb{R}$ is an independent variable, $y = y(x)$ is an unknown function, F is a given function on $n+2$ variables. The number n - the maximal order of the derivative in (1.1), is called the *order* of the ODE. The equation (1.1) is called *differential* because it contains the derivatives[2] of the unknown function. It is called *ordinary*[3] because the derivatives $y^{(k)}$ are ordinary as opposed to partial derivatives. There is a theory of partial differential equations where the unknown function depends on more than one variable and, hence, the partial derivatives are involved, but this is a topic of another lecture course.

The ODEs arise in many areas of Mathematics, as well as in Sciences and Engineering. In most applications, one needs to find explicitly or numerically a solution $y(x)$ of (1.1) satisfying some additional conditions. There are only a few types of the ODEs where one can explicitly find all the solutions. However, for quite a general class of ODEs one can prove various properties of solutions without evaluating them, including existence of solutions, uniqueness, smoothness, etc.

In Introduction we will be concerned with various examples and specific classes of ODEs of the first and second order, postponing the general theory to the next Chapters.

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 a real argument x, and $f(x, y)$ is a given function of two real variables.

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*[4] of the function f and of the equation (1.2). Then the expression $f(x, y)$ makes sense whenever $(x, y) \in D$.

Definition. A real valued function $y(x)$ defined on an interval $I \subset \mathbb{R}$, is called a (*particular*) solution of (1.2) if

1. $y(x)$ is differentiable at any $x \in I$,
2. the point $(x, y(x))$ belongs to D for any $x \in I$,
3. and the identity $y'(x) = f(x, y(x))$ holds for all $x \in I$.

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

[1] Die gewöhnliche Differentialgleichungen

[2] Ableitung

[3] gewöhnlich

[4] Definitionsbereich

Here and below by an interval we mean any set of the form

$$\begin{aligned}
(a,b) &= \{x \in \mathbb{R} : a < x < b\} \\
[a,b] &= \{x \in \mathbb{R} : a \leq x \leq b\} \\
[a,b) &= \{x \in \mathbb{R} : a \leq x < b\} \\
(a,b] &= \{x \in \mathbb{R} : a < x \leq b\},
\end{aligned}$$

where a, b are real or $\pm\infty$ and $a < b$.

Usually a given ODE cannot be solved explicitly. We will consider some classes of $f(x,y)$ when one find the general solution to (1.2) in terms of indefinite integration[5].

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*[6] of f. Assuming that f is a continuous[7] function on an interval I, we obtain the general solution on I by means of the indefinite integration:

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

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

Example. Consider the ODE

$$y' = y.$$

Let us first find all positive solutions, that is, assume that $y(x) > 0$. Dividing the ODE by y and noticing that

$$\frac{y'}{y} = (\ln y)',$$

we obtain the equivalent equation

$$(\ln y)' = 1.$$

Solving this as in the previous example, we obtain

$$\ln y = \int dx = x + C,$$

whence

$$y = e^C e^x = C_1 e^x,$$

where $C_1 = e^C$. Since $C \in \mathbb{R}$ is arbitrary, $C_1 = e^C$ is any positive number. Hence, any positive solution y has the form

$$y = C_1 e^x, \quad C_1 > 0.$$

If $y(x) < 0$ for all x, then use

$$\frac{y'}{y} = (\ln(-y))'$$

and obtain in the same way

$$y = -C_1 e^x,$$

[5] unbestimmtes Integral

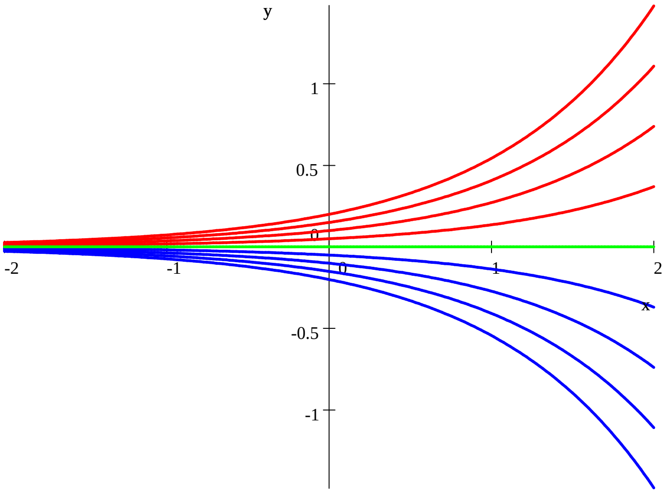
[6] Stammfunction

[7] stetig

where $C_1 > 0$. Combine these two cases together, we obtain that any solution $y(x)$ that remains positive or negative, has the form

$$y(x) = Ce^x,$$

where $C > 0$ or $C < 0$. Clearly, $C = 0$ suits as well since $y = 0$ is a solution. The next plot contains the integrals curves of such solutions:



Claim *The family of solutions $y = Ce^x$, $C \in \mathbb{R}$, is the general solution of the ODE $y' = y$.*

The constant C that parametrizes the solutions, is referred to as a *parameter*. It is clear that the particular solutions are distinguished by the values of the parameter.

Proof. Let $y(x)$ be a solution defined on an open interval I. Assume that $y(x)$ takes a positive value somewhere in I, and let (a, b) be a maximal open interval where $y(x) > 0$. Then either $(a, b) = I$ or one of the points a, b belongs to I, say $a \in I$ and $y(a) = 0$. By the above argument, $y(x) = Ce^x$ in (a, b), where $C > 0$. Since $e^x \neq 0$, this solution does not vanish at a. Hence, the second alternative cannot take place and we conclude that $(a, b) = I$, that is, $y(x) = Ce^x$ in I.

The same argument applies if $y(x) < 0$ for some x. Finally, of $y(x) \equiv 0$ then also $y(x) = Ce^x$ with $C = 0$. ■

1.1 Separable ODE

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

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

Any separable equation can be solved by means of the following theorem.

Theorem 1.1 (The method of separation of variables) *Let $f(x)$ and $g(y)$ be continuous functions on open 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 defined on some subinterval of I, solves the differential equation (1.3) *if and only if it satisfies the identity*

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

for all x in the domain of y, where C is a real constant.

For example, consider again the ODE $y'=y$ in the domain $x\in\mathbb{R}$, $y>0$. Then $f\left(x\right)=1$ and $g\left(y\right)=y\neq 0$ so that Theorem 1.1 applies. We have

$$F\left(x\right)=\int f\left(x\right)dx=\int dx=x$$

and

$$G\left(y\right)=\int\frac{dy}{g\left(y\right)}=\int\frac{dy}{y}=\ln y$$

where we do not write the constant of integration because we need only one primitive function. The equation (1.4) becomes

$$\ln y=x+C,$$

whence we obtain $y=C_1e^x$ as in the previous example. Note that Theorem 1.1 does not cover the case when $g\left(y\right)$ may vanish, which must be analyzed separately when needed.

Proof. Let $y\left(x\right)$ solve (1.3). Since $g\left(y\right)\neq 0$, we can divide (1.3) by $g\left(y\right)$, which yields

$$\frac{y'}{g\left(y\right)}=f\left(x\right). \tag{1.5}$$

Observe that by the hypothesis $f\left(x\right)=F'\left(x\right)$ and $\frac{1}{g'(y)}=G'\left(y\right)$, which implies by the chain rule

$$\frac{y'}{g\left(y\right)}=G'\left(y\right)y'=\left(G\left(y\left(x\right)\right)\right)'.$$

Hence, the equation (1.3) is equivalent to

$$G\left(y\left(x\right)\right)'=F'\left(x\right), \tag{1.6}$$

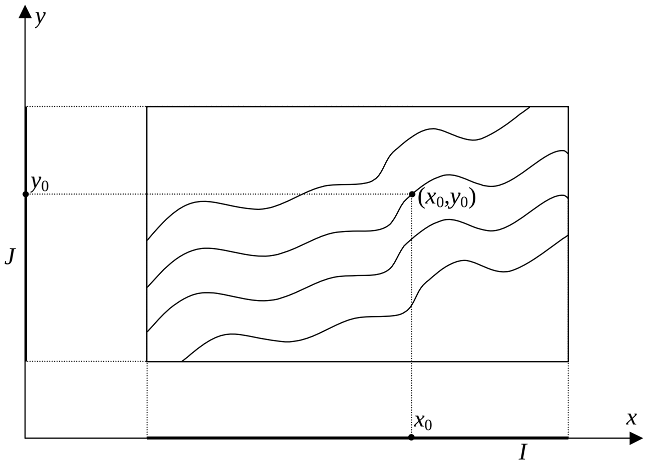
which implies (1.4).

Conversely, if function y satisfies (1.4) and is known to be differentiable in its domain then differentiating (1.4) in x, we obtain (1.6); arguing backwards, we arrive at (1.3). The only question that remains to be answered is why $y\left(x\right)$ is differentiable. Since the function $g\left(y\right)$ does not vanish, it is either positive or negative in the whole domain. Then the function $G\left(y\right)$, whose derivative is $\frac{1}{g(y)}$, is either strictly increasing or strictly decreasing in the whole domain. In the both cases, the inverse function G^{-1} is defined and is differentiable. It follows from (1.4) that

$$y\left(x\right)=G^{-1}\left(F\left(x\right)+C\right). \tag{1.7}$$

Since both F and G^{-1} are differentiable, we conclude by the chain rule that y is also differentiable, which finishes the proof. ■

Corollary. *Under the conditions of* Theorem 1.1*, for all $x_0 \in I$ and $y_0 \in J$ there exists a unique value of the constant C such that the solution $y(x)$ of* (1.3) *defined by* (1.7) *satisfies the condition $y(x_0) = y_0$.*



The condition $y(x_0) = y_0$ is called the *initial condition*[8].

Proof. Setting in (1.4) $x = x_0$ and $y = y_0$, we obtain $G(y_0) = F(x_0) + C$, which allows to uniquely determine the value of C, that is, $C = G(y_0) - F(x_0)$. Let us prove that this value of C determines by (1.7) a solution $y(x)$. The only problem is to check that the right hand side of (1.7) is defined on an interval containing x_0 (a priori it may happen so that the the composite function $G^{-1}(F(x) + C)$ has empty domain). For $x = x_0$ the right hand side of (1.7) is

$$G^{-1}(F(x_0) + C) = G^{-1}(G(y_0)) = y_0$$

so that the function $y(x)$ is defined at $x = x_0$. Since both functions G^{-1} and $F + C$ are continuous and defined on open intervals, their composition is defined on an open set. Since this set contains x_0, it contains also an interval around x_0. Hence, the function y is defined on an interval around x_0, which finishes the proof. ■

One can rephrase the statement of Corollary as follows: for all $x_0 \in I$ and $y_0 \in J$ there exists a unique solution[9] $y(x)$ of (1.3) that satisfies in addition the initial condition $y(x_0) = y_0$; that is, 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.

In applications of Theorem 1.1, it is necessary to find the functions F and G. Technically it is convenient to combine the evaluation of F and G with other computations as follows. The first step is always dividing (1.3) by g to obtain (1.5). Then integrate the both sides in x to obtain

$$\int \frac{y' dx}{g(y)} = \int f(x)\, dx. \tag{1.8}$$

Then we need to evaluate the integral in the right hand side. If $F(x)$ is a primitive of f then we write

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

[8] Anfangsbedingung

[9] The domain of this solution is determined by (1.7) and is, hence, the maximal possible.

In the left hand side of (1.8), we have $y'dx = dy$. Hence, we can change variables in the integral replacing function $y\left(x\right)$ by an independent variable y. We obtain

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

Combining the above lines, we obtain the identity (1.4).

Assume that in the separable equation $y' = f(x)\, g(y)$ the function $g(y)$ vanishes at a sequence of points, say $y_1, y_2, \ldots$, enumerated in the increasing order. Obviously, the constant function $y(x) = y_k$ is a solution of this ODE. The method of separation of variables allows to evaluate solutions in any domain $y \in (y_k, y_{k+1})$, where g does not vanish. Then the structure of the general solution requires an additional investigation.

Example. Consider the ODE

$$y' - xy^2 = 2xy,$$

with domain $(x, y) \in \mathbb{R}^2$. Rewriting it in the form

$$y' = x\left(y^2 + 2y\right),$$

we see that it is separable. The function $g(y) = y^2 + 2y$ vanishes at two points $y = 0$ and $y = -2$. Hence, we have two constant solutions $y \equiv 0$ and $y \equiv 2$. Now restrict the ODE to the domain where $g(y) \neq 0$, that is, to one of the domains

$$\mathbb{R} \times (-\infty, -2), \quad \mathbb{R} \times (-2, 0), \quad \mathbb{R} \times (0, +\infty).$$

In any of these domains, we can use the method of separation of variables, which yields

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

whence

$$\frac{y}{y+2} = C_1 e^{x^2}$$

where $C_1 = \pm e^{2C}$. Clearly, C_1 is any non-zero number here. However, since $y \equiv 0$ is a solution, C_1 can be 0 as well. Renaming C_1 to C, we obtain

$$\frac{y}{y+2} = Ce^{x^2}$$

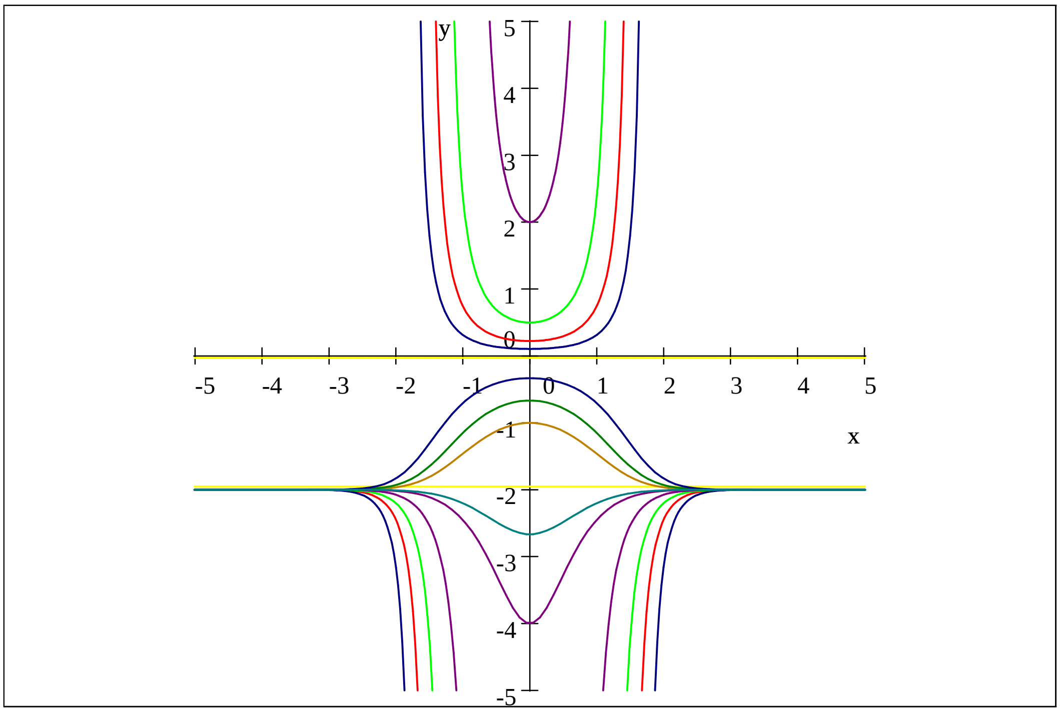
where C is any real number, whence it follows that

$$y = \frac{2Ce^{x^2}}{1 - Ce^{x^2}}. \tag{1.9}$$

We obtain the following family of solutions

$$\frac{2Ce^{x^2}}{1 - Ce^{x^2}} \quad \text{and} \quad y \equiv -2, \tag{1.10}$$

the integral curves of which are shown on the diagram:



Note that the constructed integral curves never intersect each other. Indeed, by Theorem 1.1, through any point (x_0, y_0) with $y_0 \neq 0, -2$ there is a unique integral curve from the family (1.9) with $C \neq 0$. If $y_0 = 0$ or $y_0 = -2$ then the same is true with the family (1.10), because if $C \neq 0$ then the function (1.9) never takes values 0 and -2. This implies as in the previous Section that we have found all the solutions so that (1.10) represent the general solution.

Let us show how to find a particular solution that satisfies the prescribed initial condition, say $y(0) = -4$. Substituting $x = 0$ and $y = -4$ into (1.9), we obtain an equation for C:

$$\frac{2C}{1-C} = -4$$

whence $C = 2$. Hence, the particular solution is

$$y = \frac{4e^{x^2}}{1-2e^{x^2}}.$$

Example. Consider the equation

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

which is defined for all $x, y \in \mathbb{R}$. 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

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

(the restriction $x > -C$ comes from the previous line). Similarly, in the domain $y < 0$, we obtain

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

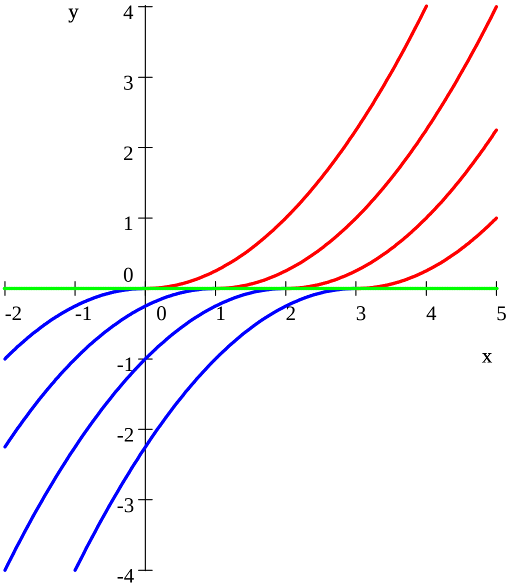
whence

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

and

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

We obtain the following integrals curves:



We see that the integral curves in the domains $y > 0$ and $y < 0$ touch the line $y = 0$. This allows us to construct more solutions as follows: for any couple of reals $a < b$, consider the function

$$y(x) = \begin{cases} -\frac{1}{4}(x-a)^2, & x < a, \\ 0, & a \le x \le b, \\ \frac{1}{4}(x-b)^2, & x > b, \end{cases} \tag{1.11}$$

which is obviously a solution with the domain $\mathbb{R}$. If we allow a to be $-\infty$ and b to be $+\infty$ with the obvious meaning of (1.11) in these cases, then (1.11) represents the general solution to $y' = \sqrt{|y|}$. Clearly, through any point $(x_0, y_0) \in \mathbb{R}^2$ there are infinitely many integral curves of the given equation.

1.2 Linear ODE of 1st order

Consider the ODE of the form

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

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* (1.12) *has the form*

$$y(x) = e^{-A(x)} \int b(x) e^{A(x)} dx, \tag{1.13}$$

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

Note that the function $y(x)$ given by (1.13) 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,

$$y(x) = u(x) e^{-A(x)}. \tag{1.14}$$

Substituting this to the equation (1.12) 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 (1.14), we finish the proof. ■

One may wonder how one can guess to make the change (1.14). Here is the motivation. Consider first the case when $b(x) \equiv 0$. In this case, the equation (1.12) 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 (1.12) take the above solution to the homogeneous equation and replace a constant C by a function $C(x)$ (or which was denoted by $u(x)$ in the proof), 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.

Example. Consider the equation

$$y' + \frac{1}{x} y = e^{x^2} \tag{1.15}$$

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 $a(x)$),

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

where C is an arbitrary constant.

Alternatively, one can solve first the homogeneous equation

$$y' + \frac{1}{x} y = 0,$$

using the separable of variables:

$$\begin{aligned} \frac{y'}{y} &= -\frac{1}{x} \\ (\ln y)' &= -(\ln x)' \\ \ln y &= -\ln x + C_1 \\ y &= \frac{C}{x}. \end{aligned}$$

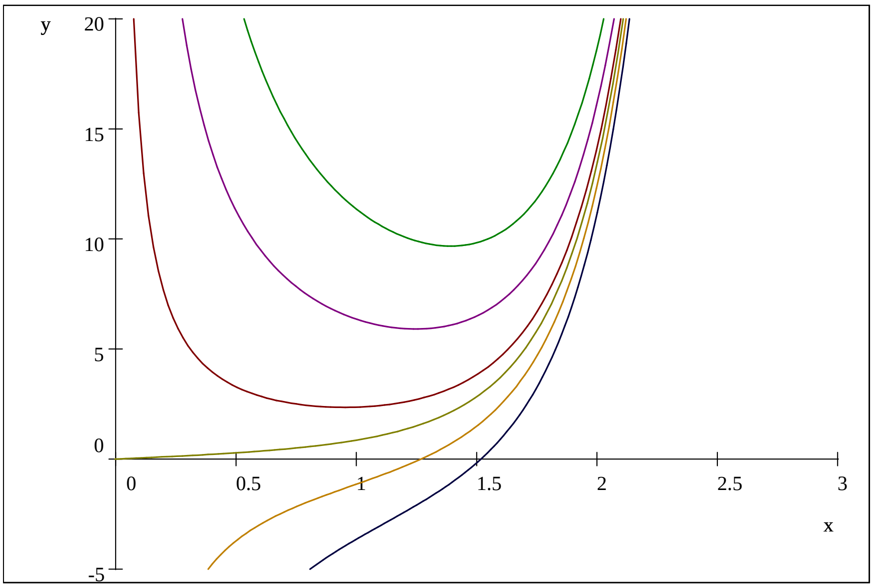
Next, replace the constant C by a function $C(x)$ and substitute into (1.15):

$$\begin{aligned} \left(\frac{C(x)}{x} \right)' + \frac{1}{x} \frac{C}{x} &= e^{x^2}, \\ \frac{C'x - C}{x^2} + \frac{C}{x^2} &= e^{x^2} \\ \frac{C'}{x} &= e^{x^2} \\ C' &= e^{x^2} x \\ C(x) &= \int e^{x^2} x dx = \frac{1}{2} \left(e^{x^2} + C_0 \right). \end{aligned}$$

Hence,

$$y = \frac{C(x)}{x} = \frac{1}{2x} \left(e^{x^2} + C_0 \right),$$

where C_0 is an arbitrary constant. The integral curves are shown on the following diagram:



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.‘ ■

1.3 Quasi-linear ODEs and 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 called differentiable at a point $(x, y) \in \Omega$ if there exist real numbers a, b such that

$$F(x + dx, y + dy) - F(x, y) = adx + bdy + o(|dx| + |dy|),$$

as $|dx| + |dy| \to 0$. Here dx and dy the increments of x and y, respectively, which are considered as new independent variables (the differentials). The linear function $adx + bdy$ of the variables dx, dy is called the differential of F at (x, y) and is denoted by dF, that is,

$$dF = adx + bdy. \tag{1.16}$$

In general, a and b are functions of (x, y).

Recall also the following relations between the notion of a differential and partial derivatives:

1. If F is differentiable at some point (x, y) and its differential is given by (1.16) then the partial derivatives $F_x = \frac{\partial F}{\partial x}$ and $F_y = \frac{\partial F}{\partial y}$ exist at this point and
$$F_x = a, \qquad F_y = b.$$
2. If F is continuously differentiable in Ω, that is, the partial derivatives F_x and F_y exist in Ω and are continuous functions, then F is differentiable at any point in Ω.

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

$$dF = adx + bdy, \tag{1.17}$$

and *inexact* otherwise. If the form is exact then the function F from (1.17) is called the *integral* of the form.

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

Lemma 1.3 *If functions a, b are continuously differentiable in Ω then the necessary condition for the form $adx + bdy$ to be exact is the identity*

$$a_y = b_x. \tag{1.18}$$

Proof. Indeed, if F is an integral of the form $adx + bdy$ then $F_x = a$ and $F_y = b$, whence it follows that the derivatives F_x and F_y are continuously differentiable. By a Schwarz theorem from Analysis, this implies that $F_{xy} = F_{yx}$ whence $a_y = b_x$. ■

Definition. The differential form $adx + bdy$ is called *closed*[10] in Ω if it satisfies the condition $a_y = b_x$ in Ω.

Hence, Lemma 1.3 says that any exact form must be closed. The converse is in general not true, as will be shown later on. However, since it is easier to verify the closedness than the exactness, it is desirable to know under what additional conditions the closedness implies the exactness. Such a result will be stated and proved below.

Example. The form $ydx - xdy$ is not closed because $a_y = 1$ while $b_x = -1$. Hence, it is inexact.

The form $ydx + xdy$ is exact because it has an integral $F(x, y) = xy$. Hence, it is also closed, which can be easily verified directly by (1.18).

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). Again, the closedness can be verified by a straight differentiation, while the exactness requires construction (or guessing) of the integral.

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

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

as it is stated in the next Theorem.

The ODE (1.19) is called *quasi-linear* because it is linear with respect to y' but not necessarily linear with respect to y. Using $y' = \frac{dy}{dx}$, one can write (1.19) in the form

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

which explains why the equation (1.19) is related to the differential form $adx + bdy$. We say that the equation (1.19) is exact (or closed) if the form $adx + bdy$ is exact (or closed).

Theorem 1.4 *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 and let $y(x)$ be a differentiable function defined on an interval $I \subset \mathbb{R}$ such that $(x, y(x)) \in \Omega$ for any $x \in I$ (that is, the graph of y is contained in Ω). Then y is a solution of the equation* (1.19) *if and only if*

$$F(x, y(x)) = \text{const} \ \ \textit{on } I \tag{1.20}$$

(that is, if function F remains constant on the graph of y).

The identity (1.20) can be considered as (an implicit form of) the general solution of (1.19). The function F is also referred to as the integral of the ODE (1.19).

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

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

[10] geschlossen

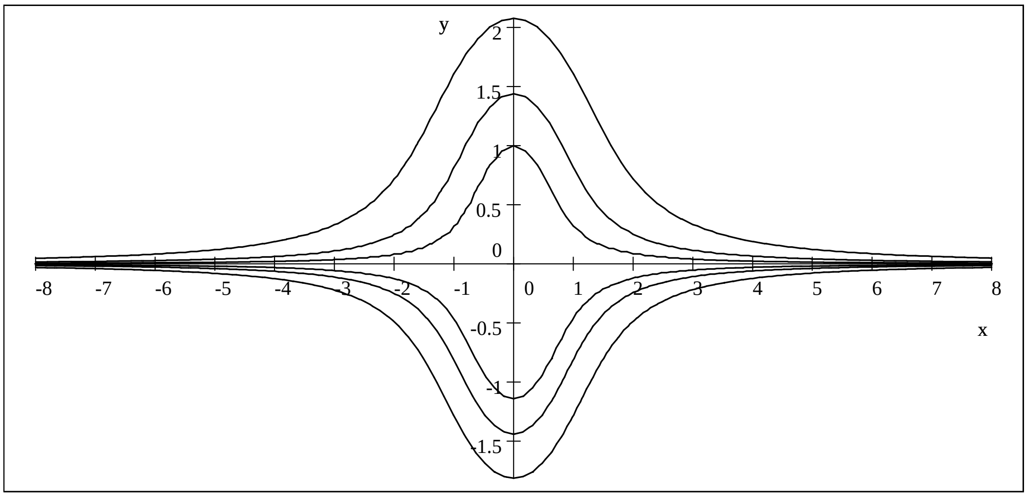
Hence, the equation $a + by' = 0$ is equivalent to $\frac{d}{dx}F(x, y(x)) = 0$ on I, and the latter is equivalent to $F(x, y(x)) = \text{const}$. ■

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, hence, is equivalent to

$$x^2 y + \frac{y^3}{3} = C,$$

because the left hand side is the integral of the corresponding differential form (cf. the previous Example). Some integral curves of this equation are shown on the diagram:

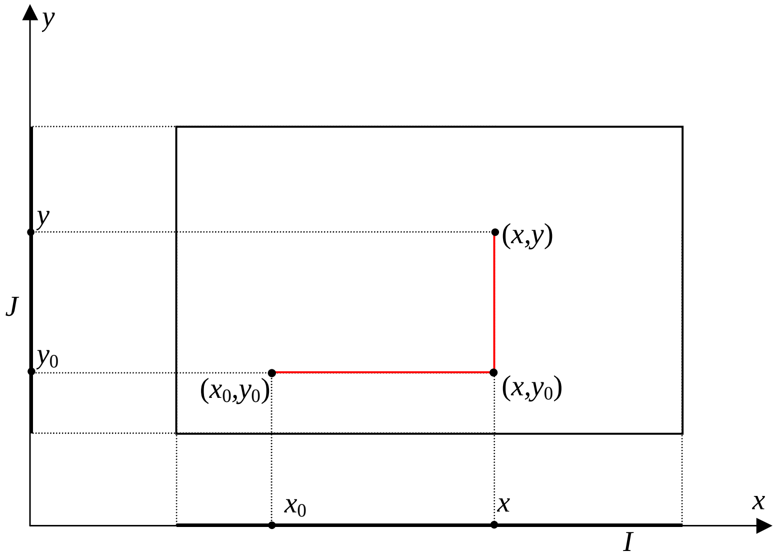


We say that a set $\Omega \subset \mathbb{R}^2$ is a *rectangle* (box) if it has the form $I \times J$ where I and J are intervals in $\mathbb{R}$. A rectangle is open if both I and J are open intervals. The following theorem provides the answer to the question how to decide whether a given differential form is exact in a rectangle.

Theorem 1.5 (The Poincaré lemma) *Let Ω be an open rectangle in $\mathbb{R}^2$. Let a, b be continuously differentiable functions on Ω such that $a_y \equiv b_x$. Then the differential form $adx + bdy$ is exact in Ω.*

It follows from Lemma 1.3 and Theorem 1.5 that if Ω is a rectangle then the form $adx + bdy$ is exact in Ω if and only if it is closed.

Proof. First we try to obtain an explicit formula for the integral F assuming that it exists. Then we use this formula to prove the existence of the integral. Fix some reference point (x_0, y_0) and assume without loss of generality that and $F(x_0, y_0) = 0$ (this can always be achieved by adding a constant to F). For any point $(x, y) \in \Omega$, also the point (x, y_0) belongs Ω; 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 (see the diagram).



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

$$F(x, y) = \int_{x_0}^{x} a(s, y_0)\, ds + \int_{y_0}^{y} b(x, t)\, dt. \tag{1.21}$$

Now we start the actual proof where we assume that the form $adx + bdy$ is closed and use the formula (1.21) to *define* function $F(x, y)$. We need to show that F is the integral; this will not only imply that the form $adx + bdy$ is exact but will give an effective way of evaluating the integral,

Due to the continuous differentiability of a and b, in order to prove that F is indeed the integral of the form $adx + bdy$, it suffices to verify the identities

$$F_x = a \quad \text{and} \quad F_y = b.$$

It is easy to see from (1.21) that

$$F_y = \frac{\partial}{\partial y} \int_{y_0}^{y} b(x, t)\, dt = b(x, y).$$

Next, we have

$$\begin{aligned} F_x &= \frac{\partial}{\partial x} \int_{x_0}^{x} a(s, y_0)\, ds + \frac{\partial}{\partial x} \int_{y_0}^{y} b(x, t)\, dt \\ &= a(x, y_0) + \int_{y_0}^{y} \frac{\partial}{\partial x} b(x, t)\, dt. \end{aligned} \tag{1.22}$$

The fact that the integral and the derivative $\frac{\partial}{\partial x}$ can be interchanged will be justified below (see Lemma 1.6). Using the hypothesis $b_x = a_y$, we obtain from (1.22)

$$\begin{aligned} F_x &= a(x, y_0) + \int_{y_0}^{y} a_y(x,t)\,dt \\ &= a(x,y_0) + (a(x,y) - a(x,y_0)) \\ &= a(x,y)\,, \end{aligned}$$

which finishes the proof. ■

Now we state and prove a lemma that justifies (1.22).

Lemma 1.6 *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 hence 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 $g(x,t)$, are interchangeable.

Proof of Lemma 1.6. We need to show that, for all $x \in I$,

$$\frac{f(y) - f(x)}{y - x} \to \int_{\alpha}^{\beta} g_x(x,t)\,dt \text{ as } y \to x,$$

which amounts to

$$\int_{\alpha}^{\beta} \frac{g(y,t) - g(x,t)}{y - x} dt \to \int_{\alpha}^{\beta} g_x(x,t)\,dt \text{ as } y \to x.$$

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

$$\frac{g(y,t) - g(x,t)}{y - x} \to g_x(x,t) \text{ as } y \to x. \tag{1.23}$$

Consider all parts of (1.23) as functions of t, with fixed x and with y 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* (*gleichmässig*) in the whole interval $[\alpha, \beta]$, that is, if

$$\sup_{t \in [\alpha,\beta]} \left| \frac{g(y,t) - g(x,t)}{y - x} - g_x(x,t) \right| \to 0 \quad \text{as } y \to x. \tag{1.24}$$

By the mean value theorem, for any $t \in [\alpha, \beta]$, there is $\xi \in [x, y]$ such that

$$\frac{g(y,t) - g(x,t)}{y - x} = g_x(\xi, t)\,.$$

Hence, the difference quotient in (1.24) can be replaced by $g_x(\xi, t)$. To proceed further, recall that a continuous function on a compact set is uniformly continuous. In particular, the function $g_x(x,t)$ is uniformly continuous on $I \times J$, that is, for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$x, \xi \in I, |x - \xi| < \delta \text{ and } t, s \in J, |t - s| < \delta \Rightarrow |g_x(x,t) - g_x(\xi, s)| < \varepsilon. \tag{1.25}$$

If $|x - y| < \delta$ then also $|x - \xi| < \delta$ and, by (1.25) with $s = t$,

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

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

$$\sup_{t \in J} \left| \frac{g(y,t) - g(x,t)}{y - x} - g_x(x,t) \right| \leq \varepsilon,$$

whence (1.24) follows. ■

Consider some examples to Theorem 1.5.

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

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

we conclude by Theorem 1.5 that the given form is exact. The integral F can be found by (1.21) 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

$$\frac{-ydx + xdy}{x^2 + y^2} \tag{1.26}$$

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 1.5 we conclude that the given form is exact in any rectangular subdomain of Ω. However, Ω itself is not a rectangle, and let us show that the form is inexact in Ω. To that end, 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'

$$d\theta = \frac{-ydx + xdy}{x^2 + y^2}, \tag{1.27}$$

that is, θ is the integral of (1.26) in Ω'. 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 (1.27) follows by differentiation of the arctan:

$$d\theta = \frac{1}{1 + (y/x)^2}\frac{xdy - ydx}{x^2} = \frac{-ydx + xdy}{x^2 + y^2}.$$

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 (1.27) 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 (1.27) follows again. Since Ω' is the union of the three half-planes $\{x > 0\}$, $\{y > 0\}$, $\{y < 0\}$, we conclude that (1.27) holds in Ω' and, hence, the form (1.26) is exact in Ω'.

Now we can prove that the form (1.26) is inexact in Ω. Assume from the contrary that it 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[11] $F = \theta + \text{const}$ 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: π and $-\pi$ respectively.

The moral of this example is that the statement of Theorem 1.5 is not true for an arbitrary open set Ω. It is possible to show that the statement of Theorem 1.5 is true if and only if the set Ω is *simply connected*, that is, if any closed curve in Ω can be continuously deformed to a point while staying in Ω. Obviously, the rectangles are simply connected (as well as Ω'), while the set $\Omega = \mathbb{R}^2 \setminus \{0\}$ is not simply connected.

[11] We use the following fact 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.

1.4 Integrating factor

Consider again the quasilinear equation

$$a(x,y) + b(x,y)\,y' = 0 \tag{1.28}$$

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 an 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 (1.28) 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 (1.28) by M the problem amounts to the case of Theorem 1.4.

Example. Consider the ODE

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

in the domain $\{x > 0, y > 0\}$ and write it in the form

$$ydx - \left(4x^2y + x\right)dy = 0.$$

Clearly, this equation is not exact. However, dividing it by x^2, we obtain the equation

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

which is already exact in any rectangular domain because

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

Taking in (1.21) $x_0 = y_0 = 1$, we obtain the integral of the form as follows:

$$F(x,y) = \int_1^x \frac{1}{s^2}ds - \int_1^y \left(4t + \frac{1}{x}\right)dt = 3 - 2y^2 - \frac{y}{x}.$$

By Theorem 1.4, the general solution is given by the identity

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

There are no regular methods for finding the integrating factors.

1.5 Second order ODE

For higher order ODEs we will use different notation: the independent variable will be denoted by t and the unknown function by $x(t)$. In this notation, a general second order ODE, resolved with respect to x'' has the form

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

where f is a given function of three variables. 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 its coordinate at time t be $x(t)$. The velocity[12] of the particle is $v(t) = x'(t)$ and the acceleration[13] is $a(t) = x''(t)$. The Newton's second law says that at any time

$$mx'' = F, \tag{1.29}$$

where m is the mass of the particle and F is the force[14] acting on the particle. In general, F is a function of t, x, x', that is, $F = F(t, x, x')$ so that (1.29) can be regarded as a second order ODE for $x(t)$.

The force F is called *conservative* if F depends only on the position x. For example, the gravitational, elastic, and electrostatic forces are conservative, while friction[15] and the air resistance[16] are non-conservative as they depend on the velocity. Assuming that $F = F(x)$, let $U(x)$ be a primitive function of $-F(x)$. The function U is called the *potential* of the force F. Multiplying the equation (1.29) by x' and integrating in t, we obtain

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

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

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

and

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

The sum $\frac{mv^2}{2} + U(x)$ is called the *mechanical energy* of the particle (which is the sum of the *kinetic* energy and the *potential* energy). Hence, we have obtained the *law of conservation of energy*[17]: the total mechanical energy of the particle in a conservative field remains constant.

[12] Geschwindigkeit

[13] Beschleunigung

[14] Kraft

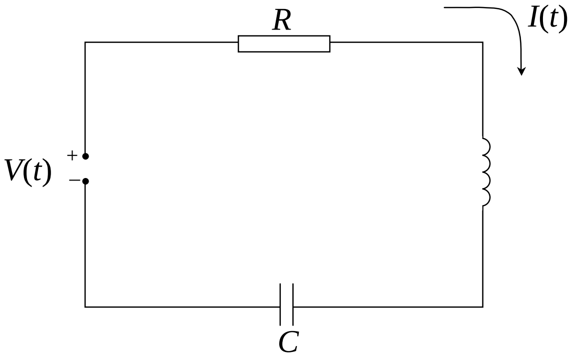
[15] Reibung

[16] Strömungswiderstand

[17] Energieerhaltungssatz

1.5.2 Electrical circuit

Consider an RLC-circuit that is, an electrical circuit[18] where a resistor, an inductor and a capacitor are connected in a series:



Denote by R the resistance[19] of the resistor, by L the inductance[20] of the inductor, and by C the capacitance[21] of the capacitor. Let the circuit contain a power source with the voltage $V(t)$[22] depending in time t. Denote by $I(t)$ the current[23] in the circuit at time t. Using the laws of electromagnetism, we obtain that the voltage drop[24] v_R on the resistor R is equal to

$$v_R = RI$$

(Ohm's law[25]), and the voltage drop v_L on the inductor is equal to

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

(Faraday's law). The voltage drop v_C on the capacitor is equal to

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

where Q is the charge[26] of the capacitor; also we have $Q' = I$. By Kirchhoff's law, we obtain

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

whence

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

[18] Schaltung
[19] Widerstand
[20] Induktivität
[21] Kapazität
[22] Spannung
[23] Strom
[24] Spannungsfall
[25] Gesetz
[26] Ladungsmenge

Differentiating in t, we obtain

$$LI'' + RI' + \frac{I}{C} = V', \tag{1.30}$$

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

1.6 Higher order ODE and normal systems

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

$$y^{(n)} = F\left(t, y, ..., y^{(n-1)}\right), \tag{1.31}$$

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

Let $x\left(t\right)$ 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 $x'\left(t\right)$ is defined component-wise by

$$x' = \left(x_1', x_2', ..., x_n'\right).$$

Consider now a *vector ODE of the first order*

$$x' = f\left(t, x\right), \tag{1.32}$$

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[27] of $\mathbb{R}^{n+1}$. Here the couple (t, x) is identified with a point in $\mathbb{R}^{n+1}$ as follows:

$$(t, x) = \left(t, x_1, ..., x_n\right).$$

Denoting by f_k the components of f, we can rewrite the vector equation (1.32) as a system of n scalar equations

$$\begin{cases} x_1' = f_1\left(t, x_1, ..., x_n\right) \\ ... \\ x_k' = f_k\left(t, x_1, ..., x_n\right) \\ ... \\ x_n' = f_n\left(t, x_1, ..., x_n\right) \end{cases} \tag{1.33}$$

A system of ODEs of the form (1.32) or (1.33) is called the *normal system.* As in the case of the scalar ODEs, define a solution of the normal system (1.32) as a differentiable function $x : I \to \mathbb{R}^n$, where I is an interval in $\mathbb{R}$, such that $\left(t, x\left(t\right)\right) \in \Omega$ for all $t \in I$ and $x'\left(t\right) = f\left(t, x\left(t\right)\right)$ for all $t \in I$.

Let us show how the equation (1.31) can be reduced to the normal system (1.33). Indeed, with any function $y\left(t\right)$ 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

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

[27] Teilmenge

Obviously,

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

and using (1.31) we obtain a system of equations

$$\begin{cases} x_1' = x_2 \\ x_2' = x_3 \\ ... \\ x_{n-1}' = x_n \\ x_n' = F(t, x_1, ...x_n) \end{cases} \tag{1.34}$$

Obviously, we can rewrite this system as a vector equation (1.32) with the function f as follows:

$$f(t, x) = (x_2, x_3, ..., x_n, F(t, x_1, ..., x_n)). \tag{1.35}$$

Conversely, the system (1.34) implies

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

so that we obtain equation (1.31) with respect to $y = x_1$. Hence, the equation (1.31) is equivalent to the vector equation (1.32) with function f defined by (1.35).

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 normal system (1.32) with

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

As we have seen in the case of the first order scalar ODE, adding the initial condition allows usually to uniquely identify the solution. The problem of finding a solution satisfying the initial condition is called the *initial value problem*, shortly IVP. What initial value problem is associated with the vector equation (1.32) and the scalar higher order equation (1.31)? Motivated by the 1st order scalar ODE, one can presume that it makes sense to consider the following IVP also for the 1st order vector ODE:

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

where $(t_0, x_0) \in \Omega$ is a given point. In particular, x_0 is a given vector from $\mathbb{R}^n$, which is called the initial value of $x(t)$, and $t_0 \in \mathbb{R}$ is the initial time. For the equation

(1.31), 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:

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

1.7 Some Analysis in $\mathbb{R}^n$

Here we briefly revise some necessary facts from Analysis in $\mathbb{R}^n$.

1.7.1 Norms in $\mathbb{R}^n$

Recall that a *norm* in $\mathbb{R}^n$ is a function $N : \mathbb{R}^n \to [0, +\infty)$ with the following properties:

1. $N(x) = 0$ if and only if $x = 0$.
2. $N(cx) = |c| N(x)$ for all $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$.
3. $N(x + y) \leq N(x) + N(y)$ for all $x, y \in \mathbb{R}^n$.

If $N(x)$ satisfies 2 and 3 but not necessarily 1 then $N(x)$ is called a *semi-norm*.

For example, the function $|x|$ is a norm in $\mathbb{R}$. Usually one uses the notation $\|x\|$ for a norm instead of $N(x)$.

Example. For any $p \geq 1$, the *p-norm* in $\mathbb{R}^n$ is defined by

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}. \tag{1.36}$$

In particular, for $p = 1$ we have

$$\|x\|_1 = \sum_{k=1}^{n} |x_k|,$$

and for $p = 2$

$$\|x\|_2 = \left(\sum_{k=1}^{n} x_k^2\right)^{1/2}.$$

For $p = \infty$ set

$$\|x\|_\infty = \max_{1 \leq k \leq n} |x_k|.$$

It is known that the p-norm for any $p \in [1, \infty]$ is indeed a norm.

It follows from the definition of a norm that in $\mathbb{R}$ any norm has the form $\|x\| = c\,|x|$ where c is a positive constant. In $\mathbb{R}^n$, $n \geq 2$, there is a great variety of non-proportional norms. However, it is known that all possible norms in $\mathbb{R}^n$ are equivalent in the following sense: if $N_1(x)$ and $N_2(x)$ are two norms in $\mathbb{R}^n$ then there are positive constants C' and C'' such that

$$C'' \leq \frac{N_1(x)}{N_2(x)} \leq C' \text{ for all } x \neq 0. \tag{1.37}$$

For example, it follows from the definitions of $\|x\|_p$ and $\|x\|_\infty$ that

$$1 \leq \frac{\|x\|_p}{\|x\|_\infty} \leq n^{1/p}.$$

For most applications, the relation (1.37) means that the choice of a specific norm is unimportant.

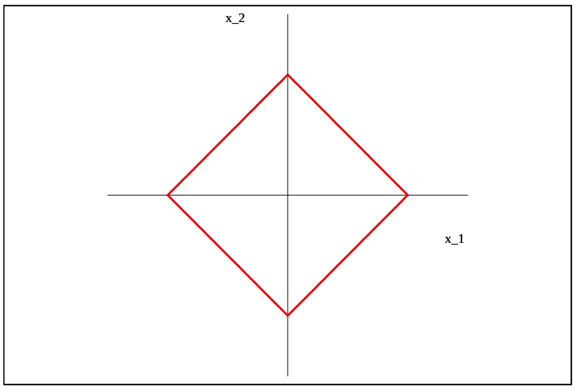
Fix a norm $\|x\|$ in $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$ and $r > 0$, define an *open ball*

$$B(x, r) = \{y \in \mathbb{R}^n : \|x - y\| < r\},$$

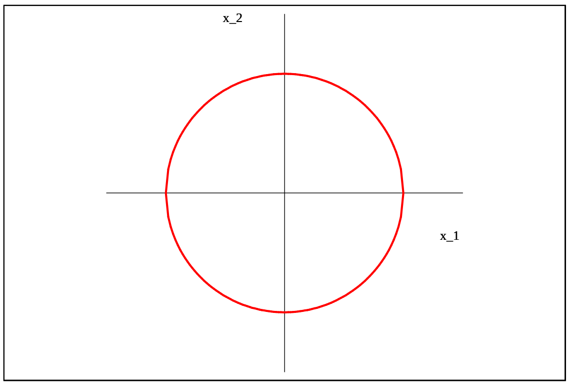
and a *closed ball*

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

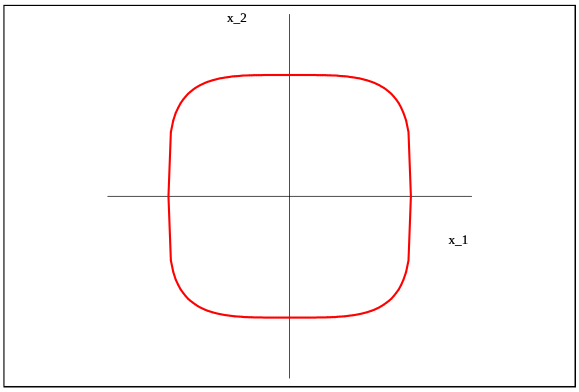
For example, in $\mathbb{R}$ with $\|x\| = |x|$ we have $B(x, r) = (x - r, x + r)$ and $\overline{B}(x, r) = [x - r, x + r]$. Below are sketches of the ball $B(0, r)$ in $\mathbb{R}^2$ for different norms:
the 1-norm (a diamond ball):



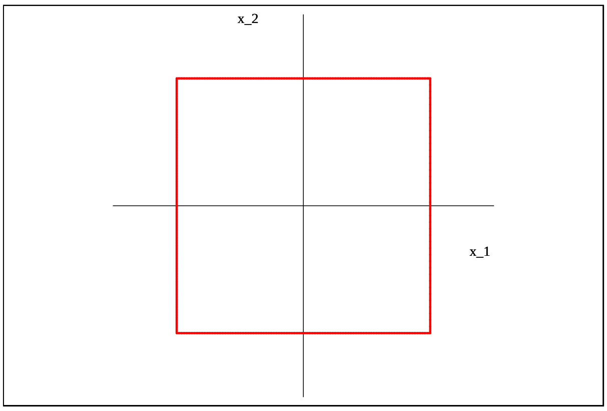
the 2-norm (a round ball):



the 4-norm:



the ∞-norm (a square ball):



1.7.2 Continuous mappings

Let S be a subset of $\mathbb{R}^n$ and f be a mapping[28] from S to $\mathbb{R}^m$. The mapping f is called continuous[29] at $x \in S$ if $f(y) \to f(x)$ as $y \to x$ that is,

$$\|f(y) - f(x)\| \to 0 \text{ as } \|y - x\| \to 0.$$

Here in the expression $\|y - x\|$ we use a norm in $\mathbb{R}^n$ whereas in the expression $\|f(y) - f(x)\|$ we use a norm in $\mathbb{R}^m$, where the norms are arbitrary, but fixed. Thanks to the equivalence of the norms in the Euclidean spaces, the definition of a continuous mapping does not depend on a particular choice of the norms.

Note that any norm $N(x)$ in $\mathbb{R}^n$ is a continuous function because by the triangle inequality

$$|N(y) - N(x)| \leq N(y - x) \to 0 \text{ as } y \to x.$$

Recall that a set $S \subset \mathbb{R}^n$ is called

- open[30] if for any $x \in S$ there is $r > 0$ such that the ball $B(x, r)$ is contained in S;
- closed[31] if S^c is open;
- compact if S is bounded and closed.

An important fact from Analysis is that if S is a compact subset of $\mathbb{R}^n$ and $f : S \to \mathbb{R}^m$ is a continuous mapping then the image[32] $f(S)$ is a compact subset of $\mathbb{R}^m$. In particular, of f is a continuous numerical functions on S, that is $f : S \to \mathbb{R}$ then f is bounded on S and attains on S its maximum and minimum.

1.7.3 Linear operators and matrices

Let us consider a special class of mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$ that are called *linear operators* or linear mappings. Namely, a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping with the following properties:

1. $A(x + y) = Ax + Ay$ for all $x, y \in \mathbb{R}^n$;
2. $A(cx) = cAx$ for all $c \in \mathbb{R}$ and $x \in \mathbb{R}^n$.

Any linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented by a $m \times n$ matrix (a_{ij}) where i is the row[33] index, $i = 1, ..., m$, and j is the column[34] index $j = 1, ..., n$, as follows:

$$(Ax)_i = \sum_{j=1}^{n} a_{ij} x_j$$

[28] Abbildung

[29] stetig

[30] offene Menge

[31] abgeschlossene Menge

[32] Das Bild

[33] Zeile

[34] Spalte

(the elements a_{ij} of the matrix (a_{ij}) are called the components of the operator A). In other words, the column vector Ax is obtained from the column-vector x by the left multiplication with the matrix (a_{ij}):

$$\begin{pmatrix} (Ax)_1 \\ \dots \\ (Ax)_i \\ \dots \\ (Ax)_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & \dots & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & a_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \dots \\ x_j \\ \dots \\ x_n \end{pmatrix}$$

The class of all linear operators from $\mathbb{R}^n$ to $\mathbb{R}^m$ is denoted by $\mathbb{R}^{m\times n}$ or by $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. One defines in the obvious way the addition $A+B$ of operators in $\mathbb{R}^{m\times n}$ and the multiplication by a constant cA with $c \in \mathbb{R}$:

1. $(A+B)(x) = Ax + Bx,$
2. $(cA)(x) = c(Ax).$

Clearly, $\mathbb{R}^{m\times n}$ with these operations is a linear space over $\mathbb{R}$. Since any $m \times n$ matrix has mn components, it can be identified with an element in $\mathbb{R}^{mn}$; consequently, $\mathbb{R}^{m\times n}$ is linearly isomorphic to $\mathbb{R}^{mn}$.

Let us fix some norms in $\mathbb{R}^n$ and $\mathbb{R}^m$. For any linear operator $A \in \mathbb{R}^{m\times n}$, define its *operator norm* by

$$\|A\| = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|}{\|x\|}, \tag{1.38}$$

where $\|x\|$ is the norm in $\mathbb{R}^n$ and $\|Ax\|$ is the norm in $\mathbb{R}^m$. We claim that always $\|A\| < \infty$. Indeed, it suffices to verify this if the norm in $\mathbb{R}^n$ is the 1-norm and the norm in $\mathbb{R}^m$ is the ∞-norm. Using the matrix representation of the operator A as above, we obtain

$$\|Ax\|_\infty = \max_i |(Ax)_i| \leq \max_{i,j} |a_{ij}| \sum_{j=1}^n |x_j| = a\|x\|_1$$

where $a = \max_{i,j} |a_{ij}|$. It follows that $\|A\| \leq a < \infty$.

Hence, one can define $\|A\|$ as the minimal possible real number such that the following inequality is true:

$$\|Ax\| \leq \|A\| \|x\| \text{ for all } x \in \mathbb{R}^n.$$

As a consequence, we obtain that any linear mapping $A \in \mathbb{R}^{m\times n}$ is continuous, because

$$\|Ay - Ax\| = \|A(y-x)\| \leq \|A\| \|y - x\| \to 0 \text{ as } y \to x.$$

Let us verify that the operator norm is a norm in the linear space $\mathbb{R}^{m\times n}$. Indeed, by (1.38) we have $\|A\| \geq 0$; moreover, if $A \neq 0$ then there is $x \in \mathbb{C}^n$ such that $Ax \neq 0$, whence $\|Ax\| > 0$ and

$$\|A\| \geq \frac{\|Ax\|}{\|x\|} > 0.$$

The triangle inequality can be deduced from (1.38) as follows:

$$\begin{aligned}\|A+B\| = & \sup_x \frac{\|(A+B)\,x\|}{\|x\|} \leq \sup_x \frac{\|Ax\|+\|Bx\|}{\|x\|} \\ \leq & \sup_x \frac{\|Ax\|}{\|x\|} + \sup_x \frac{\|Bx\|}{\|x\|} \\ = & \|A\|+\|B\|.\end{aligned}$$

Finally, the scaling property trivially follows from (1.38):

$$\|\lambda A\| = \sup_x \frac{\|(\lambda A)\,x\|}{\|x\|} = \sup_x \frac{|\lambda|\,\|Ax\|}{\|x\|} = |\lambda|\,\|A\|.$$

Similarly, one defines the notion of a norm in $\mathbb{C}^n$, the space of linear operators $\mathbb{C}^{m\times n}$ and the operator norm in $\mathbb{C}^{m\times n}$. All the properties of the norms in the complex spaces can be either proved in the same way as in the real spaces, or simply deduced from the real case by using the natural identification of $\mathbb{C}^n$ with $\mathbb{R}^{2n}$.

2 Linear equations and systems

A normal linear system of ODEs is the following vector equation

$$x' = A(t)\,x + B(t)\,, \tag{2.1}$$

where $A : I \to \mathbb{R}^{n\times n}$, $B : I \to \mathbb{R}^n$, I is an interval in $\mathbb{R}$, and $x = x(t)$ is an unknown function with values in $\mathbb{R}^n$.

In other words, for each $t \in I$, $A(t)$ is a linear operator in $\mathbb{R}^n$, while $A(t)\,x$ and $B(t)$ are the vectors in $\mathbb{R}^n$. In the coordinate form, (2.1) is equivalent to the following system of linear equations

$$x_i' = \sum_{l=1}^{n} A_{ij}(t)\,x_j + B_i(t)\,, \quad i = 1, ..., n,$$

where A_{ij} and B_i are the components of A and B, respectively. We will consider the ODE (2.1) only when $A(t)$ and $B(t)$ are continuous in t, that is, when the mappings $A : I \to \mathbb{R}^{n\times n}$ and $B : I \to \mathbb{R}^n$ are continuous. It is easy to show that the continuity of these mappings is equivalent to the continuity of all the components $A_{ij}(t)$ and $B_i(t)$, respectively

2.1 Existence of solutions for normal systems

Theorem 2.1 *In the above notation, let $A(t)$ and $B(t)$ be continuous in $t \in I$. Then, for any $t_0 \in I$ and $x_0 \in \mathbb{R}^n$, the IVP*

$$\begin{cases} x' = A(t)\,x + B(t) \\ x(t_0) = x_0 \end{cases} \tag{2.2}$$

has a unique solution $x(t)$ defined on I, and this solution is unique.

Before we start the proof, let us prove the following useful lemma.

Lemma 2.2 (The 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*

$$z(t) \leq C + L \int_{t_0}^{t} z(s)\, ds \tag{2.3}$$

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

$$z(t) \leq C \exp\left(L(t - t_0)\right) \tag{2.4}$$

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

Proof. It suffices to prove the statement the case when C is strictly positive, which implies the validity of the statement also in the case $C = 0$. Indeed, if (2.3) holds with $C = 0$ then it holds with any $C > 0$. Therefore, (2.4) 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 (2.3) 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 (2.3) that $z \leq F$ whence

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

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

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

for all $t \in [t_0, t_1]$. It follows that

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

Using again (2.3), that is, $z \leq F$, we obtain (2.4). ■

Proof of Theorem 2.1. Let us fix some bounded closed interval $[\alpha, \beta] \subset I$ such that $t_0 \in [\alpha, \beta]$. The proof consists of the following three parts.

1. the uniqueness of a solution on $[\alpha, \beta]$;
2. the existence of a solution on $[\alpha, \beta]$;
3. the existence and uniqueness of a solution on I.

We start with the observation that if $x\left(t\right)$ is a solution of (2.2) on I then, for all $t\in I$,

$$\begin{aligned} x\left(t\right) &= x_0+\int_{t_0}^{t} x'\left(s\right)ds \\ &= x_0+\int_{t_0}^{t}\left(A\left(s\right)x\left(s\right)+B\left(s\right)\right)ds. \end{aligned} \tag{2.5}$$

In particular, if $x\left(t\right)$ and $y\left(t\right)$ are two solutions of IVP (2.2) on the interval $[a,\beta]$, then they both satisfy (2.5) for all $t\in[\alpha,\beta]$ whence

$$x\left(t\right)-y\left(t\right)=\int_{t_0}^{t}A\left(s\right)\left(x\left(s\right)-y\left(s\right)\right)ds.$$

Setting

$$z\left(t\right)=\left\|x\left(t\right)-y\left(t\right)\right\|$$

and using that

$$\left\|A\left(y-x\right)\right\|\leq\left\|A\right\|\left\|y-x\right\|=\left\|A\right\|z,$$

we obtain

$$z\left(t\right)\leq\int_{t_0}^{t}\left\|A\left(s\right)\right\|z\left(s\right)ds,$$

for all $t\in[t_0,\beta]$, and a similar inequality for $t\in[\alpha,t_0]$ where the order of integration should be reversed. Set

$$a=\sup_{s\in[\alpha,\beta]}\left\|A\left(s\right)\right\|. \tag{2.6}$$

Since $s\mapsto A\left(s\right)$ is a continuous function, $s\mapsto\left\|A\left(s\right)\right\|$ is also continuous; hence, it is bounded on the interval $[\alpha,\beta]$ so that $a<\infty$. It follows that, for all $t\in[t_0,\beta]$,

$$z\left(t\right)\leq a\int_{t_0}^{t}z\left(s\right)ds,$$

whence by Lemma 2.2 $z\left(t\right)\leq 0$ and, hence, $z\left(t\right)=0$. By a similar argument, the same holds for all $t\in[\alpha,t_0]$ so that $z\left(t\right)\equiv 0$ on $[\alpha,\beta]$, which implies the identity of the solutions $x\left(t\right)$ and $y\left(t\right)$ on $[\alpha,\beta]$.

In the second part, consider a sequence $\{x_k\left(t\right)\}_{k=0}^{\infty}$ of functions on $[\alpha,\beta]$ defined inductively by

$$x_0\left(t\right)\equiv x_0$$

and

$$x_k\left(t\right)=x_0+\int_{t_0}^{t}\left(A\left(s\right)x_{k-1}\left(s\right)+B\left(s\right)\right)ds,\quad k\geq 1. \tag{2.7}$$

We will prove that the sequence $\{x_k\}_{k=0}^{\infty}$ converges on $[\alpha,\beta]$ to a solution of (2.2) as $k\to\infty$.

Using the identity (2.7) and

$$x_{k-1}\left(t\right)=x_0+\int_{t_0}^{t}\left(A\left(s\right)x_{k-2}\left(s\right)+B\left(s\right)\right)ds$$

we obtain, for any $k \geq 2$ and $t \in [t_0, \beta]$,

$$\begin{aligned}\|x_k(t) - x_{k-1}(t)\| &\leq \int_{t_0}^{t} \|A(s)\| \|x_{k-1}(s) - x_{k-2}(s)\| \, ds \\ &\leq a \int_{t_0}^{t} \|x_{k-1}(s) - x_{k-2}(s)\| \, ds\end{aligned}$$

where a is defined by (2.6). Denoting

$$z_k(t) = \|x_k(t) - x_{k-1}(t)\|,$$

we obtain the recursive inequality

$$z_k(t) \leq a \int_{t_0}^{t} z_{k-1}(s) \, ds.$$

In order to be able to use it, let us first estimate $z_1(t) = \|x_1(t) - x_0(t)\|$. By definition, we have, for all $t \in [t_0, \beta]$,

$$z_1(t) = \left\| \int_{t_0}^{t} (A(s) x_0 + B(s)) \, ds \right\| \leq b(t - t_0),$$

where

$$b = \sup_{s \in [\alpha, \beta]} \|A(s) x_0 + B(s)\| < \infty.$$

It follows by induction that

$$\begin{aligned} z_2(t) &\leq ab \int_{t_0}^{t} (s - t_0) \, ds = ab \frac{(t - t_0)^2}{2}, \\ z_3(t) &\leq a^2 b \int_{t_0}^{t} \frac{(s - t_0)^2}{2} ds = a^2 b \frac{(t - t_0)^3}{3!}, \\ &\dots \\ z_k(t) &\leq a^{k-1} b \frac{(t - t_0)^k}{k!}. \end{aligned}$$

Setting $c = \max(a, b)$ and using the same argument for $t \in [\alpha, t_0]$, rewrite this inequality in the form

$$\|x_k(t) - x_{k-1}(t)\| \leq \frac{(c|t - t_0|)^k}{k!},$$

for all $t \in [\alpha, \beta]$. Since the series

$$\sum_k \frac{(c|t - t_0|)^k}{k!}$$

is the exponential series and, hence, is convergent for all t, in particular, uniformly[35] in any bounded interval of t, we obtain by the comparison test, that the series

$$\sum_{k=1}^{\infty} \|x_k(t) - x_{k-1}(t)\|$$

[35] gleichmässig

converges uniformly in $t \in [\alpha, \beta]$, which implies that also the series

$$\sum_{k=1}^{\infty} \left(x_k\left(t\right) - x_{k-1}\left(t\right)\right)$$

converges uniformly in $t \in [\alpha, \beta]$. Since the N-th partial sum of this series is $x_N\left(t\right) - x_0$, we conclude that the sequence $\{x_k\left(t\right)\}$ converges uniformly in $t \in [\alpha, \beta]$ as $k \to \infty$. Setting

$$x\left(t\right) = \lim_{k\to\infty} x_k\left(t\right)$$

and passing in the identity

$$x_k\left(t\right) = x_0 + \int_{t_0}^{t} \left(A\left(s\right) x_{k-1}\left(s\right) + B\left(s\right)\right) ds$$

to the limit as $k \to \infty$, obtain

$$x\left(t\right) = x_0 + \int_{t_0}^{t} \left(A\left(s\right) x\left(s\right) + B\left(s\right)\right) ds. \tag{2.8}$$

This implies that $x\left(t\right)$ solves the given IVP (2.2) on $[\alpha, \beta]$. Indeed, $x\left(t\right)$ is continuous on $[\alpha, \beta]$ as a uniform limit of continuous functions; it follows that the right hand side of (2.8) is a differentiable function of t, whence

$$x' = \frac{d}{dt}\left(x_0 + \int_{t_0}^{t} \left(A\left(s\right) x\left(s\right) + B\left(s\right)\right) ds.\right) = A\left(t\right) x\left(t\right) + B\left(t\right).$$

Finally, it is clear from (2.8) that $x\left(t_0\right) = x_0$.

$\boxed{\text{Lecture 6}}$ 04.05.2009 Prof. A. Grigorian, ODE, SS 2009

Having constructed the solution of (2.2) on any bounded closed interval $[\alpha,\beta]\subset I$, let us extend it to the whole interval I as follows. For any interval I, there is an increasing sequence of bounded closed intervals $\{[\alpha_l,\beta_l]\}_{l=1}^{\infty}$ such that their union is I; furthermore, we can assume that $t_0\in[\alpha_l,\beta_l]$ for all l. Denote by $x_l(t)$ be the solution of (2.2) on $[\alpha_l,\beta_l]$. Then $x_{l+1}(t)$ is also the solution of (2.2) on $[\alpha_l,\beta_l]$ whence it follows by the uniqueness statement of the first part that $x_{l+1}(t)=x_l(t)$ on $[\alpha_l,\beta_l]$. Hence, in the sequence $\{x_l(t)\}$ any function is an extension of the previous function to a larger interval, which implies that the function

$$x(t):=x_l(t)\ \text{ if } t\in[\alpha_l,\beta_l]$$

is well-defined for all $t\in I$ and is a solution of IVP (2.2). Finally, this solution is unique on I because the uniqueness takes place in each of the intervals $[\alpha_l,\beta_l]$. $\blacksquare$

2.2 Existence of solutions for higher order linear ODE

Consider a *scalar linear* ODE of the order n, that is, the ODE

$$x^{(n)}+a_1(t)\,x^{(n-1)}+....+a_n(t)\,x=b(t)\,, \tag{2.9}$$

where all functions $a_k(t)\,,b(t)$ are defined on an interval $I\subset\mathbb{R}$.

Following the general procedure, this ODE can be reduced to the normal linear system as follows. Consider the vector function

$$\mathbf{x}(t)=\left(x(t)\,,x'(t)\,,...,x^{(n-1)}(t)\right) \tag{2.10}$$

so that

$$\mathbf{x}_1=x,\quad \mathbf{x}_2=x',...,\ \mathbf{x}_{n-1}=x^{(n-2)},\quad \mathbf{x}_n=x^{(n-1)}.$$

Then (2.9) is equivalent to the system

$$\begin{aligned}
\mathbf{x}_1' &= \mathbf{x}_2\\
\mathbf{x}_2' &= \mathbf{x}_3\\
&\ \ ...\\
\mathbf{x}_{n-1}' &= \mathbf{x}_n\\
\mathbf{x}_n' &= -a_1\mathbf{x}_n-a_2\mathbf{x}_{n-1}-...-a_n\mathbf{x}_1+b
\end{aligned}$$

that is, to the vector ODE

$$\mathbf{x}'=A(t)\,\mathbf{x}+B(t) \tag{2.11}$$

where

$$A=\begin{pmatrix} 0 & 1 & 0 & ... & 0\\ 0 & 0 & 1 & ... & 0\\ ... & ... & ... & ... & ...\\ 0 & 0 & 0 & ... & 1\\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1\end{pmatrix}\ \text{ and }\ B=\begin{pmatrix}0\\0\\...\\0\\b\end{pmatrix}. \tag{2.12}$$

Theorem 2.1 implies then the following.

Corollary. *Assume that all functions $a_k(t), b(t)$ in* (2.9) *are continuous on I. Then, for any $t_0 \in I$ and any vector $(x_0, x_1, ..., x_{n-1}) \in \mathbb{R}^n$ there is a unique solution $x(t)$ of the ODE* (2.9) *with the initial conditions*

$$\begin{cases} x(t_0) = x_0 \\ x'(t_0) = x_1 \\ ... \\ x^{(n-1)}(t_0) = x_{n-1} \end{cases} \tag{2.13}$$

Proof. Indeed, setting $\mathbf{x}_0 = (x_0, x_1, ..., x_{n-1})$, we can rewrite the IVP (2.9)-(2.13) in the form

$$\begin{cases} \mathbf{x}' = A\mathbf{x} + B \\ \mathbf{x}(t_0) = \mathbf{x}_0 \end{cases}$$

which is the IVP considered in Theorem 2.1. ■

2.3 Space of solutions of linear ODEs

The normal linear system $x' = A(t)x + B(t)$ is called *homogeneous* if $B(t) \equiv 0$, and *inhomogeneous* otherwise.

Consider an inhomogeneous normal linear system

$$x' = A(t)x + B(t), \tag{2.14}$$

where $A(t): I \to \mathbb{R}^{n \times n}$ and $B(t): I \to \mathbb{R}^n$ are continuous mappings on an open interval $I \subset \mathbb{R}$, and the associated homogeneous equation

$$x' = A(t)x. \tag{2.15}$$

Denote by $\mathcal{A}$ the set of all solutions of (2.15) defined on I.

Theorem 2.3 (a)*The set $\mathcal{A}$ is a linear space*[36] *over $\mathbb{R}$ and $\dim \mathcal{A} = n$.Consequently, if $x_1(t), ..., x_n(t)$ are n linearly independent*[37] *solutions to* (2.15) *then the general solution of* (2.15) *has the form*

$$x(t) = C_1x_1(t) + ... + C_nx_n(t), \tag{2.16}$$

where $C_1, ..., C_n$ are arbitrary constants.

(b) *If $x_0(t)$ is a particular solution of* (2.14) *and $x_1(t), ..., x_n(t)$ are n linearly independent solutions to* (2.15) *then the general solution of* (2.14) *is given by*

$$x(t) = x_0(t) + C_1x_1(t) + ... + C_nx_n(t). \tag{2.17}$$

Proof. (a) The set of all functions $I \to \mathbb{R}^n$ is a linear space with respect to the operations addition and multiplication by a constant. Zero element is the function which is constant 0 on I. We need to prove that the set of solutions $\mathcal{A}$ 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.

[36] linearer Raum

[37] linear unabhängig

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 $cx \in \mathcal{A}$ for any $c \in \mathbb{R}$. Hence, $\mathcal{A}$ is a linear space.

Fix $t_0 \in I$ and consider the mapping $\Phi : \mathcal{A} \to \mathbb{R}^n$ given by $\Phi(x) = x(t_0)$; that is, $\Phi(x)$ is the value of $x(t)$ at $t = t_0$. This mapping is obviously linear. It is surjective since by Theorem 2.1 for any $v \in \mathbb{R}^n$ there is a solution $x(t)$ with the initial condition $x(t_0) = v$. Also, this mapping is injective because $x(t_0) = 0$ implies $x(t) \equiv 0$ by the uniqueness of the solution. Hence, Φ is a linear isomorphism between $\mathcal{A}$ and $\mathbb{R}^n$, whence it follows that $\dim \mathcal{A} = \dim \mathbb{R}^n = n$.

Consequently, if $x_1, ..., x_n$ are linearly independent functions from $\mathcal{A}$ then they form a basis in $\mathcal{A}$. It follows that any element of $\mathcal{A}$ is a linear combination of $x_1, ..., x_n$, that is, any solution to $x' = A(t)x$ has the form (2.16).

(b) We claim that a function $x(t) : I \to \mathbb{R}^n$ solves (2.14) if and only if the function $y(t) = x(t) - x_0(t)$ solves (2.15). Indeed, the homogeneous equation for y is equivalent to

$$\begin{aligned} (x - x_0)' &= A(x - x_0), \\ x' &= Ax + (x_0' - Ax_0). \end{aligned}$$

Using $x_0' = Ax_0 + B$, we see that the latter equation is equivalent to (2.14). By part (a), the function $y(t)$ solves (2.15) if and only if it has the form

$$y = C_1 x_1(t) + ... + C_n x_n(t), \tag{2.18}$$

whence it follows that all solutions to (2.14) are given by (2.17). ■

Consider now a scalar ODE

$$x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = b(t) \tag{2.19}$$

where all functions $a_1, ..., a_n, f$ are continuous on an interval I, and the associated homogeneous ODE

$$x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = 0. \tag{2.20}$$

Denote by $\widetilde{\mathcal{A}}$ the set of all solutions of (2.20) defined on I.

Corollary. (a) *The set* $\widetilde{\mathcal{A}}$ *is a linear space over* $\mathbb{R}$ *and* $\dim \widetilde{\mathcal{A}} = n$*.Consequently, if* $x_1, ..., x_n$ *are* n *linearly independent solutions of* (2.20) *then the general solution of* (2.20) *has the form*

$$x(t) = C_1 x_1(t) + ... + C_n x_n(t),$$

where $C_1, ..., C_n$ *are arbitrary constants.*

(b) *If* $x_0(t)$ *is a particular solution of* (2.19) *and* $x_1, ..., x_n$ *are* n *linearly independent solutions of* (2.20) *then the general solution of* (2.19) *is given by*

$$x(t) = x_0(t) + C_1 x_1(t) + ... + C_n x_n(t).$$

Proof. (a) The fact that $\widetilde{\mathcal{A}}$ is a linear space is obvious (cf. the proof of Theorem 2.3). To prove that $\dim \widetilde{\mathcal{A}} = n$, consider the associated normal system

$$\mathbf{x}' = A(t)\,\mathbf{x}, \tag{2.21}$$

where

$$\mathbf{x} = \left(x, x', ..., x^{(n-1)}\right) \tag{2.22}$$

and $A(t)$ is given by (2.12). Denoting by $\mathcal{A}$ the space of solutions of (2.21), we see that the identity (2.22) defines a linear mapping from $\widetilde{\mathcal{A}}$ to $\mathcal{A}$. This mapping is obviously injective (if $\mathbf{x}(t) \equiv 0$ then $x(t) \equiv 0$) and surjective, because any solution $\mathbf{x}$ of (2.21) gives back a solution $x(t)$ of (2.20). Hence, $\widetilde{\mathcal{A}}$ and $\mathcal{A}$ are linearly isomorphic. Since by Theorem 2.3 $\dim \mathcal{A} = n$, it follows that $\dim \widetilde{\mathcal{A}} = n$.The rest is obvious.

(b) The proof uses the same argument as in Theorem 2.3 and is omitted. ■

Frequently it is desirable to consider *complex valued* solutions of ODEs (for example, this will be the case in the next section). The above results can be easily generalized in this direction as follows. Consider again a normal system

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

where $x(t)$ is now an unknown function with values in $\mathbb{C}^n$, $A : I \to \mathbb{C}^{n\times n}$ and $B : I \to \mathbb{C}^n$, where I is an interval in $\mathbb{R}$; that is, for any $t \in I$, $A(t)$ is a linear operator in $\mathbb{C}^n$, and $B(t)$ is a vector from $\mathbb{C}^n$. Assuming that $A(t)$ and $B(t)$ are continuous, one obtains the following extension of Theorem 2.1: for any $t_0 \in I$ and $x_0 \in \mathbb{C}^n$, the IVP

$$\begin{cases} x' = Ax + B \\ x(t_0) = x_0 \end{cases} \tag{2.23}$$

has a solution $x : I \to \mathbb{C}^n$, and this solution is unique. Alternatively, the complex valued problem (2.23) can be reduced to a real valued problem by identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$, and the claim follows by a direct application of Theorem 2.1 to the real system of order $2n$.

Also, similarly to Theorem 2.3, one shows that the set of all solutions to the homogeneous system $x' = Ax$ is a linear space over $\mathbb{C}$ and its dimension is n. The same applies to higher order scalar ODE with complex coefficients.

2.4 Solving linear homogeneous ODEs with constant coefficients

Consider the scalar ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = 0, \tag{2.24}$$

where $a_1, ..., a_n$ are real (or complex) constants. We discuss here the methods of constructing of n linearly independent solutions of (2.24), which will give then the general solution of (2.24).

It will be convenient to obtain the complex valued general solution $x(t)$ and then to extract the real valued general solution, if necessary. 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 (2.24) and noticing that $x^{(k)} = \lambda^k e^{\lambda t}$, we obtain after cancellation by $e^{\lambda t}$ the following equation for λ:

$$\lambda^n + a_1 \lambda^{n-1} + ... + a_n = 0.$$

This equation is called the *characteristic equation* of (2.24) and the polynomial $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + ... + a_n$ is called the *characteristic polynomial* of (2.24). Hence, if λ is the root[38] of the characteristic polynomial then the function $e^{\lambda t}$ solves (2.24). We try to obtain in this way n independent solutions.

Theorem 2.4 *If the characteristic polynomial $P(\lambda)$ of* (2.24) *has n distinct complex roots $\lambda_1, ..., \lambda_n$, then the following n functions*

$$e^{\lambda_1 t}, ..., e^{\lambda_n t} \tag{2.25}$$

are linearly independent complex solutions of (2.24). *Consequently, the general complex solution of* (2.24) *is given by*

$$x(t) = C_1 e^{\lambda_1 t} + ... + C_n e^{\lambda_n t}, \tag{2.26}$$

where C_j are arbitrary complex numbers.

Let $a_1, ...a_n$ be reals. If $\lambda = \alpha + i\beta$ is a non-real root of $P(\lambda)$ then $\overline{\lambda} = \alpha - i\beta$ is also a root of $P(\lambda)$, and the functions $e^{\lambda t}$, $e^{\overline{\lambda} t}$ in the sequence (2.25) *can be replaced by the real-valued functions $e^{\alpha t}\cos\beta t$, $e^{\alpha t}\sin\beta t$. By doing so to any couple of complex conjugate roots, one obtains n real-valued linearly independent solutions of* (2.24), *and the general real solution of* (2.24) *is their linear combination with real coefficients.*

Example. Consider the ODE

$$x'' - 3x' + 2x = 0.$$

The characteristic polynomial is $P(\lambda) = \lambda^2 - 3\lambda + 2$, which has the roots $\lambda_1 = 1$ and $\lambda_2 = 2$. Hence, the linearly independent solutions are e^t and e^{2t}, and the general solution is $C_1e^t + C_2e^{2t}$. More precisely, we obtain the general real solution if C_1, C_2 vary in $\mathbb{R}$, and the general complex solution if $C_1, C_2 \in \mathbb{C}$.

[38] Nullstelle

Example. Consider 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 independent 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 $C_1 \cos t + C_2 \sin t$.

Example. Consider the ODE $x''' - x = 0$. The characteristic polynomial is $P(\lambda) = \lambda^3 - 1 = (\lambda - 1)(\lambda^2 + \lambda + 1)$ 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

$$C_1 e^t + e^{-\frac{1}{2}t} \left(C_2 \cos \frac{\sqrt{3}}{2}t + C_3 \sin \frac{\sqrt{3}}{2}t \right).$$

Proof of Theorem 2.4. Since we know already that $e^{\lambda_k t}$ are solutions of (2.24), it suffices to prove that the functions in the list (2.25) are linearly independent. The fact that the general solution has the form (2.26) follows then from Corollary to Theorem 2.3.

Let us prove by induction in n that the functions $e^{\lambda_1 t}, ..., e^{\lambda_n t}$ are linearly independent whenever $\lambda_1, ..., \lambda_n$ are distinct complex numbers. 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 $t \in \mathbb{R}$,

$$C_1 e^{\lambda_1 t} + ... + C_n e^{\lambda_n t} = 0, \tag{2.27}$$

and prove that $C_1 = ... = C_n = 0$. Dividing (2.27) by $e^{\lambda_n t}$ and setting $\mu_j = \lambda_j - \lambda_n$, 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 $C_j \mu_j = 0$ when by $\mu_j \neq 0$ we conclude $C_j = 0$, for all $j = 1, ..., n-1$. Substituting into (2.27), we obtain also $C_n = 0$.

Let $a_1, .., a_n$ be reals. Since the complex conjugations commutes with addition and multiplication of numbers, the identity $P(\lambda) = 0$ implies $P(\overline{\lambda}) = 0$ (since a_k are real, we have $\overline{a}_k = a_k$). Next, we have

$$e^{\lambda t} = e^{\alpha t} (\cos \beta t + i \sin \beta t) \text{ and } e^{\overline{\lambda} t} = e^{\alpha t} (\cos \beta t - \sin \beta t) \tag{2.28}$$

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

$$e^{\alpha t}\cos\beta t=\frac{1}{2}\left(e^{\lambda t}+e^{\overline{\lambda}t}\right)\quad\text{and}\quad e^{\alpha t}\sin\beta t=\frac{1}{2i}\left(e^{\lambda t}-e^{\overline{\lambda}t}\right). \tag{2.29}$$

Hence, replacing in the sequence $e^{\lambda_1 t},....,e^{\lambda_n t}$ the functions $e^{\lambda t}$ and $e^{\overline{\lambda}t}$ by $e^{\alpha t}\cos\beta t$ and $e^{\alpha t}\sin\beta t$ preserves the linear independence of the sequence.

After replacing all couples $e^{\lambda t}$, $e^{\overline{\lambda}t}$ by the real-valued solutions as above, we obtain n linearly independent real-valued solutions of (2.24), which hence form a basis in the space of all solutions. Taking their linear combinations with real coefficients, we obtain all real-valued solutions. ■

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. 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

$$P(\lambda)=(\lambda-\lambda_0)^m Q(\lambda)\ \text{ for all }\lambda\in\mathbb{C},$$

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 fact that m is maximal possible is equivalent to $Q(\lambda)\neq 0$. The fundamental theorem of algebra can be stated as follows: if $\lambda_1,...,\lambda_r$ are all distinct roots of $P(\lambda)$ and the multiplicity of λ_j is m_j then

$$m_1+...+m_r=n$$

and, hence,

$$P(\lambda)=(\lambda-\lambda_1)^{m_1}...(\lambda-\lambda_r)^{m_r}\ \text{ for all }\lambda\in\mathbb{C}.$$

In order to obtain n independent solutions to the ODE (2.24), each root λ_j should give rise to m_j independent solutions. This is indeed can be done as is stated in the following theorem.

Theorem 2.5 *Let $\lambda_1,...,\lambda_r$ be all the distinct complex roots of the characteristic polynomial $P(\lambda)$ with the multiplicities $m_1,...,m_r$, respectively. Then the following n functions are linearly independent solutions of* (2.24)*:*

$$\left\{t^k e^{\lambda_j t}\right\},\ j=1,...,r,\ k=0,...,m_j-1. \tag{2.30}$$

Consequently, the general solution of (2.24) *is*

$$x(t)=\sum_{j=1}^{r}\sum_{k=0}^{m_j-1}C_{kj}t^k e^{\lambda_j t}\ , \tag{2.31}$$

where C_{kj} are arbitrary complex constants.

If $a_1,...,a_n$ are real and $\lambda=\alpha+i\beta$ is a non-real root of $P(\lambda)$ of multiplicity m, then $\overline{\lambda}=\alpha-i\beta$ is also a root of the same multiplicity m, and the functions $t^k e^{\lambda t}$, $t^k e^{\overline{\lambda}t}$ in the sequence (2.30) *can be replaced by the real-valued functions $t^k e^{\alpha t}\cos\beta t$, $t^k e^{\alpha t}\sin\beta t$, for any $k=0,...,m-1$.*

Remark. Setting

$$P_j(t) = \sum_{k=1}^{m_j-1} C_{jk} t^k ,$$

we obtain from (2.31)

$$x(t) = \sum_{j=1}^{r} P_j(t) e^{\lambda_j t}. \tag{2.32}$$

Hence, any solution to (2.24) has the form (2.32) where P_j is an arbitrary polynomial of t of the degree at most $m_j - 1$.

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 2.5, the functions e^t and te^t are linearly independent solutions, and the general solution is

$$x(t) = (C_1 + C_2 t) e^t.$$

Example. 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

$$x(t) = (C_1 + C_2 t) e^t + \left(C_3 + C_4 t + C_5 t^2\right) e^{-t}.$$

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 following 5 function are linearly independent solutions:

$$1,\ e^{it},\ te^{it},\ e^{-it},\ te^{-it}. \tag{2.33}$$

The general complex solution is then

$$C_1 + (C_2 + C_3 t) e^{it} + (C_4 + C_5 t) e^{-it}.$$

Replacing in the sequence (2.33) e^{it}, e^{-it} by $\cos t, \sin t$ and te^{it}, te^{-it} by $t\cos t, t\sin t$, we obtain the linearly independent real solutions

$$1,\ \cos t,\ t\cos t,\ \sin t,\ t\sin t,$$

and the general real solution

$$C_1 + (C_2 + C_3 t)\cos t + (C_4 + C_5 t)\sin t.$$

We make some preparation for the proof of Theorem 2.5. Given a polynomial

$$P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + ... + a_n$$

with complex coefficients, associate with it the differential operator

$$\begin{aligned} P\left(\frac{d}{dt}\right) &= a_0\left(\frac{d}{dt}\right)^n + a_1\left(\frac{d}{dt}\right)^{n-1} + ... + a_0 \\ &= a_0\frac{d^n}{dt^n} + a_1\frac{d^{n-1}}{dt^{n-1}} + ... + a_0, \end{aligned}$$

where we use the convention that the "product" of differential operators is the composition. That is, the operator $P\left(\frac{d}{dt}\right)$ acts on a smooth enough function $f(t)$ by the rule

$$P\left(\frac{d}{dt}\right)f = a_0 f^{(n)} + a_1 f^{(n-1)} + ... + a_0 f$$

(here the constant term a_0 is understood as a multiplication operator).

For example, the ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = 0 \tag{2.34}$$

can be written shortly in the form

$$P\left(\frac{d}{dt}\right)x = 0$$

where $P(\lambda) = \lambda^n + a_1\lambda^{n-1} + ... + a_n$ is the characteristic polynomial of (2.34).

As an example of usage of this notation, let us prove the following identity:

$$P\left(\frac{d}{dt}\right)e^{\lambda t} = P(\lambda)e^{\lambda t}. \tag{2.35}$$

Indeed, it suffices to verify it for $P(\lambda) = \lambda^k$ and then use the linearity of this identity. For such $P(\lambda) = \lambda^k$, we have

$$P\left(\frac{d}{dt}\right)e^{\lambda t} = \frac{d^k}{dt^k}e^{\lambda t} = \lambda^k e^{\lambda t} = P(\lambda)e^{\lambda t},$$

which was to be proved. If λ is a root of P then (2.35) implies that $e^{\lambda t}$ is a solution to (2.24), which has been observed and used above.

The following lemma is a far reaching generalization of (2.35).

Lemma 2.6 *If $f(t), g(t)$ are n times differentiable functions on an interval then, for any polynomial $P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \ldots + a_n$ of the order at most n, the following identity holds:*

$$P\left(\frac{d}{dt}\right)(fg) = \sum_{j=0}^{n} \frac{1}{j!} f^{(j)} P^{(j)}\left(\frac{d}{dt}\right) g. \tag{2.36}$$

Example. Let $P(\lambda) = \lambda^2 + \lambda + 1$. Then $P'(\lambda) = 2\lambda + 1$, $P'' = 2$, and (2.36) becomes

$$\begin{aligned}(fg)'' + (fg)' + fg &= fP\left(\frac{d}{dt}\right)g + f'P'\left(\frac{d}{dt}\right)g + \frac{1}{2}f''P''\left(\frac{d}{dt}\right)g \\ &= f(g'' + g' + g) + f'(2g' + g) + f''g.\end{aligned}$$

It is an easy exercise to see directly that this identity is correct.

Proof. It suffices to prove the identity (2.36) in the case when $P(\lambda) = \lambda^k$, $k \leq n$, because then for a general polynomial (2.36) will follow by taking linear combination of those for λ^k. If $P(\lambda) = \lambda^k$ then, for $j \leq k$

$$P^{(j)} = k(k-1)\ldots(k-j+1)\lambda^{k-j}$$

and $P^{(j)} \equiv 0$ for $j > k$. Hence,

$$\begin{aligned}P^{(j)}\left(\frac{d}{dt}\right) &= k(k-1)\ldots(k-j+1)\left(\frac{d}{dt}\right)^{k-j}, \ j \leq k, \\ P^{(j)}\left(\frac{d}{dt}\right) &= 0, \ \ j > k,\end{aligned}$$

and (2.36) becomes

$$(fg)^{(k)} = \sum_{j=0}^{k} \frac{k(k-1)\ldots(k-j+1)}{j!} f^{(j)} g^{(k-j)} = \sum_{j=0}^{k} \binom{k}{j} f^{(j)} g^{(k-j)}. \tag{2.37}$$

The latter identity is known from Analysis and is called the *Leibniz formula*[39]. ■

Lemma 2.7 *A complex number λ is a root of a polynomial P with the multiplicity m if and only if*

$$P^{(k)}(\lambda) = 0 \text{ for all } k = 0, \ldots, m-1 \text{ and } P^{(m)}(\lambda) \neq 0. \tag{2.38}$$

[39] If $k = 1$ then (2.37) amounts to the familiar product rule

$$(fg)' = f'g + fg'.$$

For arbitrary $k \in \mathbb{N}$, (2.37) is proved by induction in k.

Proof. If P has a root λ with multiplicity m then we have the identity

$$P(z) = (z-\lambda)^m Q(z) \quad \text{for all } z \in \mathbb{C},$$

where Q is a polynomial such that $Q(\lambda) \neq 0$. We use this identity for $z \in \mathbb{R}$. For any natural k, we have by the Leibniz formula

$$P^{(k)}(z) = \sum_{j=0}^{k} \binom{k}{j} \left((z-\lambda)^m\right)^{(j)} Q^{(k-j)}(z).$$

If $k < m$ then also $j < m$ and

$$\left((z-\lambda)^m\right)^{(j)} = \text{const}\,(z-\lambda)^{m-j},$$

which vanishes at $z = \lambda$. Hence, for $k < m$, we have $P^{(k)}(\lambda) = 0$. For $k = m$ we have again that all the derivatives $\left((z-\lambda)^m\right)^{(j)}$ vanish at $z = \lambda$ provided $j < k$, while for $j = k$ we obtain

$$\left((z-\lambda)^m\right)^{(k)} = \left((z-\lambda)^m\right)^{(m)} = m! \neq 0.$$

Hence,

$$P^{(m)}(\lambda) = \left((z-\lambda)^m\right)^{(m)} Q(\lambda) \neq 0.$$

Conversely, if (2.38) holds then by the Taylor formula for a polynomial, we have

$$\begin{aligned} P(z) &= P(\lambda) + \frac{P'(\lambda)}{1!}(z-\lambda) + \dots + \frac{P^{(n)}(\lambda)}{n!}(z-\lambda)^n \\ &= \frac{P^{(m)}(\lambda)}{m!}(z-\lambda)^m + \dots + \frac{P^{(n)}(\lambda)}{n!}(z-\lambda)^n \\ &= (z-\lambda)^m Q(z) \end{aligned}$$

where

$$Q(z) = \frac{P^{(m)}(\lambda)}{m!} + \frac{P^{(m+1)}(\lambda)}{(m+1)!}(z-\lambda) + \dots + \frac{P^{(n)}(\lambda)}{n!}(z-\lambda)^{n-m}.$$

Obviously, $Q(\lambda) = \frac{P^{(m)}(\lambda)}{m!} \neq 0$, which implies that λ is a root of multiplicity m. ■

Lemma 2.8 *If $\lambda_1, ..., \lambda_r$ are distinct complex numbers and if, for some polynomials $P_j(t)$,*

$$\sum_{j=1}^{r} P_j(t)\, e^{\lambda_j t} = 0 \quad \text{for all } t \in \mathbb{R}, \tag{2.39}$$

then $P_j(t) \equiv 0$ for all j.

Proof. Induction in r. If $r = 1$ then there is nothing to prove. Let us prove the inductive step from $r-1$ to r. Dividing (2.39) by $e^{\lambda_r t}$ and setting $\mu_j = \lambda_j - \lambda_r$, we obtain the identity

$$\sum_{j=1}^{r-1} P_j(t)\, e^{\mu_j t} + P_r(t) = 0. \tag{2.40}$$

Choose some integer $k > \deg P_r$, where $\deg P$ as the maximal power of t that enters P with non-zero coefficient. Differentiating the above identity k times, we obtain

$$\sum_{j=1}^{r-1} Q_j(t) e^{\mu_j t} = 0,$$

where we have used the fact that $(P_r)^{(k)} = 0$ and

$$\left(P_j(t) e^{\mu_j t}\right)^{(k)} = Q_j(t) e^{\mu t}$$

for some polynomial Q_j (this for example follows from the Leibniz formula). By the inductive hypothesis, we conclude that all $Q_j \equiv 0$, which implies that

$$\left(P_j e^{\mu_j t}\right)^{(k)} = 0.$$

Hence, the function $P_j e^{\mu_j t}$ must be equal to a polynomial, say $R_j(t)$. We need to show that $P_j \equiv 0$. Assuming the contrary, we obtain the identity

$$e^{\mu_j t} = \frac{R_j(t)}{P_j(t)} \tag{2.41}$$

which holds for all t except for the (finitely many) roots of $P_j(t)$. If μ_j is real and $\mu_j > 0$ then (2.41) is not possible since $e^{\mu_j t}$ goes to ∞ as $t \to +\infty$ faster than the rational function[40] on the right hand side of (2.41). If $\mu_j < 0$ then the same argument applies when $t \to -\infty$. Let now μ_j be complex, say $\mu_j = \alpha + i\beta$ with $\beta \neq 0$. Then we obtain from (2.41) that

$$e^{\alpha t} \cos \beta t = \operatorname{Re} \frac{R_j(t)}{P_j(t)} \quad \text{and} \quad e^{\alpha t} \sin \beta t = \operatorname{Im} \frac{R_j(t)}{P_j(t)}.$$

It is easy to see that the real and imaginary parts of a rational function are again rational functions. Since $\cos \beta t$ and $\sin \beta t$ vanish at infinitely many points and any rational function vanishes at finitely many points unless it is identical constant, we conclude that $R_j \equiv 0$ and hence $P_j \equiv 0$, for any $j = 1, .., r-1$. Substituting into (2.40), we obtain that also $P_r \equiv 0$. ■

Proof of Theorem 2.5. Let $P(\lambda)$ be the characteristic polynomial of (2.34). We first prove that if λ is a root of multiplicity m then the function $t^k e^{\lambda t}$ solves (2.34) for any $k = 0, ..., m-1$. By Lemma 2.6, we have

$$\begin{aligned} P\left(\frac{d}{dt}\right)\left(t^k e^{\lambda t}\right) &= \sum_{j=0}^{n} \frac{1}{j!} \left(t^k\right)^{(j)} P^{(j)}\left(\frac{d}{dt}\right) e^{\lambda t} \\ &= \sum_{j=0}^{n} \frac{1}{j!} \left(t^k\right)^{(j)} P^{(j)}(\lambda) e^{\lambda t}. \end{aligned}$$

If $j > k$ then the $\left(t^k\right)^{(j)} \equiv 0$. If $j \leq k$ then $j < m$ and, hence, $P^{(j)}(\lambda) = 0$ by hypothesis. Hence, all the terms in the above sum vanish, whence

$$P\left(\frac{d}{dt}\right)\left(t^k e^{\lambda t}\right) = 0,$$

[40] A rational function is by definition the ratio of two polynomials.

that is, the function $x(t) = t^k e^{\lambda t}$ solves (2.34).

If $\lambda_1, ..., \lambda_r$ are all distinct complex roots of $P(\lambda)$ and m_j is the multiplicity of λ_j then it follows that each function in the following sequence

$$\left\{t^k e^{\lambda_j t}\right\}, \ j = 1, ..., r, \ k = 0, ..., m_j - 1, \tag{2.42}$$

is a solution of (2.34). Let us show that these functions are linearly independent. Clearly, each linear combination of functions (2.42) has the form

$$\sum_{j=1}^{r} \sum_{k=0}^{m_j - 1} C_{jk} t^k e^{\lambda_j t} = \sum_{j=1}^{r} P_j(t) e^{\lambda_j t} \tag{2.43}$$

where $P_j(t) = \sum_{k=0}^{m_j-1} C_{jk} t^k$ are polynomials. If the linear combination is identical zero then by Lemma 2.8 $P_j \equiv 0$, which implies that all C_{jk} are 0. Hence, the functions (2.42) are linearly independent, and by Theorem 2.3 the general solution of (2.34) has the form (2.43).

Let us show that if $\lambda = \alpha + i\beta$ is a complex (non-real) root of multiplicity m then $\overline{\lambda} = \alpha - i\beta$ is also a root of the same multiplicity m. Indeed, by Lemma 2.7, λ satisfies the relations (2.38). Applying the complex conjugation and using the fact that the coefficients of P are real, we obtain that the same relations hold for $\overline{\lambda}$ instead of λ, which implies that $\overline{\lambda}$ is also a root of multiplicity m.

The last claim that every couple $t^k e^{\lambda t}, t^k e^{\overline{\lambda} t}$ in (2.42) can be replaced by real-valued functions $t^k e^{\alpha t} \cos \beta t$, $t^k e^{\alpha t} \sin \beta t$, follows from the observation that the functions $t^k e^{\alpha t} \cos \beta t$, $t^k e^{\alpha t} \sin \beta t$ are linear combinations of $t^k e^{\lambda t}, t^k e^{\overline{\lambda} t}$, and vice versa, which one sees from the identities

$$e^{\alpha t} \cos \beta t = \frac{1}{2}\left(e^{\lambda t} + e^{\overline{\lambda} t}\right), \quad e^{\alpha t} \sin \beta t = \frac{1}{2i}\left(e^{\lambda t} - e^{\overline{\lambda} t}\right),$$

$$e^{\lambda t} = e^{\alpha t}(\cos \beta t + i \sin \beta t), \qquad e^{\overline{\lambda} t} = e^{\alpha t}(\cos \beta t - i \sin \beta t),$$

multiplied by t^k (compare the proof of Theorem 2.4). ■

2.5 Solving linear inhomogeneous ODEs with constant coefficients

Here we consider the ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = f(t), \tag{2.44}$$

where the function $f(t)$ is a *quasi-polynomial*, that is, f has the form

$$f(t) = \sum_j R_j(t) e^{\mu_j t}$$

where $R_j(t)$ are polynomials, μ_j 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 before, denote by $P(\lambda)$ the characteristic polynomial of (2.44), that is

$$P(\lambda) = \lambda^n + a_1\lambda^{n-1} + ... + a_n.$$

Then the equation (2.44) can be written shortly in the form $P\left(\frac{d}{dt}\right) x = f$, which will be used below. We start with the following observation.

Claim. *If $f = f_1 + ... + f_k$ and $x_1(t), ..., x_k(t)$ are solutions to the equation $P\left(\frac{d}{dt}\right) x_j = f_j$, then $x = x_1 + ... + x_k$ solves the equation $P\left(\frac{d}{dt}\right) x = f$.*

Proof. This is trivial because

$$P\left(\frac{d}{dt}\right) x = P\left(\frac{d}{dt}\right) \sum_j x_j = \sum_j P\left(\frac{d}{dt}\right) x_j = \sum_j f_j = f.$$

■

Hence, we can assume that the function f in (2.44) is of the form $f(t) = R(t)e^{\mu t}$ where $R(t)$ is a polynomial. As we know, the general solution of the inhomogeneous equation (2.44) is obtained as a sum of the general solution of the associated homogeneous equation and a particular solution of (2.44). Hence, we focus on finding a particular solution of (2.44).

To illustrate the method, which will be used in this Section, consider first the following particular case:

$$P\left(\frac{d}{dt}\right)x = e^{\mu t} \tag{2.45}$$

where μ is *not* a root of the characteristic polynomial $P(\lambda)$ (*non-resonant case*). We claim that (2.45) has a particular solution in the form $x(t) = ae^{\mu t}$ where a is a complex constant to be chosen. Indeed, we have by (2.35)

$$P\left(\frac{d}{dt}\right)\left(e^{\mu t}\right) = P(\mu)e^{\mu t},$$

whence

$$P\left(\frac{d}{dt}\right)\left(ae^{\mu t}\right) = e^{\mu t}$$

provided

$$a = \frac{1}{P(\mu)}. \tag{2.46}$$

Example. Let us find a particular solution to the ODE

$$x'' + 2x' + x = e^t.$$

Note that $P(\lambda) = \lambda^2 + 2\lambda + 1$ and $\mu = 1$ is not a root of P. Look for a 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 (2.46), that is,

$$a = \frac{1}{P(\mu)} = \frac{1}{1+2+1} = \frac{1}{4}.$$

Hence, the answer is $x(t) = \frac{1}{4}e^t$.

Consider another equation:

$$x'' + 2x' + x = \sin t \tag{2.47}$$

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 a 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 (2.47).

Example. Consider yet another ODE

$$x'' + 2x' + x = e^{-t}\cos t. \tag{2.48}$$

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 (2.48) is $x(t) = -e^{-t}\cos t$.

Example. Finally, let us combine the above examples into one:

$$x'' + 2x' + x = 2e^t - \sin t + e^{-t}\cos t. \tag{2.49}$$

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 the homogeneous ODE $x'' + 2x' + x = 0$ is

$$x(t) = (C_1 + C_2 t)e^{-t},$$

we obtain the general solution to (2.49)

$$x(t) = (C_1 + C_2 t)e^{-t} + \frac{1}{2}e^t + \frac{1}{2}\cos t - e^{-t}\cos t.$$

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 2.9 *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*

$$P\left(\frac{d}{dt}\right)x = R(t)\,e^{\mu t} \tag{2.50}$$

has a particular 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).

In the case when $k = 0$ and $R(t) \equiv a$, the ODE $P\left(\frac{d}{dt}\right)x = ae^{\mu t}$ has a particular solution

$$x(t) = \frac{a}{P^{(m)}(\mu)}t^m e^{\mu t}. \tag{2.51}$$

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

$$x(t) = at^2 e^{-t}$$

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

$$a\left(\left(t^2e^{-t}\right)'' + 2\left(t^2e^{-t}\right)' + t^2e^{-t}\right) = e^{-t} \tag{2.52}$$

Expanding the expression in the brackets, we obtain the identity

$$\left(t^2e^{-t}\right)'' + 2\left(t^2e^{-t}\right)' + t^2e^{-t} = 2e^{-t},$$

so that (2.52) becomes $2a = 1$ and $a = \frac{1}{2}$. Hence, a particular solution is

$$x(t) = \frac{1}{2}t^2e^{-t}.$$

Alternatively, this solution can be obtained by (2.51):

$$x(t) = \frac{1}{P''(-1)}t^2e^{-t} = \frac{1}{2}t^2e^{-t}.$$

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

$$2b + 6at = t$$

so that $b = 0$ and $a = \frac{1}{6}$. The final answer is

$$x\left(t\right) = \frac{t^3}{6}e^{-t}.$$

Proof of Theorem 2.9. Let us prove that the equation

$$P\left(\frac{d}{dt}\right)x = R\left(t\right)e^{\mu t}$$

has a solution in the form

$$x\left(t\right) = t^m Q\left(t\right)e^{\mu t}$$

where m is the multiplicity of μ and $\deg Q = k = \deg R$. Using Lemma 2.6, we have

$$\begin{aligned} P\left(\frac{d}{dt}\right)x &= P\left(\frac{d}{dt}\right)\left(t^m Q\left(t\right)e^{\mu t}\right) = \sum_{j\geq 0}\frac{1}{j!}\left(t^m Q\left(t\right)\right)^{(j)}P^{(j)}\left(\frac{d}{dt}\right)e^{\mu t} \\ &= \sum_{j\geq 0}\frac{1}{j!}\left(t^m Q\left(t\right)\right)^{(j)}P^{(j)}\left(\mu\right)e^{\mu t}. \end{aligned} \tag{2.53}$$

By Lemma 2.6, the summation here runs from $j = 0$ to $j = n$ but we can allow any $j \geq 0$ because for $j > n$ the derivative $P^{(j)}$ is identical zero anyway. Furthermore, since $P^{(j)}\left(\mu\right) = 0$ for all $j \leq m - 1$, we can restrict the summation to $j \geq m$. Set

$$y\left(t\right) = \left(t^m Q\left(t\right)\right)^{(m)} \tag{2.54}$$

and observe that $y\left(t\right)$ is a polynomial of degree k, provided so is $Q\left(t\right)$. Conversely, for any polynomial $y\left(t\right)$ of degree k, there is a polynomial $Q\left(t\right)$ of degree k such that (2.54) holds. Indeed, integrating (2.54) m times without adding constants and then dividing by t^m, we obtain $Q\left(t\right)$ as a polynomial of degree k.

It follows from (2.53) that y must satisfy the ODE

$$\sum_{j\geq m}\frac{P^{(j)}\left(\mu\right)}{j!}y^{(j-m)} = R\left(t\right),$$

which we rewrite in the form

$$b_0 y + b_1 y' + ... + b_l y^{(l)} + ... = R\left(t\right) \tag{2.55}$$

where $b_l = \frac{P^{(l+m)}(\mu)}{(l+m)!}$ (in fact, the index $l = j - m$ can be restricted to $l \leq k$ since $y^{(l)} \equiv 0$ for $l > k$). Note that

$$b_0 = \frac{P^{(m)}(\mu)}{m!} \neq 0. \tag{2.56}$$

Hence, the problem amounts to the following: given a polynomial $R(t)$ of degree k, prove that there exists a polynomial $y(t)$ of degree k that satisfies (2.55). Let us prove the existence of y by induction in k.

The inductive basis for $k = 0$. In this case, $R(t)$ is a constant, say $R(t) = a$, and $y(t)$ is sought as a constant too, say $y(t) = c$. Then (2.55) becomes $b_0 c = a$ whence $c = a/b_0$. Here we use that $b_0 \neq 0$.

The inductive step from the values smaller than k to k. Represent y in the from

$$y = ct^k + z(t), \tag{2.57}$$

where z is a polynomial of degree $< k$. Substituting (2.57) into (2.55), we obtain the equation for z

$$b_0 z + b_1 z' + \ldots + b_l z^{(l)} + \ldots = R(t) - \left(b_0 c t^k + b_1 \left(ct^k\right)' + \ldots\right) =: \widetilde{R}(t).$$

Setting $R(t) = at^k +$ lower order terms and choosing c from the equation $b_0 c = a$, that is, $c = a/b_0$, we see that the term t^k cancels out and $\widetilde{R}(t)$ is a polynomial of degree $< k$. By the inductive hypothesis, the equation

$$b_0 z + b_1 z' + \ldots + b_l z^{(l)} + \ldots = \widetilde{R}(t)$$

has a solution $z(t)$ that is a polynomial of degree $< k$. Hence, the function $y = ct^k + z$ solves (2.55) and is a polynomial of degree k.

If $k = 0$, that is, $R(t) \equiv a$ is a constant then (2.55) yields

$$y = \frac{a}{b_0} = \frac{m!a}{P^{(m)}(\mu)}.$$

The equation (2.54) becomes

$$\left(t^m Q(t)\right)^{(m)} = \frac{m!a}{P^{(m)}(\mu)}$$

whence after m integrations we find one of solutions for $Q(t)$:

$$Q(t) = \frac{a}{P^{(m)}(\mu)}.$$

Therefore, the ODE $P\left(\frac{d}{dt}\right)x = ae^{\mu t}$ has a particular solution

$$x(t) = \frac{a}{P^{(m)}(\mu)} t^m e^{\mu t}.$$

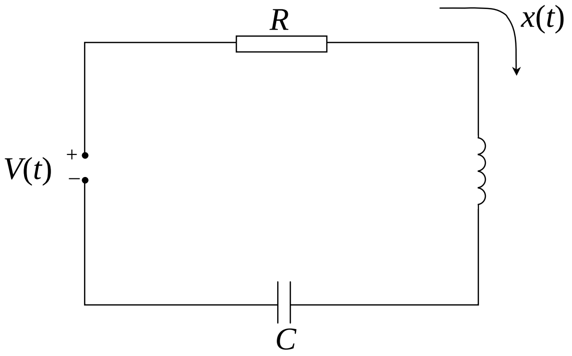
■

2.6 Second order ODE with periodic right hand side

Consider a second order ODE

$$x'' + px' + qx = f(t), \tag{2.58}$$

which occurs in various physical phenomena. For example, (2.58) describes the movement of a point body of mass m along the axis x, where the term px' comes from the friction forces, qx - from the elastic forces, and $f(t)$ is an external time-dependant force. Another physical situation that is described by (2.58), is 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. We have seen that $x(t)$ satisfied the following ODE

$$Lx'' + Rx' + \frac{x}{C} = V'.$$

If $L \neq 0$ then dividing by L we obtain an ODE of the form (2.58) where $p = R/L$, $q = 1/(LC)$, and $f = V'/L$.

As an example of application of the above methods of solving linear ODEs with constant coefficients, we investigate here the ODE

$$x'' + px' + qx = A \sin \omega t, \tag{2.59}$$

where A, ω are given positive reals. The function $A \sin \omega t$ is a model for a more general periodic force $f(t)$, 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 power source is an electrical socket with the alternating current[41] (AC). The number ω is called the *frequency*[42] of the external force (note that the period $= \frac{2\pi}{\omega}$) or the external frequency, and the number A is called the *amplitude*[43] (the maximum value) of the external force.

Assume in the sequel that $p \geq 0$ and $q > 0$, which is physically most interesting case. To find a particular solution of (2.59), let us consider first the ODE with complex right hand side:

$$x'' + px' + qx = Ae^{i\omega t}. \tag{2.60}$$

[41] Wechselstrom

[42] Frequenz

[43] Amplitude

Let $P(\lambda) = \lambda^2 + p\lambda + q$ be the characteristic polynomial. Consider first the non-resonant case when $i\omega$ is not a root of $P(\lambda)$. Searching the solution in the from $ce^{i\omega t}$, we obtain by Theorem 2.9

$$c = \frac{A}{P(i\omega)} = \frac{A}{-\omega^2 + pi\omega + q} =: a + ib$$

and the particular solution of (2.60) 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 of (2.59)

$$x(t) = a \sin \omega t + b \cos \omega t. \tag{2.61}$$

Rewrite (2.61) in the form

$$x(t) = B(\cos\varphi \sin\omega t + \sin\varphi\cos\omega t) = B\sin(\omega t + \varphi),$$

where

$$B = \sqrt{a^2+b^2} = |c| = \frac{A}{\sqrt{(q-\omega^2)^2 + \omega^2 p^2}}, \tag{2.62}$$

and the angle $\varphi \in [0, 2\pi)$ is determined from the identities

$$\cos\varphi = \frac{a}{B}, \quad \sin\varphi = \frac{b}{B}.$$

The number B is the amplitude of the solution, and φ is called the *phase*.

To obtain the general solution of (2.59), we need to add to (2.61) the general solution of 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 the following cases.

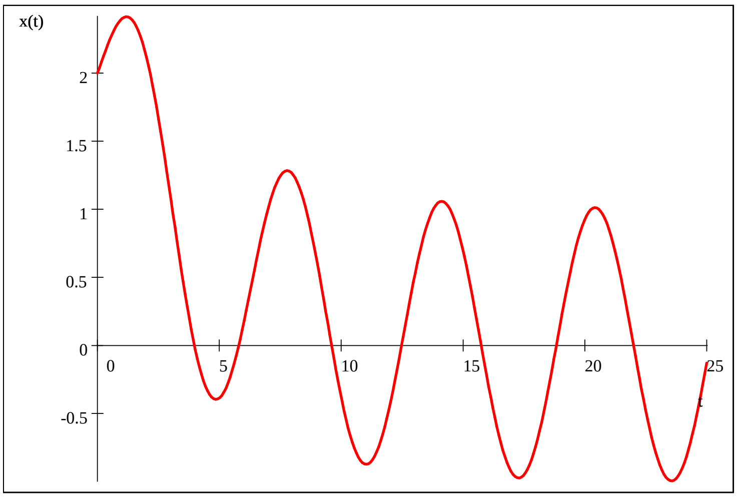
λ_1 and λ_2 are real. Since $p \geq 0$ and $q > 0$, we see that $\lambda_1, \lambda_2 < 0$. The general solution of the homogeneous equation has the from

$$\begin{array}{lccc} C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} & \text{if } \lambda_1 & \neq & \lambda_2, \\ (C_1 + C_2 t)\, e^{\lambda_1 t} & \text{if } \lambda_1 & = & \lambda_2. \end{array}$$

In the both cases, it decays exponentially in t as $t \to +\infty$. Hence, the general solution of (2.59) has the form

$$x(t) = B\sin(\omega t + \varphi) + \text{exponentially decaying terms.}$$

As we see, when $t \to \infty$ the leading term of $x(t)$ is the above particular solution $B\sin(\omega t + \varphi)$. For the electrical circuit this means that the current quickly stabilizes and becomes periodic with the same frequency ω as the external force. Here is a plot of such a function $x(t) = \sin t + 2e^{-t/4}$:



λ_1 and λ_2 are complex.
Let $\lambda_{1,2} = \alpha \pm i\beta$ 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

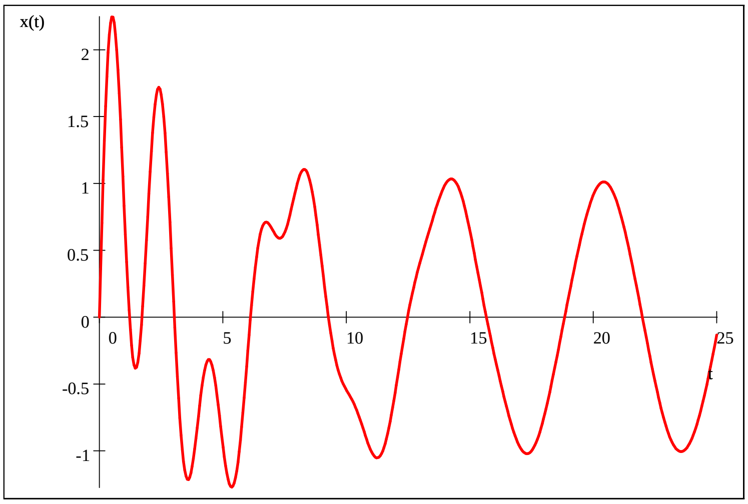
$$e^{\alpha t} \left(C_1 \cos \beta t + C_2 \sin \beta t\right) = Ce^{\alpha t} \sin \left(\beta t + \psi\right),$$

where C and ψ are arbitrary reals. 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 oscillates with the natural frequency β.

Hence, the general solution to (2.59) is

$$x\left(t\right) = B \sin \left(\omega t + \varphi\right) + Ce^{\alpha t} \sin \left(\beta t + \psi\right).$$

Consider two further sub-cases. If $\alpha < 0$ then the leading term is again $B \sin \left(\omega t + \varphi\right)$. Here is a plot of a particular example of this type: $x\left(t\right) = \sin t + 2e^{-t/4} \sin \pi t$.



If $\alpha = 0$, then $p = 0$, $q = \beta^2$, and the equation has the form

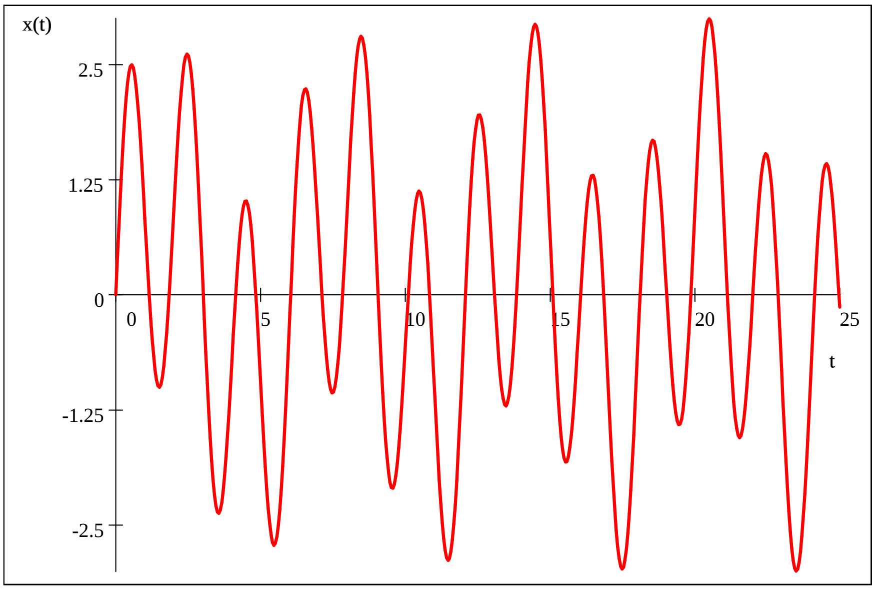
$$x'' + \beta^2 x = A \sin \omega t.$$

The roots of the characteristic polynomial are $\pm i\beta$. The assumption that $i\omega$ is not a root means that $\omega \neq \beta$. 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. If ω and β are incommensurable[44] then $x\left(t\right)$ is not periodic unless $C = 0$. Here is a particular example of such a function: $x\left(t\right) = \sin t + 2 \sin \pi t$.

[44] inkommensurabel



Strictly speaking, in practice the electrical circuits with $p = 0$ do not occur since the resistance is always positive.

Let us come back to the formula (2.62) for the amplitude B and, as an example of its application, consider the following question: for what value of the external frequency ω the amplitude B is maximal? 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 of 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 the resonant frequency, that is, if $\omega = \omega_0$, 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 larger than ω_0. In terms of ω_0 and β, we can write

$$\begin{aligned} B^2 &= \frac{A^2}{\omega^4 - 2\omega_0^2\omega^2 + q^2} = \frac{A^2}{\left(\omega^2 - \omega_0^2\right)^2 + q^2 - \omega_0^4} \\ &= \frac{A^2}{\left(\omega^2 - \omega_0^2\right)^2 + p^2\beta^2}, \end{aligned}$$

where we have used that

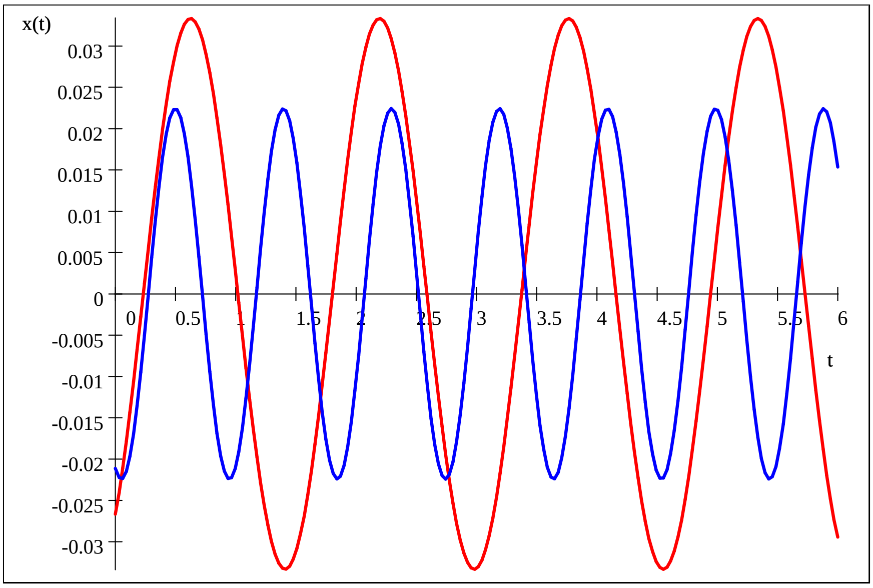
$$q^2 - \omega_0^4 = q^2 - \left(q - \frac{p^2}{2}\right)^2 = qp^2 - \frac{p^4}{4} = p^2\beta^2.$$

In particular, the maximum amplitude, that occurs at $\omega = \omega_0$, is $B_{\max} = \frac{A}{p\beta}$.

Consider a numerical example of this type. For the ODE

$$x'' + 6x' + 34x = \sin \omega t,$$

the natural frequency is $\beta = \sqrt{q - p^2/4} = 5$ and the resonant frequency is $\omega_0 = \sqrt{q - p^2/2} = 4$. The particular solution (2.61) in the case $\omega = \omega_0 = 4$ is $x(t) = \frac{1}{50}\sin 4t - \frac{2}{75}\cos 4t$ with amplitude $B = B_{\max} = 1/30$, and in the case $\omega = 7$ it is $x(t) = -\frac{5}{663}\sin 7t - \frac{14}{663}\cos 7t$ with the amplitude $B \approx 0.224$. The plots of these two functions are shown below:



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 (2.59) has the form

$$x'' + \omega^2 x = A \sin \omega t. \tag{2.63}$$

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 by Theorem 2.9

$$c = \frac{A}{P'(i\omega)} = \frac{A}{2i\omega}.$$

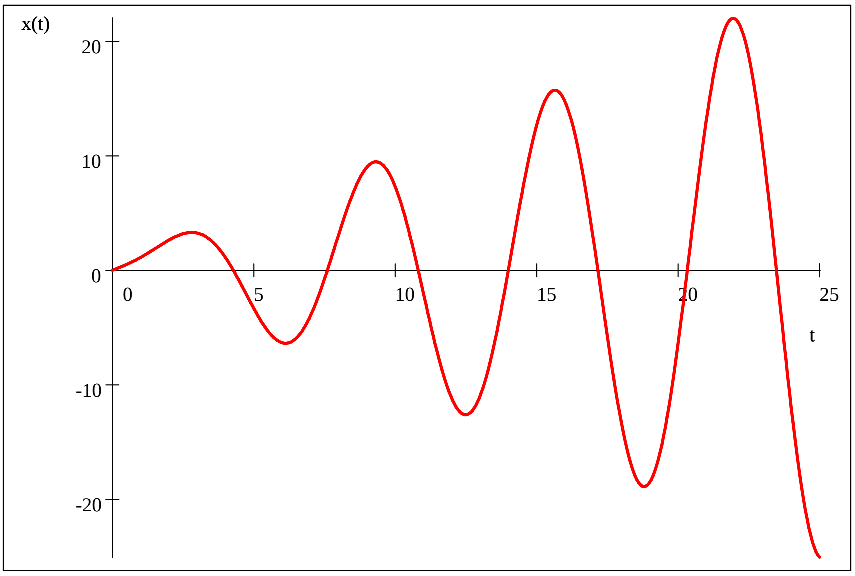
Hence, the complex particular solution is

$$x(t) = \frac{At}{2i\omega}e^{i\omega t} = -i\frac{At}{2\omega}\cos \omega t + \frac{At}{2\omega}\sin \omega t.$$

Using its imaginary part, we obtain the general solution of (2.63) as follows:

$$x(t) = -\frac{At}{2\omega}\cos \omega t + C \sin(\omega t + \psi).$$

Here is a plot of such a function $x(t) = -t \cos t + 2 \sin t$:



In this case 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 falls apart over time or the mathematical model becomes unsuitable for describing the physical system.

2.7 The method of variation of parameters

2.7.1 A system of the 1st order

Consider again a general linear system

$$x' = A(t)x + B(t) \tag{2.64}$$

where as before $A(t) : I \to \mathbb{R}^{n \times n}$ and $B(t) : I \to \mathbb{R}^n$ are continuous, I being an interval in $\mathbb{R}$. We present here the method of *variation of parameters* that allows to solve (2.64), given n linearly independent solutions $x_1(t)$,..., $x_n(t)$ of the homogeneous system $x' = A(t)x$.

We start with the following observation.

Lemma 2.10 *If the solutions $x_1(t), ..., x_n(t)$ of the system $x' = A(t)x$ are linearly independent then, for any $t_0 \in I$, the vectors $x_1(t_0), ..., x_n(t_0)$ are linearly independent.*

Proof. This statement is contained implicitly in the proof of Theorem 2.3, but nevertheless we repeat the argument. Assume that for some constant $C_1, ..., C_n$

$$C_1 x_1(t_0) + ... + C_n x_n(t_0) = 0.$$

Consider the function $x(t) = C_1 x_1(t) + ... + C_n x_n(t)$. Then $x(t)$ solves the IVP

$$\begin{cases} x' = A(t)x, \\ x(t_0) = 0, \end{cases}$$

whence by the uniqueness theorem $x(t) \equiv 0$. Since the solutions $x_1, ..., x_n$ are independent, it follows that $C_1 = ... = C_n = 0$, whence the independence of vectors $x_1(t_0), ..., x_n(t_0)$ follows. ■

Example. Consider two vector functions

$$x_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} \sin t \\ \cos t \end{pmatrix},$$

which are obviously linearly independent. However, for $t = \pi/4$, we have

$$x_1(t) = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} = x_2(t)$$

so that the vectors $x_1(\pi/4)$ and $x_2(\pi/4)$ are linearly dependent. Hence, $x_1(t)$ and $x_2(t)$ cannot be solutions of the same system $x' = Ax$.

For comparison, the functions

$$x_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \text{ and } x_2(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}$$

are solutions of the same system

$$x' = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x,$$

and, hence, the vectors $x_1(t)$ and $x_2(t)$ are linearly independent for any t. This follows also from

$$\det(x_1 \mid x_2) = \det \begin{pmatrix} \cos t & -\sin t \\ \sin & \cos t \end{pmatrix} = 1 \neq 0.$$

Given n linearly independent solutions to $x' = A(t)x$, let us form a $n \times n$ matrix

$$X(t) = (x_1(t) \mid x_2(t) \mid ... \mid x_n(t))$$

where the k-th column is the column-vector $x_k(t)$, $k = 1, ..., n$. The matrix X is called the *fundamental matrix* of the system $x' = Ax$.

It follows from Lemma 2.10 that the column of $X(t)$ are linearly independent for any $t \in I$, which implies that $\det X(t) \neq 0$ and that the inverse matrix $X^{-1}(t)$ is defined for all $t \in I$. This allows us to solve the inhomogeneous system as follows.

Theorem 2.11 *The general solution to the system*

$$x' = A(t)x + B(t), \tag{2.65}$$

is given by

$$x(t) = X(t) \int X^{-1}(t) B(t)\, dt, \tag{2.66}$$

where X is the fundamental matrix of the system $x' = Ax$.

Proof. Let us look for a solution to (2.65) in the form

$$x(t) = C_1(t)\,x_1(t) + ... + C_n(t)\,x_n(t) \tag{2.67}$$

where $C_1, C_2, .., C_n$ are now unknown real-valued functions to be determined. Since for any $t \in I$, the vectors $x_1(t), ..., x_n(t)$ are linearly independent, any $\mathbb{R}^n$-valued function $x(t)$ can be represented in the form (2.67). Let $C(t)$ be the column-vector with components $C_1(t), ..., C_n(t)$. Then identity (2.67) can be written in the matrix form as

$$x = XC$$

whence $C = X^{-1}x$. It follows that $C_1, ..., C_n$ are rational functions of $x_1, ..., x_n, x$. Since $x_1, ..., x_n, x$ are differentiable, the functions $C_1, ..., C_n$ are also differentiable.

Differentiating the identity (2.67) in t and using $x_k' = Ax_k$, we obtain

$$\begin{aligned} x' &= C_1x_1' + C_2x_2' + ... + C_nx_n' \\ &\quad + C_1'x_1 + C_2'x_2 + ... + C_n'x_n \\ &= C_1Ax_1 + C_2Ax_2 + ... + C_nAx_n \\ &\quad + C_1'x_1 + C_2'x_2 + ... + C_n'x_n \\ &= Ax + C_1'x_1 + C_2'x_2 + ... + C_n'x_n \\ &= Ax + XC'. \end{aligned}$$

Hence, the equation $x' = Ax + B$ is equivalent to

$$XC' = B. \tag{2.68}$$

Solving this vector equation for C', we obtain

$$C' = X^{-1}B,$$

$$C(t) = \int X^{-1}(t)\,B(t)\,dt,$$

and

$$x(t) = XC = X(t)\int X^{-1}(t)\,B(t)\,dt.$$

■

The term "variation of parameters" comes from the identity (2.67). Indeed, if $C_1, ..., C_n$ are constant parameters then this identity determines the general solution of the homogeneous ODE $x' = Ax$. By allowing $C_1, ..., C_n$ to be variable, we obtain the general solution to $x' = Ax + B$.

Second proof. Observe first that the matrix X satisfies the following ODE

$$X' = AX.$$

Indeed, this identity holds for any column x_k of X, whence it follows for the whole matrix. Differentiating (2.66) in t and using the product rule, we obtain

$$\begin{aligned} x' &= X'(t)\int X^{-1}(t)\,B(t)\,dt + X(t)\left(X^{-1}(t)\,B(t)\right) \\ &= AX\int X^{-1}B(t)\,dt + B(t) \\ &= Ax + B(t). \end{aligned}$$

Hence, $x\left(t\right)$ solves (2.65). Let us show that (2.66) gives all the solutions. Note that the integral in (2.66) is indefinite so that it can be presented in the form

$$\int X^{-1}\left(t\right)B\left(t\right)dt=V\left(t\right)+C,$$

where $V\left(t\right)$ is a vector function and $C=\left(C_1,...,C_n\right)$ is an arbitrary constant vector. Hence, (2.66) gives

$$\begin{aligned} x\left(t\right) &= X\left(t\right)V\left(t\right)+X\left(t\right)C \\ &= x_0\left(t\right)+C_1x_1\left(t\right)+...+C_nx_n\left(t\right), \end{aligned}$$

where $x_0\left(t\right)=X\left(t\right)V\left(t\right)$ is a solution of (2.65). By Theorem 2.3 we conclude that $x\left(t\right)$ is indeed the general solution.

Example. Consider the system

$$\begin{cases} x_1'=-x_2 \\ x_2'=x_1 \end{cases}$$

or, in the vector form,

$$x'=\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}x.$$

It is easy to see that this system has two independent solutions

$$x_1\left(t\right)=\begin{pmatrix}\cos t \\ \sin t\end{pmatrix} \quad \text{and} \quad x_2\left(t\right)=\begin{pmatrix}-\sin t \\ \cos t\end{pmatrix}.$$

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

$$x'=A\left(t\right)x+B\left(t\right)$$

where $B\left(t\right)=\binom{b_1(t)}{b_2(t)}$. By (2.66), we obtain the general solution

$$\begin{aligned} x &= \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}\int\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}\begin{pmatrix} b_1\left(t\right) \\ b_2\left(t\right)\end{pmatrix}dt \\ &= \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}\int\begin{pmatrix} b_1\left(t\right)\cos t+b_2\left(t\right)\sin t \\ -b_1\left(t\right)\sin t+b_2\left(t\right)\cos t\end{pmatrix}dt. \end{aligned}$$

Consider a particular example $B\left(t\right)=\binom{1}{-i}$. Then the integral is

$$\int\begin{pmatrix}\cos t-t\sin t \\ -\sin t-t\cos t\end{pmatrix}dt=\begin{pmatrix} t\cos t+C_1 \\ -t\sin t+C_2\end{pmatrix},$$

whence

$$\begin{aligned} x &= \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}\begin{pmatrix} t\cos t+C_1 \\ -t\sin t+C_2\end{pmatrix} \\ &= \begin{pmatrix} C_1\cos t-C_2\sin t+t \\ C_1\sin t+C_2\cos t\end{pmatrix} \\ &= \begin{pmatrix} t \\ 0\end{pmatrix}+C_1\begin{pmatrix}\cos t \\ \sin t\end{pmatrix}+C_2\begin{pmatrix}-\sin t \\ \cos t\end{pmatrix}. \end{aligned}$$

2.7.2 A scalar ODE of n-th order

Consider now a scalar ODE of order n

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = f(t), \tag{2.69}$$

where $a_k(t)$ and $f(t)$ 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

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 \\ 0 \\ ... \\ f \end{pmatrix}.$$

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 solutions, we obtain the fundamental matrix

$$X = (\ \mathbf{x}_1 \mid \mathbf{x}_2 \mid \ \dots \ \mid \mathbf{x}_n\ ) = \begin{pmatrix} x_1 & x_2 & ... & x_n \\ x_1' & x_2' & ... & x_n' \\ ... & ... & ... & ... \\ x_1^{(n-1)} & x_2^{(n-1)} & ... & x_n^{(n-1)} \end{pmatrix}.$$

We need to multiply X^{-1} by B. 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

$$X^{-1}B = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ ... & ... & ... \\ y_{n1} & \cdots & y_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ ... \\ f \end{pmatrix} = \begin{pmatrix} y_{1n} f \\ ... \\ y_{nn} f \end{pmatrix} = f y_n,$$

and the general vector solution is

$$\mathbf{x} = X(t) \int f(t)\, y_n(t)\, dt.$$

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 $\int f(t)\, y_n(t)\, dt$, whence

$$x(t) = \sum_{j=1}^{n} x_j(t) \int f(t)\, y_{jn}(t)\, dt.$$

Hence, we have proved the following.

Corollary. *Let $x_1, ..., x_n$ be n linearly independent solutions of*

$$x^{(n)} + a_1(t)\, x^{(n-1)} + ... + a_n(t)\, x = 0$$

and X be the corresponding fundamental matrix. Then, for any continuous function $f(t)$, the general solution of (2.69) is given by

$$x(t) = \sum_{j=1}^{n} x_j(t) \int f(t)\, y_{jn}(t)\, dt, \tag{2.70}$$

where y_{jk} are the entries of the matrix X^{-1}.

Example. Consider the ODE

$$x'' + x = f(t). \tag{2.71}$$

The independent solutions of the homogeneous equation $x'' + x = 0$ are $x_1(t) = \cos t$ and $x_2(t) = \sin t$, so that the fundamental matrix is

$$X = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix},$$

and the inverse is

$$X^{-1} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

By (2.70), the general solution to (2.71) is

$$\begin{aligned} x(t) &= x_1(t) \int f(t)\, y_{12}(t)\, dt + x_2(t) \int f(t)\, y_{22}(t)\, dt \\ &= \cos t \int f(t)\,(-\sin t)\, dt + \sin t \int f(t) \cos t dt. \end{aligned} \tag{2.72}$$

For example, if $f(t) = \sin t$ then we obtain

$$\begin{aligned} x(t) &= \cos t \int \sin t\,(-\sin t)\, dt + \sin t \int \sin t \cos t dt \\ &= -\cos t \int \sin^2 t dt + \frac{1}{2} \sin t \int \sin 2t dt \\ &= -\cos t \left(\frac{1}{2} t - \frac{1}{4} \sin 2t + C_1 \right) + \frac{1}{4} \sin t\,(-\cos 2t + C_2) \\ &= -\frac{1}{2} t \cos t + \frac{1}{4} (\sin 2t \cos t - \sin t \cos 2t) + c_1 \cos t + c_2 \sin t \\ &= -\frac{1}{2} t \cos t + c_1 \cos t + c_2 \sin t. \end{aligned}$$

Of course, the same result can be obtained by Theorem 2.9. Consider one more example

$f(t) = \tan t$ that cannot be handled by Theorem 2.9. In this case, we obtain from (2.72)[45]

$$\begin{aligned} x &= \cos t \int \tan t\,(-\sin t)\,dt + \sin t \int \tan t \cos t dt \\ &= \cos t \left(\frac{1}{2} \ln \left(\frac{1-\sin t}{1+\sin t} \right) + \sin t \right) - \sin t \cos t + c_1 \cos t + c_2 \sin t \\ &= \frac{1}{2} \cos t \ln \left(\frac{1-\sin t}{1+\sin t} \right) + c_1 \cos t + c_2 \sin t. \end{aligned}$$

Let us show how one can use the method of variation of parameters for the equation (2.71) directly, without using the formula (2.70). We start with writing up the general solution to the homogeneous ODE $x'' + x = 0$, which is

$$x(t) = C_1 \cos + C_2 \sin t, \tag{2.73}$$

where C_1 and C_2 are constant. Next, we look for the solution of (2.71) in the form

$$x(t) = C_1(t) \cos t + C_2(t) \sin t, \tag{2.74}$$

which is obtained from (2.73) by replacing the constants by functions. To obtain the equations for the unknown functions $C_1(t), C_2(t)$, differentiate (2.74):

$$\begin{aligned} x'(t) &= -C_1(t) \sin t + C_2(t) \cos t \\ &\quad + C_1'(t) \cos t + C_2'(t) \sin t. \end{aligned} \tag{2.75}$$

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,

$$C_1' \cos t + C_2' \sin t = 0. \tag{2.76}$$

The motivation for this choice is as follows. Switching to the normal system, one must have the identity

$$\mathbf{x}(t) = C_1(t)\, \mathbf{x}_1(t) + C_2 \mathbf{x}_2(t),$$

which componentwise is

$$\begin{aligned} x(t) &= C_1(t) \cos t + C_2(t) \sin t \\ x'(t) &= C_1(t) (\cos t)' + C_2(t) (\sin t)'. \end{aligned}$$

Differentiating the first line and subtracting the second line, we obtain (2.76).

[45] 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

$$\int \frac{dt}{\cos t} = \int \frac{d \sin t}{\cos^2 t} = \int \frac{d \sin t}{1 - \sin^2 t} = \frac{1}{2} \ln \frac{1-\sin t}{1+\sin t}.$$

It follows from (2.75) and (2.76) 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 done correctly). Hence, the second equation for C_1' and C_2' is

$$-C_1' \sin t + C_2' \cos t = f(t).$$

Solving the system of linear algebraic equations

$$\begin{cases} C_1' \cos t + C_2' \sin t = 0 \\ -C_1' \sin t + C_2' \cos t = f(t) \end{cases}$$

we obtain

$$C_1' = -f(t) \sin t, \qquad C_2' = f(t) \cos t$$

whence

$$C_1 = -\int f(t) \sin t dt, \quad C_2 = \int f(t) \cos t dt.$$

Substituting into (2.74), we obtain (2.72).

2.8 Wronskian and the Liouville formula

Let I be an open interval in $\mathbb{R}$.

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.

Definition. Let $x_1, ..., x_n$ are n real-valued functions on I, which are $n-1$ times differentiable on I.. 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 2.12 (*a*) *Let $x_1, ..., x_n$ be a sequence of $\mathbb{R}^n$-valued functions that solve a linear system $x' = A(t)x$, and let $W(t)$ be their Wronskian. 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.*

(*b*) *Let $x_1, ..., x_n$ be a sequence of real-valued functions that solve a linear scalar ODE*

$$x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = 0,$$

and let $W(t)$ be their Wronskian. 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. (*a*) Indeed, if the functions $x_1, ..., x_n$ are linearly independent then, by Lemma 2.10, the vectors $x_1(t), ..., x_n(t)$ are linearly independent for any value of t, which implies $W(t) \neq 0$. If the functions $x_1, ..., x_n$ are linearly dependent then also the vectors $x_1(t), ..., x_n(t)$ are linearly dependent for any t, whence $W(t) \equiv 0$.

(*b*) 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 a vector ODE $\mathbf{x}' = A(t)\mathbf{x}$. The Wronskian of $\mathbf{x}_1, ..., \mathbf{x}_n$ 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 part (*a*). ■

Theorem 2.13 (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*

$$W(t) = W(t_0) \exp\left(\int_{t_0}^{t} \operatorname{trace} A(\tau)\, d\tau\right), \tag{2.77}$$

for all $t, t_0 \in I$.

Recall that the trace[46] $\operatorname{trace} A$ of the matrix A is the sum of all the diagonal entries of the matrix.

Corollary. *Consider a scalar 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

$$W(t) = W(t_0) \exp\left(-\int_{t_0}^{t} a_1(\tau)\, d\tau\right). \tag{2.78}$$

[46] die Spur

Proof of Theorem 2.13. Let the entries of the matrix $(x_1|\ x_2|...|x_n)$ be x_{ij} where i is the row index and j is the column index; in particular, the components of the vector x_j are $x_{1j}, x_{2j}, ..., x_{nj}$. Denote by r_i the i-th row of this matrix, that is, $r_i = (x_{i1}, x_{i2}, ..., x_{in})$; then

$$W = \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n \end{pmatrix}$$

We use the following formula for differentiation of the determinant, which follows from the full expansion of the determinant and the product rule:

$$W'(t) = \det \begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} + \det \begin{pmatrix} r_1 \\ r_2' \\ ... \\ r_n \end{pmatrix} + ... + \det \begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n' \end{pmatrix}. \tag{2.79}$$

Indeed, if $f_1(t), ..., f_n(t)$ are real-valued differentiable functions then the product rule implies by induction

$$(f_1...f_n)' = f_1'f_2...f_n + f_1f_2'...f_n + ... + f_1f_2...f_n'.$$

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, that is, (2.79).

The fact that each vector x_j satisfies the equation $x_j' = Ax_j$ can be written in the coordinate form as follows

$$x_{ij}' = \sum_{k=1}^{n} A_{ik}x_{kj}. \tag{2.80}$$

For any fixed i, the sequence $\{x_{ij}\}_{j=1}^{n}$ is nothing other than the components of the row r_i. Since the coefficients A_{ik} do not depend on j, (2.80) implies the same 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,

$$\det\begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} = A_{11}\det\begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n \end{pmatrix} = A_{11}W(t).$$

Evaluating similarly the other terms in (2.79), we obtain

$$W'(t) = (A_{11} + A_{22} + ... + A_{nn})\,W(t) = (\operatorname{trace} A)\,W(t).$$

By Lemma 2.12, $W(t)$ is either identical 0 or never zero. In the first case there is nothing to prove. In the second case, we can solve the above ODE using the method of separation of variables. Indeed, dividing it $W(t)$ and integrating in t, we obtain

$$\ln\frac{W(t)}{W(t_0)} = \int_{t_0}^{t} \operatorname{trace} A(\tau)\,d\tau$$

(note that $W(t)$ and $W(t_0)$ have the same sign so that the argument of ln is positive), whence (2.77) follows. ■

Proof of Corollary. The scalar ODE is equivalent to the normal system $\mathbf{x}' = A\mathbf{x}$ where

$$A = \begin{pmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & ... & -a_1 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x \\ x' \\ ... \\ x^{(n-1)} \end{pmatrix}.$$

Since the Wronskian of the normal system coincides with $W(t)$, (2.78) follows from (2.77) 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 (2.78)

$$\det\begin{pmatrix} x_0 & x \\ x_0' & x' \end{pmatrix} = C\exp\left(-\int a_1(t)\,dt\right),$$

that is

$$x_0x' - xx_0' = C\exp\left(-\int a_1(t)\,dt\right).$$

Using the identity

$$\frac{x_0x' - xx_0'}{x_0^2} = \left(\frac{x}{x_0}\right)'$$

we obtain the ODE

$$\left(\frac{x}{x_0}\right)' = \frac{C\exp\left(-\int a_1(t)\,dt\right)}{x_0^2}, \tag{2.81}$$

and by integrating it we obtain $\frac{x}{x_0}$ and, hence, x.

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 (2.81)

$$\left(\frac{x}{\tan t}\right)' = \frac{C}{\tan^2 t}$$

whence[47]

$$x = C\tan t\int\frac{dt}{\tan^2 t} = C\tan t\left(-t-\cot t + C_1\right).$$

The answer can also be written in the form

$$x(t) = c_1\left(t\tan t + 1\right) + c_2\tan t$$

where $c_1 = -C$ and $c_2 = CC_1$.

2.9 Linear homogeneous systems with constant coefficients

Consider a normal homogeneous system $x' = Ax$ where $A \in \mathbb{C}^{n\times n}$ is a constant $n\times n$ matrix with complex entries and $x(t)$ is a function from $\mathbb{R}$ to $\mathbb{C}^n$. As we know, the general solution of this system can be obtained as a linear combination of n linearly independent solutions. Here we will be concerned with construction of such solutions.

2.9.1 A simple case

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 $\mathbb{C}^n$ that does not depend on t. Then the equation $x' = Ax$ becomes

$$\lambda e^{\lambda t}v = e^{\lambda t}Av$$

that is, $Av = \lambda v$. Recall that any non-zero vector v, that satisfies the identity $Av = \lambda v$ for some constant λ, is called an *eigenvector* of A, and λ is called the *eigenvalue*[48]. Hence,

[47] 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

$$\int\cot^2 t dt = -t - \cot t + C.$$

[48] der Eigenwert

the function $x(t) = e^{\lambda t}v$ is a non-trivial solution to $x' = Ax$ provided v is an eigenvector of A and λ is the corresponding eigenvalue.

It is easy to see that λ is an eigenvalue if and only if the matrix $A - \lambda\,\mathrm{id}$ is not invertible, that is, when

$$\det(A - \lambda\,\mathrm{id}) = 0, \tag{2.82}$$

where id is the $n \times n$ identity matrix. This equation is called the *characteristic equation* of the matrix A and can be used to determine the eigenvalues. Then the eigenvector is determined from the equation

$$(A - \lambda\,\mathrm{id})\,v = 0. \tag{2.83}$$

Note that the eigenvector is not unique; for example, if v is an eigenvector then cv is also an eigenvector for any constant $c \neq 0$.

The function

$$P(\lambda) := \det(A - \lambda\,\mathrm{id})$$

is a polynomial of λ of order n, that is called the *characteristic polynomial* of the matrix A. Hence, the eigenvalues of A are the roots of the characteristic polynomial $P(\lambda)$.

Lemma 2.14 *If a $n \times n$ matrix A has n linearly independent eigenvectors $v_1, ..., v_n$ with the (complex) eigenvalues $\lambda_1, ..., \lambda_n$ then the following functions*

$$e^{\lambda_1 t}v_1,\ e^{\lambda_2 t}v_2, ..., e^{\lambda_n t}v_n \tag{2.84}$$

are n linearly independent solutions of the system $x' = Ax$. Consequently, the general solution of this system is

$$x(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t} v_k,$$

where $C_1, ..., C_k$ are arbitrary complex constants..

If A is a real matrix and λ is a non-real eigenvalue of A with an eigenvector v then $\overline{\lambda}$ is an eigenvalue with eigenvector $\overline{v}$, and the terms $e^{\lambda t}v, e^{\overline{\lambda} t}\overline{v}$ in (2.84) *can be replaced by the couple* $\operatorname{Re}\left(e^{\lambda t}v\right)$, $\operatorname{Im}\left(e^{\lambda t}v\right)$.

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 first claim follows from Theorem 2.3.

If $Av = \lambda v$ then applying the complex conjugation and using the fact the entries of A are real, we obtain $A\overline{v} = \overline{\lambda}\overline{v}$ so that $\overline{\lambda}$ is an eigenvalue with eigenvector $\overline{v}$. Since the functions $e^{\lambda t}v$ and $e^{\overline{\lambda} t}\overline{v}$ are solutions, their linear combinations

$$\operatorname{Re} e^{\lambda t}v = \frac{e^{\lambda t}v + e^{\overline{\lambda} t}\overline{v}}{2} \quad \text{and} \quad \operatorname{Im} e^{\lambda t}v = \frac{e^{\lambda t}v - e^{\overline{\lambda} t}\overline{v}}{2i}$$

are also solutions. Since $e^{\lambda t}v$ and $e^{\overline{\lambda} t}\overline{v}$ can also be expressed via these solutions:

$$\begin{aligned} e^{\lambda t}v &= \operatorname{Re} e^{\lambda t}v + i \operatorname{Im} e^{\lambda t}v \\ e^{\overline{\lambda} t}\overline{v} &= \operatorname{Re} e^{\lambda t}v - i \operatorname{Im} e^{\lambda t}v, \end{aligned}$$

replacing in (2.84) the terms $e^{\lambda t}, e^{\overline{\lambda} t}$ by the couple $\operatorname{Re}\left(e^{\lambda t}v\right)$, $\operatorname{Im}\left(e^{\lambda t}v\right)$ results in again n linearly independent solutions. ■

It is known from Linear Algebra that if A has n distinct eigenvalues then their eigenvectors are automatically linearly independent, and Lemma 2.14 applies. Another case when the hypotheses of Lemma 2.14 are satisfied is if A is a symmetric real matrix; in this case there is always a basis of the eigenvectors.

Example. Consider the system

$$\begin{cases} x' = y \\ y' = x \end{cases}.$$

The vector form of this system is $\mathbf{x} = A\mathbf{x}$ where $\mathbf{x} = \binom{x}{y}$ and $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The characteristic polynomial is

$$P(\lambda) = \det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 - 1,$$

the characteristic equation is $\lambda^2 - 1 = 0$, whence the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = -1$. For $\lambda = \lambda_1 = 1$ we obtain the equation (2.83) for $v = \binom{a}{b}$:

$$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

which gives only one independent equation $a - b = 0$. Choosing $a = 1$, we obtain $b = 1$ whence

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Similarly, for $\lambda = \lambda_2 = -1$ we have the equation for $v = \binom{a}{b}$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

which amounts to $a + b = 0$. Hence, the eigenvector for $\lambda_2 = -1$ is

$$v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Since the vectors v_1 and v_2 are independent, we obtain the general solution in the form

$$\mathbf{x}(t) = C_1 e^t \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2 e^{-t} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} C_1 e^t + C_2 e^{-t} \\ C_1 e^t - C_2 e^{-t} \end{pmatrix},$$

that is, $x(t) = C_1 e^t + C_2 e^{-t}$ and $y(t) = C_1 e^t - C_2 e^{-t}$.

Example. Consider the system

$$\begin{cases} x' = -y \\ y' = x \end{cases}.$$

The matrix of the system is $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and the the characteristic polynomial is

$$P(\lambda) = \det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1.$$

Hence, the characteristic equation is $\lambda^2+1=0$ whence $\lambda_1=i$ and $\lambda_2=-i$. For $\lambda=\lambda_1=i$ we obtain the equation for the eigenvector $v=\binom{a}{b}$

$$\left(\begin{array}{cc} -i & -1 \\ 1 & -i \end{array}\right)\begin{pmatrix} a \\ b \end{pmatrix}=0,$$

which amounts to the single equation $ia+b=0$. Choosing $a=i$, we obtain $b=1$, whence

$$v_1=\begin{pmatrix} i \\ 1 \end{pmatrix}$$

and the corresponding solution of the ODE is

$$\mathbf{x}_1\left(t\right)=e^{it}\begin{pmatrix} i \\ 1 \end{pmatrix}=\begin{pmatrix} -\sin t+i\cos t \\ \cos t+i\sin t \end{pmatrix}.$$

Since this solution is complex, we obtain the general solution using the second claim of Lemma 2.14:

$$\mathbf{x}\left(t\right)=C_1\operatorname{Re}\mathbf{x}_1+C_2\operatorname{Im}\mathbf{x}_1=C_1\begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}+C_2\begin{pmatrix} \cos t \\ \sin t \end{pmatrix}=\begin{pmatrix} -C_1\sin t+C_2\cos t \\ C_1\cos t+C_2\sin t \end{pmatrix}.$$

Example. Consider a normal system

$$\begin{cases} x'=y \\ y'=0. \end{cases}$$

This system is trivially solved to obtain $y=C_1$ and $x=C_1t+C_2$. However, if we try to solve it using the above method, we fail. Indeed, the matrix of the system is $A=\left(\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right)$, the characteristic polynomial is

$$P\left(\lambda\right)=\det\left(\begin{array}{cc} -\lambda & 1 \\ 0 & -\lambda \end{array}\right)=\lambda^2,$$

and the characteristic equation $P\left(\lambda\right)=0$ yields only one eigenvalue $\lambda=0$. The eigenvector $v=\binom{a}{b}$ satisfies the equation

$$\left(\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right)\begin{pmatrix} a \\ b \end{pmatrix}=0,$$

whence $b=0$. That is, the only eigenvector (up to a constant multiple) is $v=\binom{1}{0}$, and the only solution we obtain in this way is $\mathbf{x}\left(t\right)=\binom{1}{0}$. The problem lies in the properties of this matrix: it does not have a basis of eigenvectors, which is needed for this method.

In order to handle such cases, we use a different approach.

2.9.2 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 linear operator in $\mathbb{R}^n$ then we may be able to use this formula if we define what is e^{At}. In this section, we define the exponential function e^A for any linear operator A in $\mathbb{C}^n$. Denote by $\mathcal{L}(\mathbb{C}^n)$ the space of all linear operators from $\mathbb{C}^n$ to $\mathbb{C}^n$.

Fix a norm in $\mathbb{C}^n$ and recall that the operator norm of an operator $A \in \mathcal{L}(\mathbb{C}^n)$ is defined by

$$\|A\| = \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{2.85}$$

The operator norm is a norm in the linear space $\mathcal{L}(\mathbb{C}^n)$. Note that in addition to the linear operations $A + B$ and cA, that are defined for all $A, B \in \mathcal{L}(\mathbb{C}^n)$ and $c \in \mathbb{C}$, the space $\mathcal{L}(\mathbb{C}^n)$ enjoys the operation of the product of operators AB that is the composition of A and B:

$$(AB)x := A(Bx)$$

and that clearly belongs to $\mathcal{L}(\mathbb{C}^n)$ as well. The operator norm satisfies the following multiplicative inequality

$$\|AB\| \leq \|A\| \|B\|. \tag{2.86}$$

Indeed, it follows from (2.85) that $\|Ax\| \leq \|A\| \|x\|$ whence

$$\|(AB)x\| = \|A(Bx)\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\|,$$

which yields (2.86).

Definition. If $A \in \mathcal{L}(\mathbb{C}^n)$ then define $e^A \in \mathcal{L}(\mathbb{C}^n)$ by means of the identity

$$e^A = \mathrm{id} + A + \frac{A^2}{2!} + ... + \frac{A^k}{k!} + ... = \sum_{k=0}^{\infty} \frac{A^k}{k!}, \tag{2.87}$$

where id is the identity operator.

Of course, in order to justify this definition, we need to verify the convergence of the series (2.87).

Lemma 2.15 *The exponential series* (2.87) *converges for any* $A \in \mathcal{L}(\mathbb{C}^n)$.

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 (2.86) 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. ■

Theorem 2.16 *For any $A \in \mathcal{L}(\mathbb{C}^n)$ the function $F(t) = e^{tA}$ satisfies the ODE $F' = AF$. Consequently, the general solution of the ODE $x' = Ax$ is given by $x = e^{tA}v$ where $v \in \mathbb{C}^n$ is an arbitrary vector.*

Here $x = x(t)$ is as usually a $\mathbb{C}^n$-valued function on $\mathbb{R}$, while $F(t)$ is an $\mathcal{L}(\mathbb{C}^n)$-valued function on $\mathbb{R}$. Since $\mathcal{L}(\mathbb{C}^n)$ is linearly isomorphic to $\mathbb{C}^{n^2}$, we can also say that $F(t)$ is a $\mathbb{C}^{n^2}$-valued function on $\mathbb{R}$, which allows to understand the ODE $F' = AF$ in the same sense as general vectors ODE. The novelty here is that we regard $A \in \mathcal{L}(\mathbb{C}^n)$ as an operator in $\mathcal{L}(\mathbb{C}^n)$ (that is, an element of $\mathcal{L}(\mathcal{L}(\mathbb{C}^n))$) by means of the operator multiplication.

Proof. We have by definition

$$F(t) = e^{tA} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

Consider the series of the derivatives:

$$G(t) := \sum_{k=0}^{\infty} \frac{d}{dt}\left(\frac{t^k A^k}{k!}\right) = \sum_{k=1}^{\infty} \frac{t^{k-1}A^k}{(k-1)!} = A\sum_{k=1}^{\infty} \frac{t^{k-1}A^{k-1}}{(k-1)!} = AF.$$

It is easy to see (in the same way as Lemma 2.15) that this series converges locally uniformly in t, which implies that F is differentiable in t and $F' = G$. It follows that $F' = AF$.

For function $x(t) = e^{tA}v$, we have

$$x' = \left(e^{tA}\right)' v = \left(Ae^{tA}\right) v = Ax$$

so that $x(t)$ solves the ODE $x' = Ax$ for any v.

If $x(t)$ is any solution to $x' = Ax$ then set $v = x(0)$ and observe that the function $e^{tA}v$ satisfies the same ODE and the initial condition

$$e^{tA}v|_{t=0} = \operatorname{id} v = v.$$

Hence, both $x(t)$ and $e^{tA}v$ solve the same initial value problem, whence the identity $x(t) = e^{tA}v$ follows by the uniqueness part of Theorem 2.1. ■

Remark. If $v_1, ..., v_n$ are linearly independent vectors in $\mathbb{C}^n$ then the solutions $e^{tA}v_1, ..., e^{tA}v_n$ are also linearly independent. If $v_1, ..., v_n$ is the canonical basis in $\mathbb{C}^n$, then $e^{tA}v_k$ is the k-th column of the matrix e^{tA}. Hence, the columns of the matrix e^{tA} form n linearly independent solutions, that is, e^{tA} is the 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

$$e^{tA} = \operatorname{diag}\left(e^{\lambda_1 t}, ..., e^{\lambda_n t}\right).$$

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

$$e^{tA} = \operatorname{id} + tA = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Hence, the general solution to $x' = Ax$ is

$$x(t) = e^{tA}v = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} C_1 + C_2 t \\ C_2 \end{pmatrix},$$

where C_1, C_2 are the components of v.

Definition. Operators $A, B \in \mathcal{L}(\mathbb{C}^n)$ 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,

$$(A + B)^2 = A^2 + 2AB + B^2. \tag{2.88}$$

Indeed, in general we have

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2,$$

which yields (2.88) if $AB = BA$.

Lemma 2.17 *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, we have

$$(AC)B = A(CB) = A(BC) = (AB)C = (BA)C = B(AC).$$

Claim 2. *If A and B commute then so do e^A and B.*

It follows from Claim 1 that A^k and B commute for any natural k, whence

$$e^A B = \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) B = B \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) = B e^A.$$

Claim 3. *If $A(t)$ and $B(t)$ are differentiable functions from $\mathbb{R}$ to $\mathcal{L}(\mathbb{C}^n)$ then*

$$(A(t)B(t))' = A'(t)B(t) + A(t)B'(t). \tag{2.89}$$

Indeed, we have for any component

$$(AB)'_{ij} = \left(\sum_k A_{ik}B_{kj}\right)' = \sum_k A'_{ik}B_{kj} + \sum_k A_{ik}B'_{kj} = (A'B)_{ij} + (AB')_{ij} = (A'B + AB')_{ij},$$

whence (2.89) follows.

Now we can finish the proof of the lemma. Consider the function $F : \mathbb{R} \to \mathcal{L}(\mathbb{C}^n)$ defined by

$$F(t) = e^{tA}e^{tB}.$$

Differentiating it using Theorem 2.16, 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, by Theorem 2.16, the function $G(t) = e^{t(A+B)}$ satisfies the same equation

$$G' = (A+B)G.$$

Since $G(0) = F(0) = \text{id}$, 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 $e^Ae^B = e^{A+B}$. ■

Alternative proof. Let us briefly discuss a direct algebraic proof of $e^{A+B} = e^Ae^B$. 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

$$e^{A+B} = \sum_{n=0}^\infty \frac{(A+B)^n}{n!} = \sum_{n=0}^\infty \sum_{k=0}^n \frac{A^k B^{n-k}}{k!(n-k)!}.$$

On the other hand, using the Cauchy product formula for absolutely convergent series of operators, we have

$$e^Ae^B = \sum_{m=0}^\infty \frac{A^m}{m!} \sum_{l=0}^\infty \frac{B^l}{l!} = \sum_{n=0}^\infty \sum_{k=0}^n \frac{A^k B^{n-k}}{k!(n-k)!},$$

whence the identity $e^{A+B} = e^Ae^B$ follows.

2.9.3 Jordan cells

Here we show how to compute e^A provided A is a Jordan cell.

Definition. An $n \times n$ matrix J is called a *Jordan cell* if it has the form

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}, \tag{2.90}$$

where λ is any complex number. That is, all the entries on the main diagonal of J are λ, all the entries on the second diagonal are 1, and all other entries of J are 0.

Let us use Lemma 2.17 to evaluate e^{tJ}. Clearly, we have $J = \lambda \,\mathrm{id} + N$ where

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}. \tag{2.91}$$

The matrix (2.91) is called a *nilpotent Jordan cell.* Since the matrices $\lambda \,\mathrm{id}$ and N commute (because id commutes with any matrix), Lemma 2.17 yields

$$e^{tA} = e^{t\lambda \,\mathrm{id}} e^{tN} = e^{t\lambda} e^{tN}. \tag{2.92}$$

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

$$N_{ij} = \left\{ \begin{array}{ll} 1, & \text{if } j = i+1 \\ 0, & \text{otherwise} \end{array} \right. ,$$

where i is the row index and j is the column index. It follows that

$$\left(N^2\right)_{ij} = \sum_{k=1}^{n} N_{ik} N_{kj} = \left\{ \begin{array}{ll} 1, & \text{if } j = i+2 \\ 0, & \text{otherwise} \end{array} \right.$$

that is,

$$N^2 = \begin{pmatrix} 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix},$$

where the entries with value 1 are located on the 3rd diagonal, that is, two positions above the main diagonal, and all other entries are 0. Similarly, we obtain

$$N^k \;\; = \begin{array}{c} \begin{array}{ccccc} & \overset{k+1}{\searrow} & & & \end{array} \\ \begin{pmatrix} 0 & \ddots & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \end{array}$$

where the entries with value 1 are located on the $(k+1)$-st diagonal, that is, k positions above the main diagonal, provided $k < n$, and $N^k = 0$ if $k \geq n$.[49] It follows that

$$e^{tN} = \mathrm{id} + \frac{t}{1!}N + \frac{t^2}{2!}N^2 + ... + \frac{t^{n-1}}{(n-1)!}N^{n-1} = \begin{pmatrix} 1 & \frac{t}{1!} & \frac{t^2}{2!} & \ddots & \frac{t^{n-1}}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!} \\ \vdots & & \ddots & \ddots & \frac{t}{1!} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}. \quad (2.93)$$

Combining with (2.92), we obtain the following statement.

Lemma 2.18 *If J is a Jordan cell* (2.90) *then, for any $t \in \mathbb{R}$,*

$$e^{tJ} = \begin{pmatrix} e^{\lambda t} & \frac{t}{1!}e^{t\lambda} & \frac{t^2}{2!}e^{t\lambda} & \ddots & \frac{t^{n-1}}{(n-1)!}e^{t\lambda} \\ 0 & e^{t\lambda} & \frac{t}{1!}e^{t\lambda} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!}e^{t\lambda} \\ \vdots & & \ddots & \ddots & \frac{t}{1!}e^{t\lambda} \\ 0 & \cdots & \cdots & 0 & e^{t\lambda} \end{pmatrix}. \quad (2.94)$$

Taking the columns of (2.94), we obtain n linearly independent solutions of the system $x' = Jx$:

$$x_1(t) = \begin{pmatrix} e^{\lambda t} \\ 0 \\ \cdots \\ \cdots \\ 0 \end{pmatrix}, \ x_2(t) = \begin{pmatrix} \frac{t}{1!}e^{\lambda t} \\ e^{\lambda t} \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \ x_3(t) = \begin{pmatrix} \frac{t^2}{2!}e^{\lambda t} \\ \frac{t}{1!}e^{\lambda t} \\ e^{\lambda t} \\ \cdots \\ 0 \end{pmatrix}, \ \ldots, \ x_n(t) = \begin{pmatrix} \frac{t^{n-1}}{(n-1)!}e^{\lambda t} \\ \cdots \\ \frac{t^2}{2!}e^{\lambda t} \\ \frac{t}{1!}e^{\lambda t} \\ e^{\lambda t} \end{pmatrix}.$$

[49] 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”.

2.9.4 The Jordan normal form

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 2.19 *The following identity is true:*

$$e^{A\otimes B} = e^A \otimes e^B. \tag{2.95}$$

In extended notation, (2.95) means that

$$e^C = \left(\begin{array}{|c|c|}\hline e^A & 0 \\ \hline 0 & e^B \\ \hline\end{array}\right).$$

Proof. Observe first that if A_1, A_2 are $m \times m$ matrices and B_1, B_2 are $l \times l$ matrices then

$$(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1A_2) \otimes (B_1B_2). \tag{2.96}$$

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. The tensor product of a finite number of Jordan cells is called a *Jordan normal form.*

That is, if a Jordan normal form is a matrix as follows:

$$J_1 \otimes J_2 \otimes \cdots \otimes J_k = \begin{pmatrix} J_1 & & & & \\ & J_2 & & 0 & \\ & & \ddots & & \\ & 0 & & J_{k-1} & \\ & & & & J_k \end{pmatrix},$$

where J_j are Jordan cells.

Lemmas 2.18 and 2.19 allow to evaluate e^{tA} if 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 2.18, 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 2.19,

$$e^{tA} = \begin{pmatrix} e^t & te^t & 0 & 0 \\ 0 & e^t & 0 & 0 \\ 0 & 0 & e^{2t} & te^{2t} \\ 0 & 0 & 0 & e^{2t} \end{pmatrix}.$$

The columns of this matrix form 4 linearly independent solutions, and the general solution is their linear combination:

$$x(t) = \begin{pmatrix} C_1e^t + C_2te^t \\ C_2e^t \\ C_3e^{2t} + C_4te^{2t} \\ C_4e^{2t} \end{pmatrix}.$$

2.9.5 Transformation of an operator to a Jordan normal form

Given a basis $b = \{b_1, b_2, ..., b_n\}$ in $\mathbb{C}^n$ and a vector $x \in \mathbb{C}^n$, denote by x^b the column vector that represents x in this basis. That is, if x_j^b is the j-th component of x^b then

$$x = x_1^b b_1 + x_2^b b_2 + ... + x_n^b b_n = \sum_{j=1}^n x_j^b b_j.$$

Similarly, if A is a linear operator in $\mathbb{C}^n$ then denote by A^b the matrix that represents A in the basis b. It is determined by the identity

$$(Ax)^b = A^b x^b,$$

which should be true for all $x \in \mathbb{C}^n$, where in the right hand side we have the product of the $n \times n$ matrix A^b and the column-vector x^b.

Clearly, $(b_j)^b = (0, ...1, ...0)$ where 1 is at position j, which implies that $(Ab_j)^b = A^b (b_j)^b$ is the j-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 j-th column of A^b is the column vector Ab_j written in the basis $b_1, ..., b_n$.

Example. Consider the operator A in $\mathbb{C}^2$ that is given in the canonical basis $e = \{e_1, e_2\}$ by the matrix

$$A^e = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

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

$$(Ab_1)^e = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

and

$$(Ab_2)^e = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

It follows that $Ab_1 = -b_1$ and $Ab_2 = b_2$ whence

$$A^b = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

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 a Jordan normal form.*

The basis b is called the Jordan basis of A, and the matrix A^b is called the Jordan normal form of A.

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 is the disjoint union of the Jordan chains.

Since

$$A^b - \lambda\,\mathrm{id} \quad = \begin{array}{cccccccccc}
 & & & j & \cdots & \cdots & j{+}p{-}1 & & & \\
 & & & \downarrow & & & \downarrow & & & \\
\left(\right. & \ddots & & & & & & & \left.\right) & \\
 & & \ddots & & & & & & & \\
 & & & 0 & 1 & \cdots & 0 & & & \leftarrow j \\
 & & & 0 & \ddots & \ddots & \vdots & & & \cdots \\
 & & & \vdots & \ddots & \ddots & 1 & & & \cdots \\
 & & & 0 & \cdots & 0 & 0 & & & \leftarrow j{+}p{-}1 \\
 & & & & & & & \ddots & & \\
 & & & & & & & & \ddots &
\end{array}$$

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

$$\begin{aligned}
(A - \lambda \operatorname{id})\, b_j &= 0 \\
(A - \lambda \operatorname{id})\, b_{j+1} &= b_j \\
(A - \lambda \operatorname{id})\, b_{j+2} &= b_{j+1} \\
&\dots \\
(A - \lambda \operatorname{id})\, b_{j+p-1} &= b_{j+p-2}.
\end{aligned}$$

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 2.20 *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 2.18 and 2.19

$$e^{tA^b} = \begin{pmatrix}
\ddots & & & & \\
& \boxed{\ddots} & & & \\
& & \boxed{\begin{array}{cccc} e^{\lambda t} & \frac{t}{1!} e^{t\lambda} & \cdots & \frac{t^{p-1}}{(p-1)!} e^{t\lambda} \\ 0 & e^{t\lambda} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{t}{1!} e^{t\lambda} \\ 0 & \cdots & 0 & e^{t\lambda} \end{array}} & & \\
& & & \boxed{\ddots} & \\
& & & & \ddots
\end{pmatrix},$$

where the block in the middle is e^{tJ}. By Theorem 2.16, the columns of this matrix give 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,

$$x_1(t) = \begin{pmatrix} \dots \\ \boxed{\begin{matrix} e^{\lambda t} \\ 0 \\ \dots \\ \dots \\ 0 \end{matrix}} \\ \dots \end{pmatrix}, \ x_2 = \begin{pmatrix} \dots \\ \boxed{\begin{matrix} \frac{t}{1!}e^{\lambda t} \\ e^{\lambda t} \\ 0 \\ \dots \\ 0 \end{matrix}} \\ \dots \end{pmatrix}, \ \dots, \ x_p = \begin{pmatrix} \dots \\ \boxed{\begin{matrix} \frac{t^{p-1}}{(p-1)!}e^{\lambda t} \\ \dots \\ \frac{t^2}{2!}e^{\lambda t} \\ \frac{t}{1!}e^{\lambda t} \\ e^{\lambda t} \end{matrix}} \\ \dots \end{pmatrix},$$

where all the vectors are written in the basis b, the boxed rows correspond to the cell J, and all the terms outside boxes 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 2.20. ■

Let λ be an eigenvalue of an operator A. Denote by m the *algebraic multiplicity* of λ, that is, its multiplicity as a root of characteristic polynomial[50] $P(\lambda) = \det(A - \lambda \operatorname{id})$. Denote by g the *geometric multiplicity* of λ, that is the dimension of the eigenspace of λ:

$$g = \dim\ker(A - \lambda \operatorname{id}).$$

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[51] of A^b, whereas g is equal to the number of the Jordan cells with λ on the diagonal[52]. It follows that $g \le m$, and the equality occurs if and only if all the Jordan cells with the eigenvalue λ have dimension 1.

Consider some examples of application of Theorem 2.20.

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 eigenvector v is

$$(A - \lambda \operatorname{id})\, v = 0$$

[50] To compute $P(\lambda)$, one needs to write the operator A in some basis b as a matrix A_b and then evaluate $\det(A_b - \lambda \operatorname{id})$. 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 $A_b - \lambda \operatorname{id} = C(A_{b'} - \lambda \operatorname{id})C^{-1}$ whence $\det(A_b - \lambda \operatorname{id}) = \det C \det(A_{b'} - \lambda \operatorname{id}) \det C^{-1} = \det(A_{b'} - \lambda \operatorname{id})$.

[51] 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, $k = m$.

[52] Note that each Jordan cell correponds to exactly one eigenvector.

that is, setting $v = (a, b)$,

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

which is equivalent to $-a + b = 0$. Choosing $a = 1$ 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 $g_1 = 1$. Therefore, 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 2.20, we obtain the solutions

$$\begin{aligned} x_1(t) &= e^{3t} v_1 \\ x_2(t) &= e^{3t} (t v_1 + v_2) \end{aligned}$$

where v_2 is the 1st generalized eigenvector that can be determined from the equation

$$(A - \lambda \operatorname{id}) v_2 = v_1.$$

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

$$(A-\lambda_2\,\mathrm{id})\,v = 0,$$

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 2.20, 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

$$x_3(t) = e^t (tv_1 + v_2),$$

where v_2 is the first generalized eigenvector to be determined from the equation

$$(A - \lambda_2 \, \mathrm{id}) v_2 = v_1.$$

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

$$x_3(t) = e^t (tv_1 + v_2) = e^t \begin{pmatrix} -1 \\ t \\ 1 - t \end{pmatrix}.$$

Finally, the general solution is

$$x(t) = C_1 x_1 + C_2 x_2 + C_3 x_3 = \begin{pmatrix} C_1 e^{2t} - C_3 e^t \\ -2C_1 e^{2t} + (C_2 + C_3 t) e^t \\ 2C_1 e^{2t} + (C_3 - C_2 - C_3 t) e^t \end{pmatrix}.$$

Corollary. *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*

$$x(t) = e^{\lambda t} \left(u_1 + u_2 t + \ldots + u_s t^{s-1} \right) \tag{2.97}$$

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.

Remark. For practical use, one should substitute (2.97) 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 (2.97) 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 $p_1 + ... + p_g$ 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 2.20 for the eigenvalue λ is equal to m. Let us show that each of the solutions of Theorem 2.20 has the form (2.97). Indeed, each solution of Theorem 2.20 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 (2.97), 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. ■

3 The initial value problem for general ODEs

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

$$\begin{cases} x' = f(t, x), \\ x(t_0) = x_0, \end{cases} \tag{3.1}$$

where (t_0, x_0) is a given point in Ω. A function $x(t) : I \to \mathbb{R}^n$ is called a solution (3.1) if the domain I is an open interval in $\mathbb{R}$ containing t_0, $x(t)$ is differentiable in t in I, $(t, x(t)) \in \Omega$ for all $t \in I$, and $x(t)$ satisfies the identity $x'(t) = f(t, x(t))$ for all $t \in I$ 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). It is also called the integral curve of the ODE $x' = f(t, x)$.

Our aim here is to prove the unique solvability of (3.1) under reasonable assumptions about the function $f(t, x)$, and to investigate the dependence of the solution on the initial condition and on other possible parameters.

3.1 Existence and uniqueness

Let Ω be an open subset of $\mathbb{R}^{n+1}$ and $f = f(t, x)$ be a mapping from Ω to $\mathbb{R}^n$. We start with a description of the class of functions $f(t, x)$, which will be used in all main theorems.

Definition. Function $f(t, x)$ is called *Lipschitz* in x in Ω if there is a constant L such that for all $(t, x), (t, y) \in \Omega$

$$\|f(t, x) - f(t, y)\| \leq L \|x - y\|. \tag{3.2}$$

The constant L is called the *Lipschitz constant* of f in Ω.

In the view of the equivalence of any two norms in $\mathbb{R}^n$, the property to be Lipschitz does not depend on the choice of the norm (but the value of the Lipschitz constant L does).

A subset G of $\mathbb{R}^{n+1}$ will be called a *cylinder* if it has the form $G = I \times B$ where I is an interval in $\mathbb{R}$ and B is a ball (open or closed) in $\mathbb{R}^n$. The cylinder is closed if both I and B are closed, and open if both I and B are open.

Definition. Function $f(t, x)$ is called *locally Lipschitz* in x in Ω if, for any $(t_0, x_0) \in \Omega$, there exist constants $\varepsilon, \delta > 0$ such that the cylinder

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

is contained in Ω and f is Lipschitz in x in G.

Note that the Lipschitz constant may be different for different cylinders G.

Example. The function $f(t, x) = \|x\|$ is Lipschitz in x in $\mathbb{R}^n$ because by the triangle inequality

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

The Lipschitz constant is equal to 1. Note that the function $f(t,x)$ is not differentiable at $x=0$.

Many examples of locally Lipschitz functions can be obtained by using the following lemma.

Lemma 3.1 (a) *If all components f_k of f are differentiable functions in a cylinder G and all the partial derivatives $\frac{\partial f_k}{\partial x_i}$ are bounded in G then the function $f(t,x)$ is Lipschitz in x in G.*

(b) *If all partial derivatives $\frac{\partial f_k}{\partial x_j}$ exists and are continuous in Ω then $f(t,x)$ is locally Lipschitz in x in Ω.*

Proof. Let us use the following mean value property of functions in $\mathbb{R}^n$: if g is a differentiable real valued function in a ball $B\subset\mathbb{R}^n$ then, for all $x,y\in B$ there is $\xi\in[x,y]$ such that

$$g(y)-g(x)=\sum_{j=1}^{n}\frac{\partial g}{\partial x_j}(\xi)(y_j-x_j) \tag{3.3}$$

(note that the interval $[x,y]$ is contained in the ball B so that $\frac{\partial g}{\partial x_j}(\xi)$ makes sense). Indeed, consider the function

$$h(t)=g(x+t(y-x)) \quad \text{where } t\in[0,1].$$

The function $h(t)$ is differentiable on $[0,1]$ and, by the mean value theorem in $\mathbb{R}$, there is $\tau\in(0,1)$ such that

$$g(y)-g(x)=h(1)-h(0)=h'(\tau).$$

Noticing that by the chain rule

$$h'(\tau)=\sum_{j=1}^{n}\frac{\partial g}{\partial x_j}(x+\tau(y-x))(y_j-x_j)$$

and setting $\xi=x+\tau(y-x)$, we obtain (3.3).

(a) Let $G=I\times B$ where I is an interval in $\mathbb{R}$ and B is a ball in $\mathbb{R}^n$. If $(t,x),(t,y)\in G$ then $t\in I$ and $x,y\in B$. Applying the above mean value property for the k-th component f_k of f, we obtain that

$$f_k(t,x)-f_k(t,y)=\sum_{j=1}^{n}\frac{\partial f_k}{\partial x_j}(t,\xi)(x_j-y_j), \tag{3.4}$$

where ξ is a point in the interval $[x,y]\subset B$. Set

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

and note that by the hypothesis $C<\infty$. Hence, by (3.4)

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

Taking max in k, we obtain

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

Switching in the both sides to the given norm $\|\cdot\|$ and using the equivalence of all norms, we obtain that f is Lipschitz in x in G.

(b) Given a point $(t_0, x_0) \in \Omega$, choose positive ε and δ so that the cylinder

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

is contained in Ω, which is possible by the openness of Ω. Since the components f_k are continuously differentiable, they are differentiable. Since G is a closed bounded set and the partial derivatives $\frac{\partial f_k}{\partial x_j}$ are continuous, they are bounded on G. By part (a) we conclude that f is Lipschitz in x in G, which finishes the proof. ∎

Now we can state the main existence and uniqueness theorem.

Theorem 3.2 (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 is locally Lipschitz in x. Then, for any point $(t_0, x_0) \in \Omega$, the initial value problem (3.1) *has a solution.*

Furthermore, if $x(t)$ and $y(t)$ are two solutions of (3.1) *then $x(t) = y(t)$ in their common domain.*

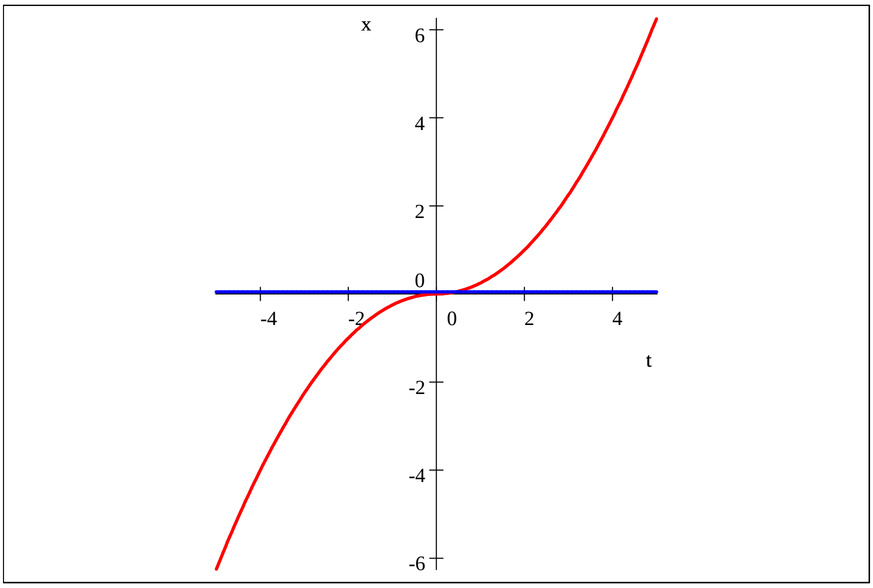
Remark. By Lemma 3.1, the hypothesis of Theorem 3.2 that f is locally Lipschitz in x could be replaced by a simpler hypotheses that all partial derivatives $\frac{\partial f_k}{\partial x_j}$ exist and are continuous in Ω. However, as we have seen above, there are examples of functions that are Lipschitz but not differentiable, and Theorem 3.2 applies to such functions as well.

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 considered in Section 1.1. 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.

The proof of Theorem 3.2 uses essentially the following abstract result.

Banach fix point theorem. *Let (X,d) be a complete metric space*[53] *and $f:X\to X$ be a contraction mapping*[54]*, that is,*

$$d\left(f\left(x\right),f\left(y\right)\right)\leq qd\left(x,y\right)$$

for some $q\in(0,1)$ and all $x,y\in X$. Then f has exactly one fixed point, that is, a point $x\in X$ such that $f\left(x\right)=x$.

Proof. Choose an arbitrary point $x\in X$ and define a sequence $\{x_n\}_{n=0}^{\infty}$ by induction using

$$x_0=x\ \text{ and } x_{n+1}=f\left(x_n\right)\text{ for any } n\geq 0.$$

Our purpose will be to show that the sequence $\{x_n\}$ converges and that the limit is a fixed point of f. We start with the observation that

$$d\left(x_{n+1},x_n\right)=d\left(f\left(x_n\right),f\left(x_{n-1}\right)\right)\leq qd\left(x_n,x_{n-1}\right).$$

It follows by induction that

$$d\left(x_{n+1},x_n\right)\leq q^nd\left(x_1,x_0\right)=Cq^n \tag{3.5}$$

where $C=d\left(x_1,x_0\right)$. Let us show that the sequence $\{x_n\}$ is Cauchy. Indeed, for any $m>n$, we obtain, using the triangle inequality and (3.5),

$$\begin{aligned}d\left(x_m,x_n\right)&\leq d\left(x_n,x_{n+1}\right)+d\left(x_{n+1},x_{n+2}\right)+...+d\left(x_{m-1},x_m\right)\\&\leq C\left(q^n+q^{n+1}+...+q^{m-1}\right)\\&\leq\frac{Cq^n}{1-q}.\end{aligned}$$

[53] Vollständiger metrische Raum

[54] Kontraktionsabbildung

Therefore, $d(x_m, x_n) \to 0$ as $n, m \to \infty$, that is, the sequence $\{x_n\}$ is Cauchy. By the definition of a complete metric space, any Cauchy sequence converges. Hence, we conclude that $\{x_n\}$ converges, and let $a = \lim_{n\to\infty} x_n$. Then

$$d(f(x_n), f(a)) \leq q d(x_n, a) \to 0 \text{ as } n \to \infty,$$

so that $f(x_n) \to f(a)$. On the other hand, $f(x_n) = x_{n+1} \to a$ as $n \to \infty$, whence it follows that $f(a) = a$, that is, a is a fixed point.

Let us prove the uniqueness of the fixed point. If b is another fixed point then

$$d(a, b) = d(f(a), f(b)) \leq q d(a, b),$$

which is only possible if $d(a, b) = 0$ and, hence, $a = b$. ■

Proof of Theorem 3.2. 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*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{3.6}$$

Here the integral of the vector valued function is understood component-wise. If x solves IVP then (3.6) 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 (3.6) follows. Conversely, if x is a continuous function that satisfies (3.6) 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 prove the existence of a solution of (3.1). Let ε, δ be the parameter from the the local Lipschitz condition at this point, that is, there is a constant L such that

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

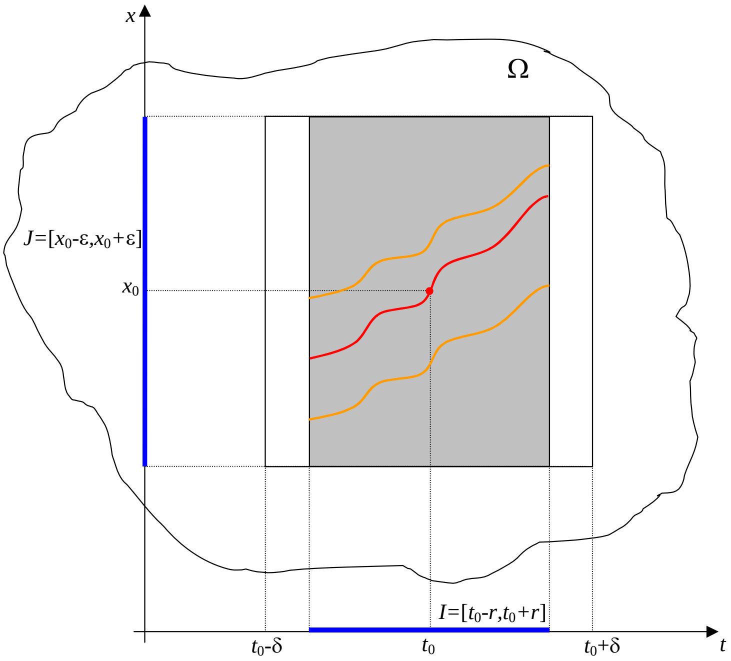
for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in \overline{B}(x_0, \varepsilon)$. Choose some $r \in (0, \delta]$ to be specified later on, and set

$$I = [t_0 - r, t_0 + r] \text{ and } J = \overline{B}(x_0, \varepsilon).$$

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

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

The following diagram illustrates the setting in the case $n = 1$:



Consider the integral operator A defined on functions $x\left(t\right)$ by

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

We would like to ensure that $x \in X$ implies $Ax \in X$. Note that, for any $x \in X$, the point $\left(s, x\left(s\right)\right)$ 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

$$\left\|Ax\left(t\right) - x_0\right\| \leq \varepsilon \text{ for all } t \in I. \tag{3.7}$$

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

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

where

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

Hence, if r is so small that $Mr \leq \varepsilon$ then (3.7) 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 $x \in X$ solves the IVP if $x = Ax$, that is, if

x is a fixed point of the mapping A. The existence of a fixed point of A will be proved using the Banach fixed point theorem, but in order to be able to apply this theorem, we must introduce a distance function[55] d on X so that (X, d) is a complete metric space and A is a contraction mapping with respect to this distance.

Define the distance function on the function family X as follows: for all $x, y \in X$, set

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

Then (X, d) is known to be a complete metric space. 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

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

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

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

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 $I = [t_0 - r, t_0 + r]$ where $r > 0$ is to be defined. Let ε and δ be the parameters from the Lipschitz condition at the point (t_0, x_0) as above. Choose $r > 0$ to satisfy all the above restrictions $r \leq \delta, r \leq \varepsilon/M, r < \frac{1}{L}$, and in addition to be 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 $x(t_0) = y(t_0) = x_0$ and both $x(t)$ and $y(t)$ are continuous functions). Then the both functions x and y belong to the space X and are fixed points of the mapping A. By the uniqueness part of the Banach fixed point theorem, we conclude that $x = y$ as elements of X, that is, $x(t) = y(t)$ for all $t \in I$.

[55] Abstand

Alternatively, the same can be proved by using the Gronwall inequality (Lemma 2.2). Indeed, the both solutions satisfy 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

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

assuming for certainty that $t_0 \le t \le t_0 + r$. 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

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

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 $x(t)$ 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.

Now we use the following fact from Analysis: 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 $S = U$. Since the set S is non-empty (it contains t_0) and is both open and closed in U, we conclude that $S = U$, which finishes the proof of uniqueness.

For the sake of completeness, let us give also the proof of the fact that S is either empty or $S = U$. 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 $b_2 \in S^c$. Continuing further on, 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[56], 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. ■

Example. The method of the proof of the existence in Theorem 3.2 suggests the following procedure of computation of the solution of IVP. We start with any function $x_0 \in X$ (using

[56] Intervallschachtelungsprinzip

the same notation as in the proof) and construct the sequence $\{x_n\}_{n=0}^{\infty}$ of functions in X using the rule $x_{n+1} = Ax_n$. The sequence $\{x_n\}$ is called the *Picard iterations,* and it converges uniformly to the solution $x(t)$.

Let us illustrate this method on the following example:

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

The operator A is given by

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

whence, setting $x_0(t) \equiv 1$, we obtain

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

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

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

and by induction

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

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

Remark. Let us summarize the proof of the existence part of Theorem 3.2 as follows. For any point $(t_0, x_0) \in \Omega$, we first choose positive constants ε, δ, L from the Lipschitz condition, so that 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 value of t,

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

Let

$$M = \sup_{(t,x)\in G} \|f(t, x)\|$$

and choose any positive r to satisfy

$$r \leq \delta,\ r \leq \frac{\varepsilon}{M},\ r < \frac{1}{L}. \tag{3.8}$$

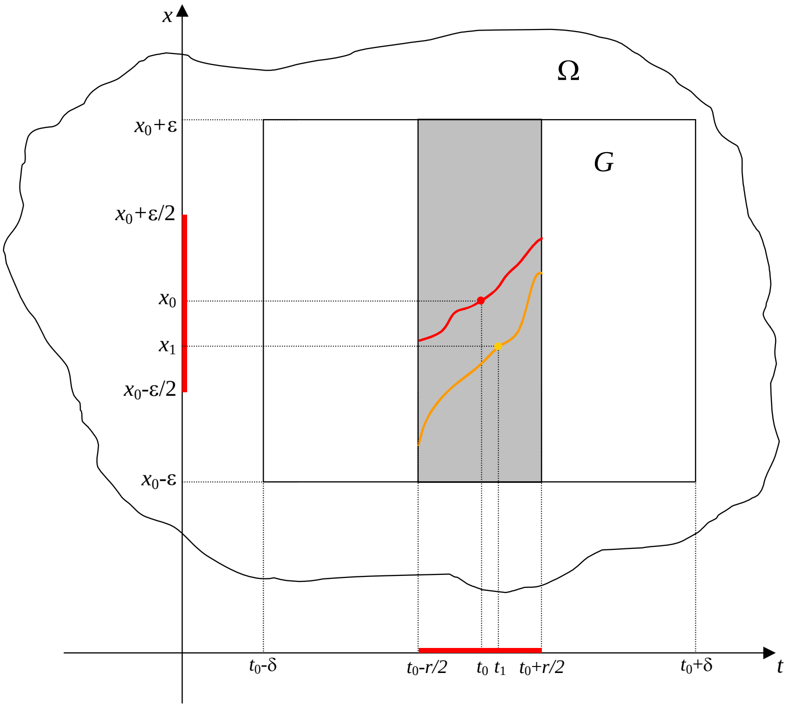
Then there exists a solution $x(t)$ to the IVP, which is defined on the interval $[t_0 - r, t_0 + r]$ and takes values in $\overline{B}(x_0, \varepsilon)$.

The fact that the domain of the solution admits the explicit estimates (3.8) can be used as follows.

Corollary. *Under the conditions of Theorem* 3.2 *for any point* $(t_0, x_0) \in \Omega$ *there are positive constants* ε *and* r *such that, for any* $t_1 \in [t_0 - r/2, t_0 + r/2]$ *and* $x_1 \in \overline{B}(x_0, \varepsilon/2)$*, the IVP*

$$\begin{cases} x' = f(t, x), \\ x(t_1) = x_1, \end{cases} \tag{3.9}$$

has a solution $x(t)$ *which is defined for all* $t \in [t_0 - r/2, t_0 + r/2]$ *and takes values in* $\overline{B}(x_0, \varepsilon)$ *(see the diagram below for the case* $n = 1$*).*



The essence of this statement is that despite the variation of the initial conditions (t_1, x_1), the solution of (3.9) is defined on the same interval $[t_0 - r/2, t_0 + r/2]$ and takes values in the same ball $\overline{B}(x_0, \varepsilon)$ provided (t_1, x_1) is close enough to (t_0, x_0).

Proof. Let $\varepsilon, \delta, L, M$ be as in the proof of Theorem 3.2 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. Therefore, choosing $r > 0$ to satisfy the conditions

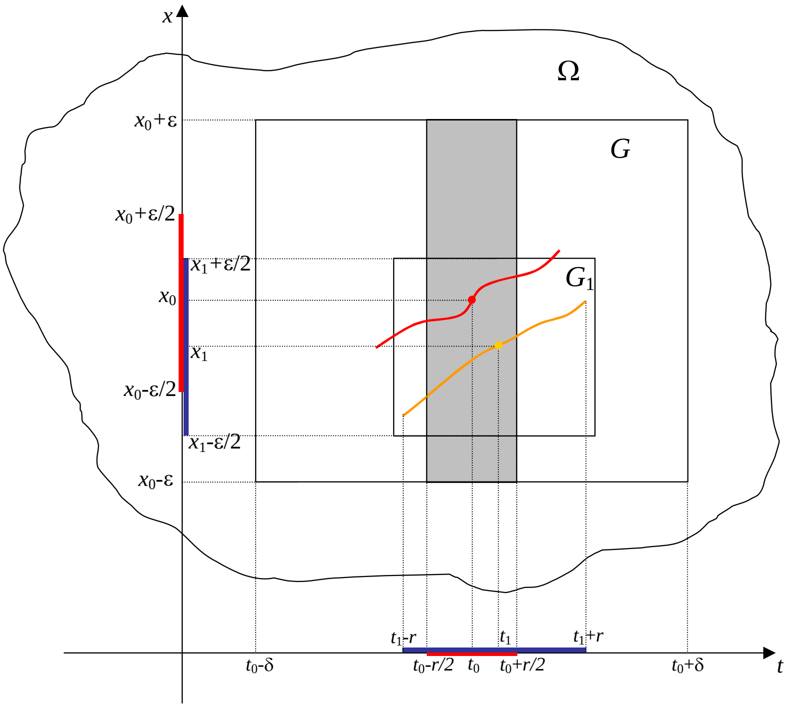
$$r \leq \delta/2, \ r \leq \frac{\varepsilon}{2M}, \ r < \frac{1}{L},$$

for example,

$$r = \min\left(\frac{\delta}{2}, \frac{\varepsilon}{2M}, \frac{1}{2L}\right),$$

we obtain that the IVP (3.9) has a solution $x(t)$ that is defined in the interval $[t_1 - r, t_1 + r]$ and takes values in $\overline{B}(x_1, \varepsilon/2) \subset \overline{B}(x, \varepsilon)$. If $t_1 \in [t_0 - r/2, t_0 + r/2]$ then $[t_0 - r/2, t_0 + r/2] \subset$

$[t_1 - r, t_1 + r]$, and we see that the solution $x(t)$ of (3.9) is defined on $[t_0 - r/2, t_0 + r/2]$ and takes value in $\overline{B}(x, \varepsilon)$, which was to be proved (see the diagram below).



■

3.2 Maximal solutions

Consider again the ODE

$$x' = f(t, x) \tag{3.10}$$

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 3.2 says that any two solutions of the same initial value problem

$$\begin{cases} x' = f(t, x), \\ x(t_0) = x_0 \end{cases} \tag{3.11}$$

coincide 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. The purpose of what follows is to determine the maximal possible domain of the solution to (3.11).

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 3.3 *Under the conditions of Theorem* 3.2*, the following is true.*

(a) The initial value problem (3.11) *has is a unique maximal solution for any initial values* $(t_0, x_0) \in \Omega$.

(b) If $x(t)$ *and* $y(t)$ *are two maximal solutions of* (3.10) *and* $x(t) = y(t)$ *for some value of* t*, then the solutions* x *and* y *are identical, 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.

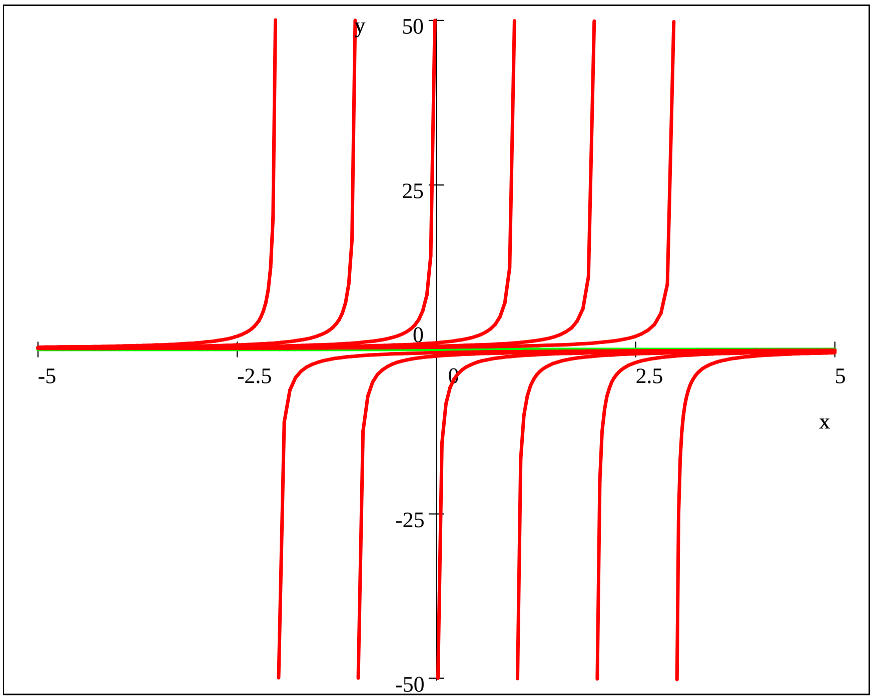
Example. 1. Consider the ODE $x' = x^2$ in the domain $\Omega = \mathbb{R}^2$. This is separable equation and can be solved as follows. Obviously, $x \equiv 0$ is a constant solution. In the domains where $x \neq 0$ we have

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

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 $x(t) = \frac{1}{C-t}$ with the maximal domains $(C, +\infty)$ and $(-\infty, C)$ (see the diagram below).



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 $x(t) = \frac{1}{C-t}$ leaves K as $t \to C$ because $x(t) \to \pm\infty$.

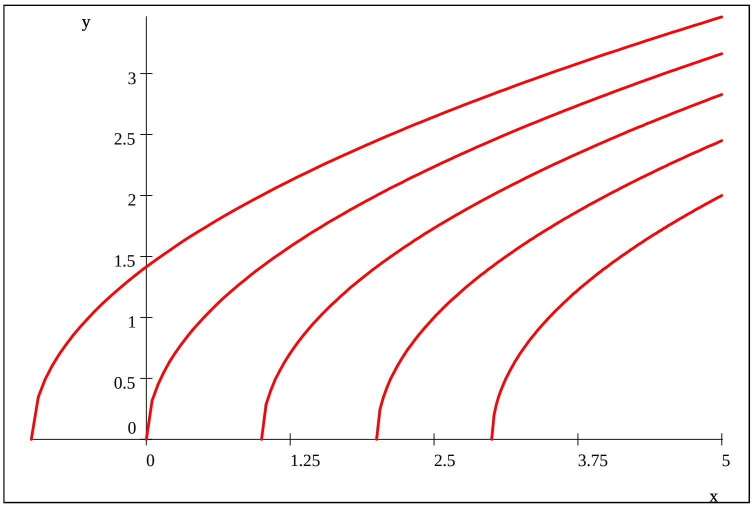
2. Consider the ODE $x' = \frac{1}{x}$ in the domain $\Omega = \mathbb{R} \times (0, +\infty)$ (that is, $t \in \mathbb{R}$ and $x > 0$). By the separation of variables, we obtain

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

whence

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

See the diagram below:



Obviously, the maximal domain of the solution is $(C, +\infty)$. The solution 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 3.3 will be preceded by a lemma.

Lemma 3.4 *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:*

$$x(t) = x_\alpha(t) \ \ \text{if}\ t \in I_\alpha. \tag{3.12}$$

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

The function $x(t)$ defined by (3.12) is referred to as the *union* of the family $\{x_\alpha(t)\}$.

Proof. First of all, let us verify that the identity (3.12) 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 3.3. (a) Let S be the set of all possible solutions to the IVP (3.11) that are defined on open intervals. Let $x(t)$ be the union of all solutions from S. By Lemma 3.4, the function $x(t)$ is also a solution to (3.11) 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 (3.11) and let $z(t)$ be the union of the solutions $x(t)$ and $y(t)$. By Lemma 3.4, $z(t)$ solves the same 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 of (3.10) 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 of (3.10) defined on (a, b) where $a < b$, 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.

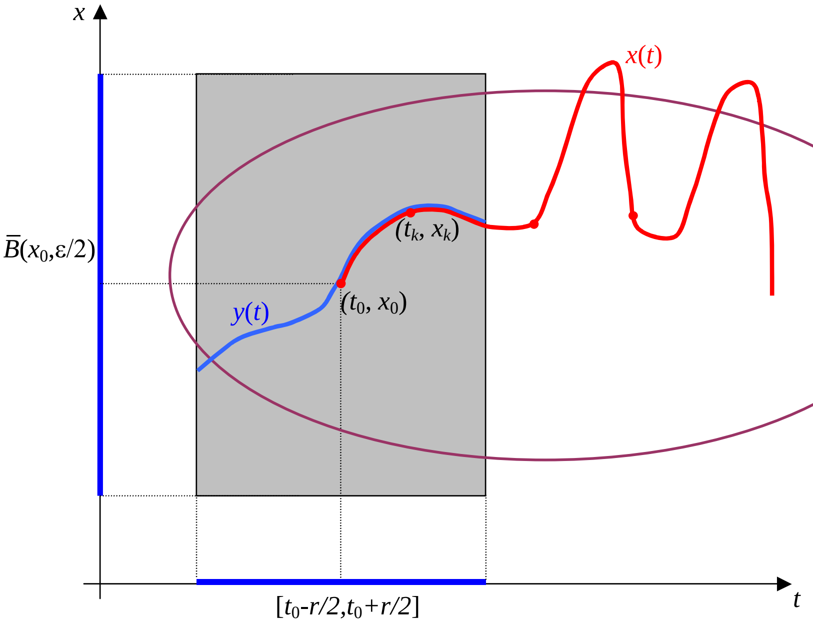
By Corollary to Theorem 3.2, for the point (t_0, x_0), there exist $r, \varepsilon > 0$ such that the IVP with the initial point inside the cylinder

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

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 3.3.

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

Let Ω be an open set in $\mathbb{R}^{n+1}$ and f, g be two mappings from Ω to $\mathbb{R}^n$. Assume in what follows that both f, g are continuous and locally Lipschitz in x in Ω, and consider two initial value problems

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} \tag{3.13}$$

and

$$\begin{cases} y' = g(t,y) \\ y(t_0) = x_0 \end{cases} \tag{3.14}$$

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

Let the function f be fixed and $x(t)$ be a fixed solution of (3.13). The function g will be treated as variable. Our purpose is to show that if g is chosen close enough to f then the solution $y(t)$ of (3.14) is close enough to $x(t)$. Apart from the theoretical interest, this result has significant practical consequences. For example, if one knows the function $f(t,x)$ only approximately (which is always the case in applications in Sciences and Engineering) then solving (3.13) approximately means solving another problem (3.14) where g is an approximation to f. Hence, it is important to know that the solution $y(t)$ of (3.14) is actually an approximation of $x(t)$.

Theorem 3.5 *Let $x(t)$ be a solution to the IVP* (3.13) *defined on a closed bounded interval $[\alpha, \beta]$ such that $\alpha < t_0 < \beta$. For any $\varepsilon > 0$, there exists $\eta > 0$ with the following property: for any function $g : \Omega \to \mathbb{R}^n$ as above and such that*

$$\sup_{\Omega} \|f - g\| \leq \eta, \tag{3.15}$$

there is a solution $y(t)$ of the IVP (3.14) *defined on $[\alpha, \beta]$, and this solution satisfies the inequality*

$$\sup_{[\alpha,\beta]} \|x(t) - y(t)\| \leq \varepsilon. \tag{3.16}$$

Proof. We start with the following estimate of the difference $\|x(t) - y(t)\|$.

Claim 1. *Let $x(t)$ and $y(t)$ be solutions of (3.13) and (3.14) respectively, and assume that they both are defined on some interval (a,b) containing t_0. Assume also that the* *graphs of $x(t)$ and $y(t)$ over (a,b) belong to a subset K of Ω such that $f(t,x)$ is Lipschitz in K in x with the Lipschitz constant L. Then*

$$\sup_{t\in(a,b)} \|x(t)-y(t)\| \le e^{L(b-a)}(b-a)\sup_K \|f-g\|. \tag{3.17}$$

Let us use the integral identities

$$x(t)=x_0+\int_{t_0}^t f(s,x(s))\,ds \quad \text{and} \quad y(t)=x_0+\int_{t_0}^t g(s,y(s))\,ds.$$

Assuming that $t_0\le t<b$ and using the triangle inequality, we obtain

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

Since the points $(s,x(s))$ and $(s,y(s))$ are in K, we obtain by the Lipschitz condition in K that

$$\|x(t)-y(t)\| \le L\int_{t_0}^t \|x(s)-y(s)\|\,ds+(b-a)\sup_K\|f-g\|. \tag{3.18}$$

Applying the Gronwall lemma to the function $z(t)=\|x(t)-y(t)\|$, we obtain

$$\begin{aligned}\|x(t)-y(t)\| &\le e^{L(t-t_0)}(b-a)\sup_K\|f-g\|\\ &\le e^{L(b-a)}(b-a)\sup_K\|f-g\|. \end{aligned}\tag{3.19}$$

In the same way, (3.19) holds for $a<t\le t_0$ so that it is true for all $t\in(a,b)$, whence (3.17) follows.

The estimate (3.17) implies that if $\sup_\Omega \|f-g\|$ is small enough then also the difference $\|x(t)-y(t)\|$ is small, which is essentially what is claimed in Theorem 3.5. However, in order to be able to use the estimate (3.17) in the setting of Theorem 3.5, we have to deal with the following two issues:

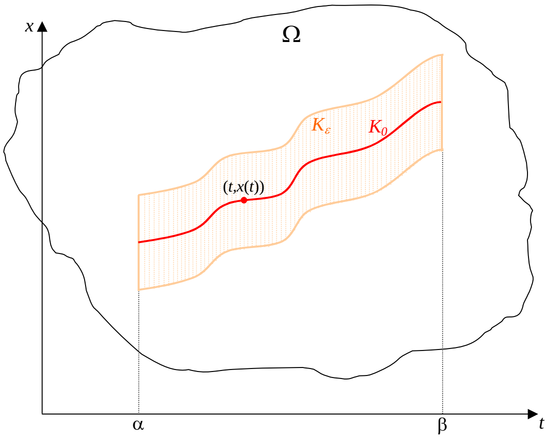
– to show that the solution $y(t)$ is defined on the interval $[\alpha,\beta]$;

– to show that the graphs of $x(t)$ and $y(t)$ are contained in a set K where f satisfies the Lipschitz condition in x.

We proceed with the construction of such a set. For any $\varepsilon\ge 0$, consider the set

$$K_\varepsilon=\left\{(t,x)\in\mathbb{R}^{n+1}:\alpha\le t\le\beta,\ \ \|x-x(t)\|\le\varepsilon\right\} \tag{3.20}$$

which can be regarded as the closed 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 the function $x(t)$ on $[\alpha,\beta]$ (see the diagram below).



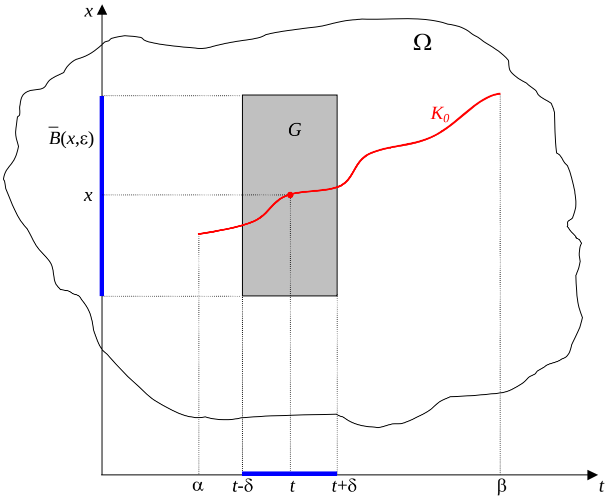
The set K_0 is compact because it is the image of the compact interval $[\alpha, \beta]$ under the continuous mapping $t \mapsto (t, x(t))$. Hence, K_0 is bounded and closed, which implies that also K_ε for any $\varepsilon > 0$ is bounded and closed. Thus, K_ε is a compact subset of $\mathbb{R}^{n+1}$ for any $\varepsilon \geq 0$.

Claim 2. *There is $\varepsilon > 0$ such that $K_\varepsilon \subset \Omega$ and f is Lipschitz in K_ε in x.*

Indeed, by the local Lipschitz condition, for any point $(t, x) \in \Omega$ (in particular, for any $(t, x) \in K_0$), there are constants $\varepsilon, \delta > 0$ such that the cylinder

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

is contained in Ω and f is Lipschitz in G (see the diagram below).



Consider also an open cylinder

$$H = (t - \delta, t + \delta) \times B\left(x, \tfrac{1}{2}\varepsilon\right).$$

Varying the point (t, x) in K_0, we obtain a cover of K_0 by such cylinders. Since K_0 is compact, there is a finite subcover, that is, a finite number of points $\{(t_i, x_i)\}_{i=1}^m$ on K_0 and numbers $\varepsilon_i, \delta_i > 0$, such that the cylinders

$$H_i = (t_i - \delta_i, t_i + \delta_i) \times B\left(x_i, \tfrac{1}{2}\varepsilon_i\right)$$

cover all K_0. Set

$$G_i = [t_i - \delta_i, t_i + \delta_i] \times \overline{B}\,(x_i, \varepsilon_i)$$

and let L_i be the Lipschitz constant of f in G_i, which exists by the choice of ε_i, δ_i.

Define ε and L as follows

$$\varepsilon = \frac{1}{2} \min_{1 \le i \le m} \varepsilon_i, \qquad L = \max_{1 \le i \le m} L_i, \tag{3.21}$$

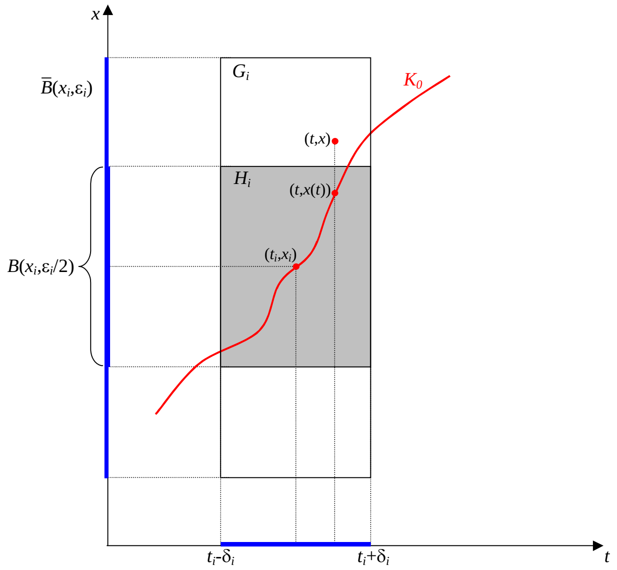
and prove that $K_\varepsilon \subset \Omega$ and that the function f is Lipschitz in K_ε with the constant L. For any point $(t, x) \in K_\varepsilon$, we have by the definition of K_ε that $t \in [\alpha, \beta]$, $(t, x\,(t)) \in K_0$ and

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

Then the point $(t, x\,(t))$ belongs to one of the cylinders H_i so that

$$t \in (t_i - \delta_i, t_i + \delta_i) \quad \text{and} \quad \|x\,(t) - x_i\| < \tfrac{1}{2}\varepsilon_i$$

(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 by (3.21) $\varepsilon \le \varepsilon_i/2$. Therefore, $x \in B\,(x_i, \varepsilon_i)$ whence it follows that $(t, x) \in G_i$ and $(t, x) \in \Omega$. Hence, we have shown that any point from K_ε belongs to Ω, which proves that $K_\varepsilon \subset \Omega$.

If $(t, x), (t, , y) \in K_\varepsilon$ then by the above argument the both points x, y belong to the same ball $B\,(x_i, \varepsilon_i)$ that is determined by the condition $(t, x\,(t)) \in H_i$. Then $(t, x), (t, , y) \in G_i$ and, since f is Lipschitz in G_i with the constant L_i, we obtain

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

where we have used the definition (3.21) of L. This shows that f is Lipschitz in x in K_ε and finishes the proof of Claim 2.

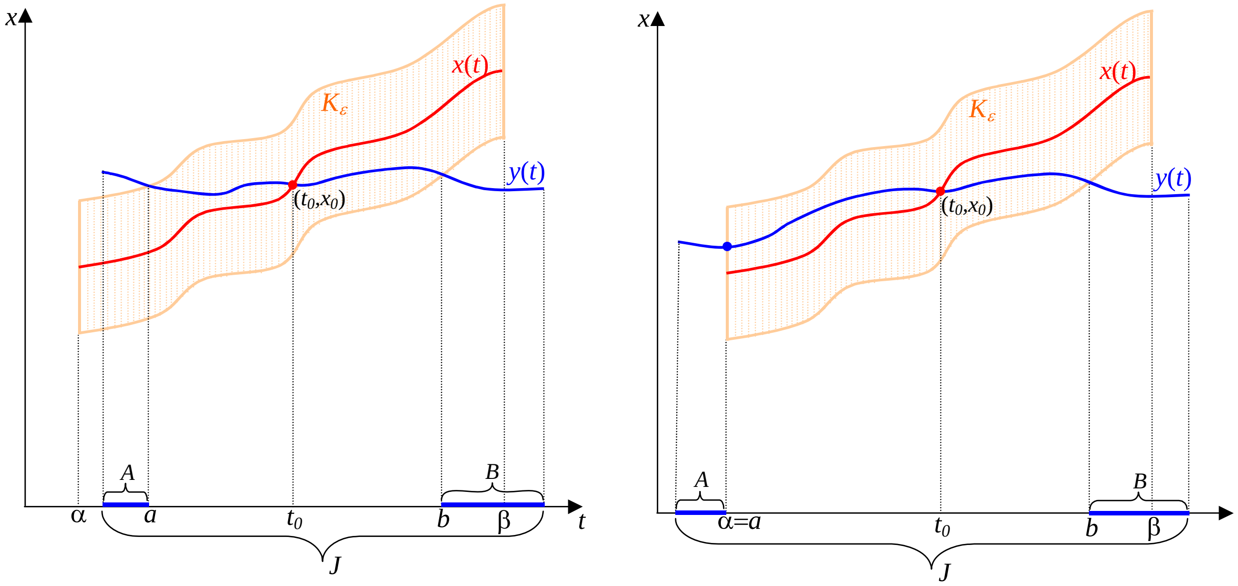
Let now $y(t)$ be the maximal solution to the IVP (3.14), and let J be its domain; in particular, J is an open interval containing t_0. We are going to spot an interval (a,b) where both functions $x(t)$ and $y(t)$ are defined and their graphs over (a,b) belong to K_ε. Since $y(t_0) = x_0$, the point $(t_0, y(t_0))$ of the graph of $y(t)$ is contained in K_ε. By Theorem 3.3, the graph of $y(t)$ leaves K_ε when t goes to the endpoints of J. It follows that the following sets are non-empty:

$$\begin{aligned} A &: = \{t \in J, t < t_0 : (t, y(t)) \notin K_\varepsilon\}, \\ B &: = \{t \in J, t > t_0 : (t, y(t)) \notin K_\varepsilon\}. \end{aligned}$$

On the other hand, in a small neighborhood of t_0, the graph of $y(t)$ is close to the point (t_0, x_0) and, hence, is in K_ε. Therefore, the sets A and B are separated out from t_0, the set A being on the left from t_0, and the set B being on the right from t_0. Set

$$a = \sup A \quad \text{and} \quad b = \inf B$$

(see the diagram below for two cases: $a > \alpha$ and $a = \alpha$).



Claim 3. *We have $a < t_0 < b$ and the interval $[a, b]$ is contained both in J and $[\alpha, \beta]$.*

Indeed, as was remarked above, the set A lies on the left from t_0 and is separated from t_0, whence $a < t_0$; similarly $b > t_0$. Since A is contained in J but is separated from the right endpoint of J, it follows that $a \in J$; similarly, $b \in J$. Any point $t \in (a, b)$ belongs to J but does not belong to A or B, whence $(t, y(t)) \in K_\varepsilon$; therefore, $t \in [\alpha, \beta]$, which implies that $a, b \in [\alpha, \beta]$.

Now we can finish the proof of Theorem 3.5 as follows. Assuming that

$$\|f - g\|_{K_\varepsilon} \leq \eta$$

and using $b - a \leq \beta - \alpha$, we obtain from (3.17) with $K = K_\varepsilon$ that

$$\sup_{t\in(a,b)} \|x(t) - y(t)\| \leq e^{L(\beta-\alpha)} (\beta - \alpha) \eta \leq \frac{\varepsilon}{2}, \tag{3.22}$$

provided we choose $\eta = \frac{\varepsilon}{2(\beta-\alpha)} e^{-L(\beta-\alpha)}$. It follows then by letting $t \to a+$ that

$$\|x(a) - y(a)\| \leq \varepsilon/2. \tag{3.23}$$

Let us deduce from (3.23) that $a = \alpha$. Indeed, for any $t \in A$, we have $(t, y(t)) \notin K_\varepsilon$, which can occur in two ways:

1. either $t < \alpha$ (exit through the left boundary of K_ε)
2. or $t \geq \alpha$ and $\|x(t) - y(t)\| > \varepsilon$ (exit through the lateral boundary of K_ε)

If $a > \alpha$ then there is a point $t \in A$ such that $t > \alpha$. At this point we have $\|x(t) - y(t)\| > \varepsilon$ whence it follows by letting $t \to a-$ that

$$\|x(a) - y(a)\| \geq \varepsilon,$$

which contradicts (3.23). Hence, we conclude that $a = \alpha$ and in the same way $b = \beta$. Since $[a, b] \subset J$, it follows that the solution $y(t)$ is defined on the interval $[\alpha, \beta]$. Finally, (3.16) follows from (3.22). ■

Using the proof of Theorem 3.5, we can refine the statement of Theorem 3.5 as follows.

Corollary. *Under the hypotheses of Theorem 3.5, let $\varepsilon > 0$ be sufficiently small so that $K_\varepsilon \subset \Omega$ and $f(t, x)$ is Lipschitz in x in K_ε with the Lipschitz constant L (where K_ε is defined by (3.20)). If $\sup_{K_\varepsilon} \|f - g\|$ is sufficiently small, then the IVP (3.14) has a solution $y(t)$ defined on $[\alpha, \beta]$, and the following estimate holds*

$$\sup_{[\alpha,\beta]} \|x(t) - y(t)\| \leq e^{L(\beta-\alpha)} (\beta - \alpha) \sup_{K_\varepsilon} \|f - g\|. \tag{3.24}$$

Proof. Indeed, as it was shown in the proof of Theorem 3.5, if $\sup_{K_\varepsilon} \|f - g\|$ is small enough then $\alpha = b$ and $\beta = b$. Hence, (3.24) follows from (3.17). ■

3.4 Continuity of solutions with respect to a parameter

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

$$\begin{cases} x' = f(t, x, s) \\ x(t_0) = x_0 \end{cases} \tag{3.25}$$

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:

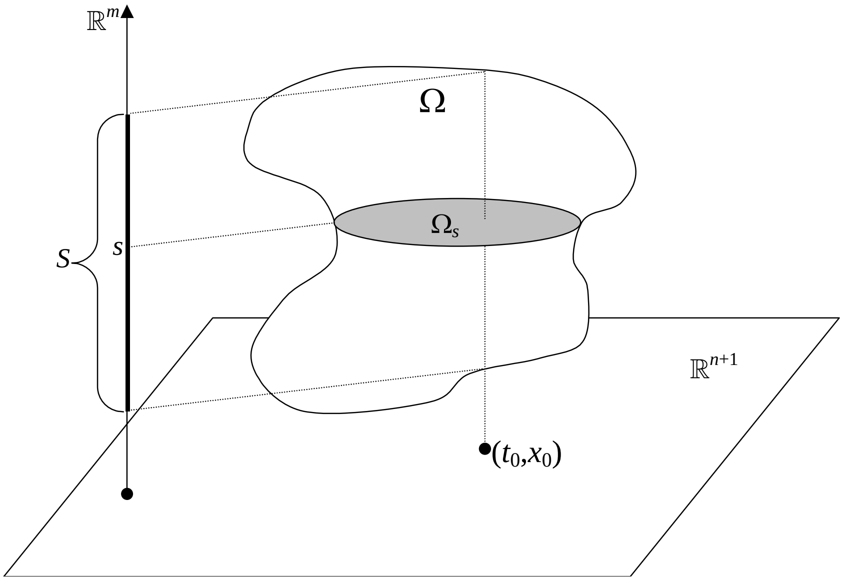
$$(t, x, s) = (t, x_1, .., x_n, s_1, ..., s_m).$$

How do we understand (3.25)? 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,

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



Then the IVP (3.25) 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 (3.25) 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

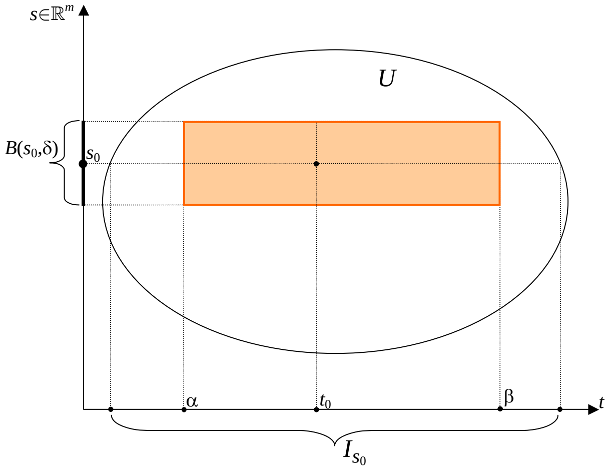
$$U = \left\{(t, s) \in \mathbb{R}^{m+1} : s \in S, t \in I_s\right\}.$$

Theorem 3.6 *Under the above assumptions, the set U is an open subset of $\mathbb{R}^{m+1}$ and the function $x(t, s) : U \to \mathbb{R}^n$ is continuous in (t, s).*

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 $\delta > 0$ such that

$$[\alpha, \beta] \times \overline{B}(s_0, \delta) \subset U, \tag{3.26}$$

which will imply that U is open. Here $\overline{B}(s_0, \delta)$ is a closed 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).



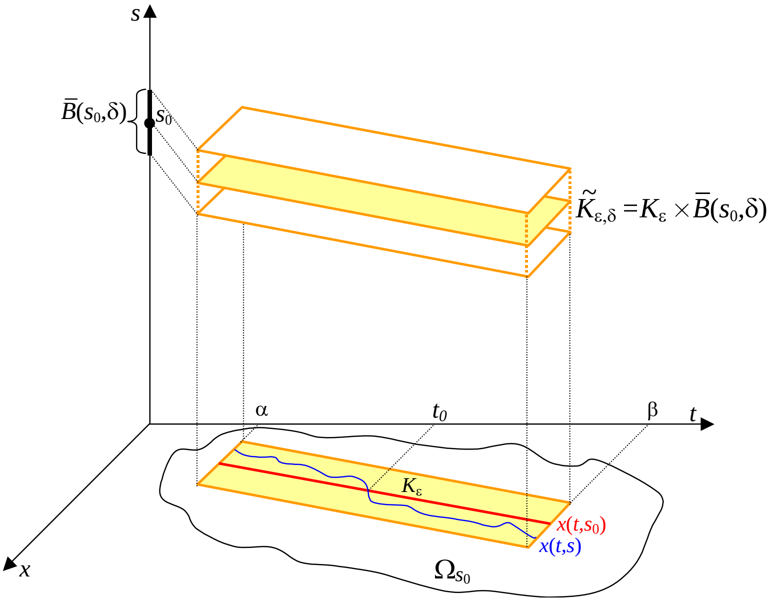
As in the proof of Theorem 3.5, consider a set

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

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

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

(see the diagram below).



s
$\overline{B}(s_0,\delta)$
s_0
$\widetilde{K}_{\varepsilon,\delta} = K_\varepsilon \times \overline{B}(s_0,\delta)$
α
t_0
β
t
K_ε
$x(t,s_0)$
$x(t,s)$
Ω_{s_0}
x

If ε and δ are small enough then $\widetilde{K}_{\varepsilon,\delta}$ is contained in Ω (cf. the proof of Theorem 3.5). Hence, for any $s \in \overline{B}(s_0, \delta)$, the function $f(t,x,s)$ is defined for all $(t,x) \in K_\varepsilon$. Since the function f is continuous on Ω, it is uniformly continuous on the compact set $\widetilde{K}_{\varepsilon,\delta}$, 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 Corollary to Theorem 3.5 with[57] $f(t,x) = f(t,x,s_0)$ and $g(t,x) = f(t,x,s)$ where $s \in \overline{B}(s_0, \delta)$, we obtain that if

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

is small enough then the solution $y(t) = x(t,s)$ is defined on $[\alpha, \beta]$. In particular, this implies (3.26) for small enough δ.

Furthermore, applying the estimate (3.24) of Corollary to Theorem 3.5, we 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 $C = e^{L(\beta-\alpha)}(\beta - \alpha)$ depends only on α, β and the Lipschitz constant L of the function $f(t,x,s_0)$ in K_ε. Letting $s \to s_0$, we obtain that

$$\sup_{t\in[\alpha,\beta]} \|x(t,s) - x(t,s_0)\| \to 0 \text{ as } s \to s_0, \tag{3.27}$$

so that $x(t,s)$ is continuous in s at $s = s_0$ uniformly in $t \in [\alpha, \beta]$. Since $x(t,s)$ is continuous in t for any fixed s, it follows that x is continuous in (t,s). Indeed, fix a point $(t_0, s_0) \in U$ (where $t_0 \in (\alpha, \beta)$ is not necessarily the value from the initial condition) and prove that $x(t,s) \to x(t_0,s_0)$ as $(t,s) \to (t_0,s_0)$. Using (3.27) and the continuity of the function $x(t,s_0)$ in t, we obtain

$$\begin{aligned} \|x(t,s) - x(t_0,s_0)\| &\le \|x(t,s) - x(t,s_0)\| + \|x(t,s_0) - x(t_0,s_0)\| \\ &\le \sup_{t\in[\alpha,\beta]} \|x(t,s) - x(t,s_0)\| + \|x(t,s_0) - x(t_0,s_0)\| \\ &\to 0 \text{ as } s \to s_0 \text{ and } t \to t_0, \end{aligned}$$

which finishes the proof. ■

As an example of application of Theorem 3.6, consider the dependence of a solution on the initial condition.

Corollary. *Consider the initial value problem*

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

where $f : \Omega \to \mathbb{R}^n$ is a continuous function in an open set $\Omega \subset \mathbb{R}^{n+1}$ that is Lipschitz in x in Ω. For any point $(t_0, x_0) \in \Omega$, denote by $x(t, t_0, x_0)$ the maximal solution of this problem. Then the function $x(t, t_0, x_0)$ is continuous jointly in (t, t_0, x_0).

[57] Since the common domain of the functions $f(t,x,s)$ and $f(t,x,s_0)$ is $(t,x) \in \Omega_{s_0} \cap \Omega_s$, Theorem 3.5 should be applied with this domain.

Proof. Consider a new function $y(t) = x(t+t_0) - x_0$ that satisfies the ODE

$$y'(t) = x'(t+t_0) = f(t+t_0, x(t+t_0)) = f(t+t_0, y(t) + x_0).$$

Considering $s := (t_0, x_0)$ as a parameter and setting

$$F(t, y, s) = f(t+t_0, y+x_0),$$

we see that y satisfies the IVP

$$\begin{cases} y' = F(t, y, s) \\ y(0) = 0. \end{cases}$$

The domain of function F consists of those points $(t, y, t_0, x_0) \in \mathbb{R}^{2n+2}$ for which

$$(t+t_0, y+x_0) \in \Omega,$$

which obviously is an open subset of $\mathbb{R}^{2n+2}$. Since the function $F(t, y, s)$ is clearly continuous in (t, y, s) and is locally Lipschitz in y, we conclude by Theorem 3.6 that the maximal solution $y = y(t, s)$ is continuous in (t, s). Consequently, the function $x(t, t_0, x_0) = y(t - t_0, t_0, x_0) + x_0$ is continuous in (t, t_0, x_0). ■

3.5 Differentiability of solutions in parameter

Consider again the initial value problem with parameter

$$\begin{cases} x' = f(t, x, s), \\ x(t_0) = x_0, \end{cases} \tag{3.28}$$

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 $(t, x, s) = (t, x_1, ..., x_n, s_1, ..., s_m)$. Let us use the following notation for the partial Jacobian matrices of f with respect to the variables x and s:

$$f_x = \partial_x f = \frac{\partial f}{\partial x} := \left(\frac{\partial f_i}{\partial x_k}\right),$$

where $i = 1, ..., n$ is the row index and $k = 1, ..., n$ is the column index, so that f_x is an $n \times n$ matrix, and

$$f_s = \partial_s f = \frac{\partial f}{\partial s} := \left(\frac{\partial f_i}{\partial s_l}\right),$$

where $i = 1, ..., n$ is the row index and $l = 1, ..., m$ is the column index, so that f_s is an $n \times m$ matrix.

Note that if f_x is continuous in Ω then, by Lemma 3.1, f is locally Lipschitz in x so that all the previous results apply. Let $x(t, s)$ be the maximal solution to (3.28). Recall that, by Theorem 3.6, 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 3.7 *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*

$$\begin{cases} y' = f_x(t, x(t, s), s)\, y + f_s(t, x(t, s), s), \\ y(t_0) = 0. \end{cases} \tag{3.29}$$

Here $\partial_s x = \left(\frac{\partial x_k}{\partial s_l}\right)$ is an $n \times m$ matrix where $k = 1, .., n$ is the row index and $l = 1, ..., m$ is the column index. Hence, $y = \partial_s x$ can be considered as a vector in $\mathbb{R}^{n \times m}$ depending on t and s. The both terms in the right hand side of (3.29) are also $n \times m$ matrices so that (3.29) makes sense. Indeed, f_s is an $n \times m$ matrix, and $f_x y$ is the product of the $n \times n$ and $n \times m$ matrices, which is again an $n \times m$ matrix. The notation $f_x\left(t, x\left(t, s\right), s\right)$ means that one first evaluates the partial derivative $f_x\left(t, x, s\right)$ and then substitute the values of s and $x = x\left(t, s\right)$ (and the same remark applies to $f_s\left(t, x\left(t, s\right), s\right)$).

The ODE in (3.29) is called the *variational equation* for (3.28) along the solution $x\left(t, s\right)$ (or the *equation in variations*).

$\boxed{\text{Lecture 19}}$ 22.06.2009 Prof. A. Grigorian, ODE, SS 2009

Note that the variational equation is linear. Indeed, for any fixed s, it can be written in the form

$$y' = a\left(t\right) y + b\left(t\right),$$

where

$$a\left(t\right) = f_x\left(t, x\left(t,s\right), s\right), \qquad b\left(t\right) = f_s\left(t, x\left(t,s\right), s\right).$$

Since f is continuous and $x\left(t,s\right)$ is continuous by Theorem 3.6, the functions $a\left(t\right)$ and $b\left(t\right)$ are continuous in t. If the domain in t of the solution $x\left(t,s\right)$ is I_s then the domain of the variational equation is $I_s \times \mathbb{R}^{n\times m}$. By Theorem 2.1, the solution $y\left(t\right)$ of (3.29) exists in the full interval I_s. Hence, Theorem 3.7 can be stated as follows: if $x\left(t,s\right)$ is the solution of (3.28) on I_s and $y\left(t\right)$ is the solution of (3.29) on I_s then we have the identity $y\left(t\right) = \partial_s x\left(t,s\right)$ for all $t \in I_s$. This provides a method of evaluating $\partial_s x\left(t,s\right)$ for a fixed s without finding $x\left(t,s\right)$ for all s.

Example. Consider the IVP with parameter

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

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

For $s = 0$ we have the IVP

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

whence we obtain $x\left(t,0\right) = -\frac{1}{t}$. Noticing that $f_x = 2x$ and $f_s = 2/t$ we obtain the variational equation along this solution

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

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

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

where $A\left(t\right)$ is a primitive of $a\left(t\right) = -2/t$, that is $A\left(t\right) = -2\ln t$. Hence,

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

The initial condition $y\left(1\right) = 0$ is satisfied for $C = -1$ so that $y\left(t\right) = 1 - t^{-2}$. By Theorem 3.7, we conclude that $\partial_s x\left(t,0\right) = 1 - t^{-2}$.

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

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

that is,

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

In particular, we obtain for small s an approximation

$$x(t,s) \approx -\frac{1}{t} + \left(1 - \frac{1}{t^2}\right) s.$$

Later we will be able to obtain more terms in the Taylor formula and, hence, to get a better approximation for $x(t,s)$.

Let us discuss the variational equation (3.29). It is easy to deduce (3.29) provided it is known that the mixed partial derivatives $\partial_s \partial_t x$ and $\partial_t \partial_s x$ exist and are equal (for example, this is the case when $x(t,s) \in C^2(U)$). Then differentiating (3.28) in s and using the chain rule, we obtain

$$\partial_t \partial_s x = \partial_s (\partial_t x) = \partial_s [f(t, x(t,s), s)] = f_x(t, x(t,s), s) \partial_s x + f_s(t, x(t,s), s),$$

which implies (3.29) after substitution $\partial_s x = y$. Although this argument is not a proof[58] of Theorem 3.7, it allows to memorize the variational equation.

Yet another point of view on the equation (3.29) is as follows. Fix a value of $s = s_0$ and set $x(t) = x(t, s_0)$. Using the differentiability of $f(t,x,s)$ in (x,s), we can write, for any fixed t,

$$f(t,x,s) = f(t, x(t), s_0) + f_x(t, x(t), s_0)(x - x(t)) + f_s(t, x(t), s_0)(s - s_0) + R,$$

where $R = o(\|x - x(t)\| + \|s - s_0\|)$ as $\|x - x(t)\| + \|s - s_0\| \to 0$. If s is close to s_0 then $x(t,s)$ is close to $x(t)$, and we obtain the approximation

$$f(t, x(t,s), s) \approx f(t, x(t), s_0) + a(t)(x(t,s) - x(t)) + b(t)(s - s_0),$$

whence

$$x'(t,s) \approx f(t, x(t), s_0) + a(t)(x(t,s) - x(t)) + b(t)(s - s_0).$$

This equation is linear with respect to $x(t,s)$ and is called the *linearized equation* at the solution $x(t)$. Substituting $f(t, x(t), s_0)$ by $x'(t)$ and dividing by $s - s_0$, we obtain that the function $z(t,s) = \frac{x(t,s) - x(t)}{s - s_0}$ satisfies the approximate equation

$$z' \approx a(t) z + b(t).$$

It is not difficult to believe that the derivative $y(t) = \partial_s x|_{s=s_0} = \lim_{s \to s_0} z(t,s)$ satisfies this equation exactly. This argument (although not rigorous) shows that the variational equation for y originates from the linearized equation for x.

How can one evaluate 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 (3.29), which we write in the form

$$\begin{cases} y' = g(t, y, s) \\ y(t_0) = 0 \end{cases} \tag{3.30}$$

where

$$g(t, y, s) = f_x(t, x(t,s), s) y + f_s(t, x(t,s), s). \tag{3.31}$$

[58] In the actual proof of Theorem 3.7, the variational equation is obtained *before* the identity $\partial_t \partial_s x = \partial_s \partial_t x$. The main technical difficulty in the proof of Theorem 3.7 lies in verifying the differentiability of x in s.

For what follows we use the notation $F(a,b,c,...) \in C^k(a,b,c,...)$ if all the partial derivatives of the order up to k of the function F with respect to the specified variables $a,b,c...$ exist and are continuous functions, in the domain of F. For example, the condition in Theorem 3.7 that f_x and f_s are continuous, can be shortly written as $f \in C^1(x,s)$, and the claim of Theorem 3.7 is that $x(t,s) \in C^1(t,s)$.

Assume now that $f \in C^2(x,s)$. Then by (3.31) we obtain that g is continuous and $g \in C^1(y,s)$, whence by Theorem 3.7 $y \in C^1(s)$. In particular, the function $z = \partial_s y = \partial_{ss} x$ is defined. Applying the variational equation to the problem (3.30), we obtain the equation for z

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

Since $g_y = f_x(t,x,s)$,

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

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

$$\begin{cases} z' = f_x(t,x,s)\, z + f_{xx}(t,x,s)\, y^2 + 2 f_{xs}(t,x,s)\, y + f_{ss}(t,x,s) \\ z'(t_0) = 0. \end{cases} \tag{3.32}$$

Note that here x must be substituted by $x(t,s)$ and y – by $y(t,s)$.

The equation (3.32) is called the *variational equation of the second order* (or *the second variational equation*). It is a linear ODE 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 can find 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 already computed that

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

Let us now evaluate $z(t) = \partial_{ss} x(t,0)$. Since

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

we obtain the second variational equation

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

Solving this equation 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

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

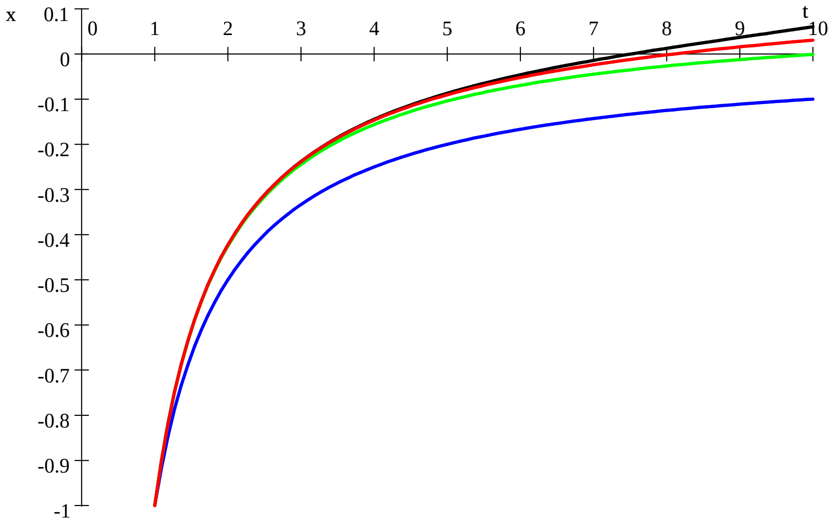
The initial condition $z(1) = 0$ yields $C = \frac{16}{3}$ whence

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

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

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

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

$$x(t,s) = x(t) + y(t)s + \frac{1}{2}z(t)s^2 + o(s^2). \tag{3.33}$$

Let us write a similar expansion for $x' = \partial_t x$, assuming that the derivatives ∂_t and ∂_s commute on x. We have

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

and in the same way

$$\partial_{ss}x' = \partial_s y' = \partial_t \partial_s y = z'.$$

Hence,

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

Substituting this into the equation

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

we obtain

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

whence

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

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

$$-1 = x(1) + sy(1) + \frac{s^2}{2}z(1) + o(s^2),$$

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.

Before we prove Theorem 3.7, let us prove some auxiliary statements 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 the case $z = 0$. If $x, y \in B(0,r)$ then $\|x\| < r$ and $\|y\| < r$ whence for any $\lambda \in [0,1]$

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

It follows that $(1-\lambda)x + \lambda y \in B(0,r)$, which was to be proved.

Lemma 3.8 (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

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

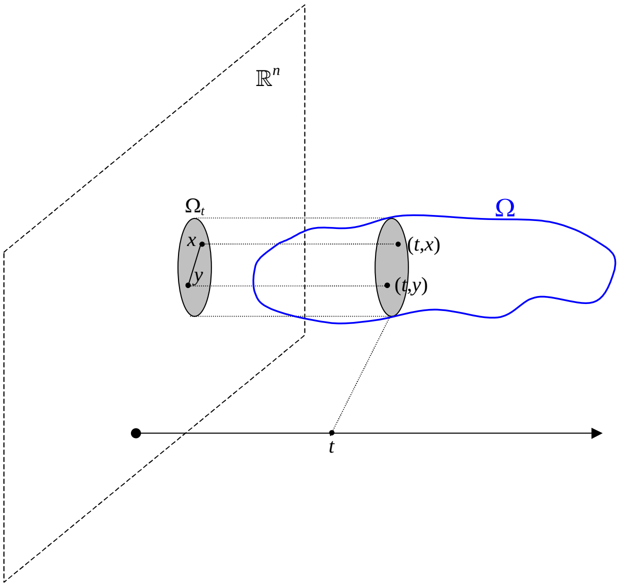
Then there exists a continuous mapping $\varphi(t,x,y) : \Omega' \to \mathbb{R}^{l\times n}$ *such that the following identity holds:*

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

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*

$$\varphi(t,x,x) = f_x(t,x). \tag{3.34}$$



Let us discuss the statement of Lemma 3.8 before the proof. The statement involves the parameter t (that can also be multi-dimensional), which is needed for the application of this Lemma for the proof of Theorem 3.7. However, the statement of Lemma 3.8 is significantly simpler when function f does not depend on t. Indeed, in this case Lemma 3.8 can be stated as follows. Let Ω be a convex open subset of $\mathbb{R}^n$ and $f:\Omega\to\mathbb{R}^l$ be a continuously differentiable function. Then there is a continuous function $\varphi(x,y):\Omega\times\Omega\to\mathbb{R}^{l\times n}$ such that the following identity is true for all $x,y\in\Omega$:

$$f(y)-f(x)=\varphi(x,y)(y-x). \tag{3.35}$$

Furthermore, we have

$$\varphi(x,x)=f_x(x). \tag{3.36}$$

Note that, by the differentiability of f, we have

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

The point of the identity (3.35) is that the term $o(\|x-y\|)$ can be eliminated replacing $f_x(x)$ by a continuous function $\varphi(t,x,y)$.

Consider some simple examples of functions $f(x)$ with $n=l=1$. 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, $\varphi(x,x)=2x=f'(x)$. A similar formula holds for $f(x)=x^k$ with any $k\in\mathbb{N}$:

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

so that $\varphi(x,y)=x^{k-1}+x^{k-2}y+...+y^{k-1}$ and $\varphi(x,x)=kx^{k-1}$.

In the case $n=l=1$, one can define $\varphi(x,y)$ to satisfy (3.35) and (3.36) 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 $f'(x)$.

This argument does not work in the case $n>1$ since one cannot divide by $y-x$. In the general case, we use a different approach.

Proof of Lemma 3.8. 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

$$\varphi_i(t, x, y) = \int_0^1 f_{x_i}(t, x + \lambda(y - x))\, d\lambda. \tag{3.37}$$

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 $B((t, x), \varepsilon)$ and $B((t, y), \varepsilon)$ in $\mathbb{R}^{n+1}$ are contained in Ω. Assuming the norm in all spaces in question is the ∞-norm, we obtain that $B((t, x, y), \varepsilon) \subset \Omega'$. The continuity of φ_i follows from the following general statement.

Lemma 3.9 *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 $u \in U$.

Proof of Lemma 3.9 Let $\{u_k\}_{k=1}^{\infty}$ be a sequence in U that converges to some $u \in U$. Then all u_k with large enough index k are contained in a closed ball $\overline{B}(u, \varepsilon) \subset U$. Since $f(\lambda, u)$ is continuous in $[a, b] \times U$, it is uniformly continuous on any compact set in this domain, in particular, in $[a, b] \times \overline{B}(u, \varepsilon)$. Hence, the convergence

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

is uniform in $\lambda \in [0, 1]$. Since the operations of integration and the uniform convergence are interchangeable, we conclude that $\varphi(u_k) \to \varphi(u)$, which proves the continuity of φ.

The proof of Lemma 3.8 is finished as follows. Consider $f_{x_i}(t, x + \lambda (y - x))$ 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 3.9 that also $\varphi_i(t, x, y)$ is continuous in (t, x, y).

Finally, if $x = y$ then $f_{x_i}(t, x + \lambda (y - x)) = f_{x_i}(t, x)$ which implies by (3.37) that

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

and, hence, $\varphi(t, x, x) = f_x(t, x)$, that is, (3.34). ■

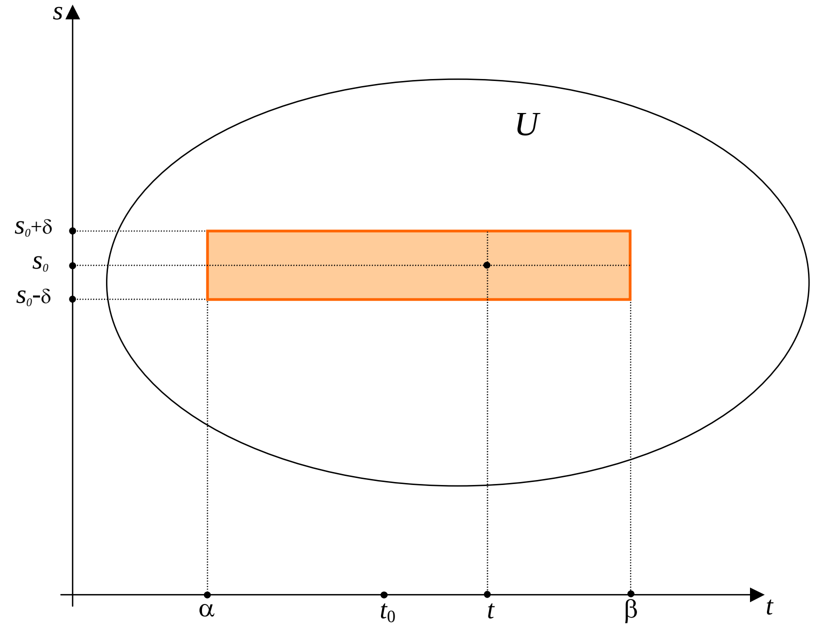
Now we are in position to prove Theorem 3.7.

Proof of Theorem 3.7. Recall that $x(t, s)$ is the maximal solution of the initial value problem (3.28), it is defined in an open set $U \in \mathbb{R}^{m+1}$ and is continuous in (t, s) in U (cf. Theorem 3.6). In the main part of the proof, we show that the partial derivative $\partial_{s_i} x$ exists in U. 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 a value s_0 of the parameter s and prove that $\partial_s x(t, s)$ exists at $s = s_0$ for any $t \in I_{s_0}$, where I_{s_0} is the domain in t of the solution $x(t) := x(t, s_0)$. Since the differentiability is a local property, it suffices to prove that $\partial_s x(t, s)$ exists at $s = s_0$ for any $t \in (\alpha, \beta)$ where $[\alpha, \beta]$ is any bounded closed interval that is a subset of I_{s_0}. For a technical reason, we will assume that (α, β) contains t_0. By Theorem 3.6, for any $\varepsilon > 0$ there is $\delta > 0$ such that the rectangle $[a, \beta] \times [s_0 - \delta, s_0 + \delta]$ is contained in U and

$$\sup_{t \in [\alpha, \beta]} \|x(t, s) - x(t)\| < \varepsilon \quad \text{for all } s \in [s_0 - \delta, s_0 + \delta] \tag{3.38}$$

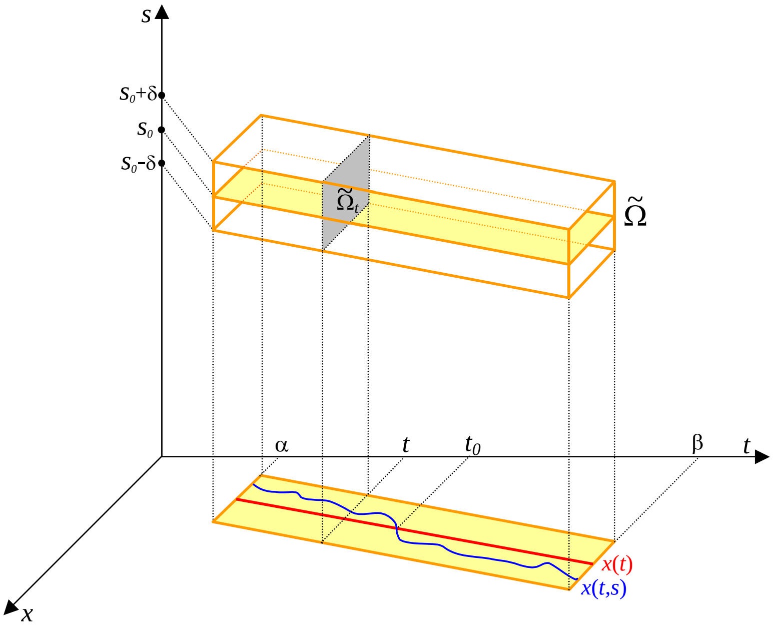
(see the diagram below).



Furthermore, ε and δ can be chosen so small that the following set

$$\widetilde{\Omega} := \left\{(t, x, s) \in \mathbb{R}^{n+m+1} : \alpha < t < \beta,\ \|x - x(t)\| < \varepsilon,\ \ |s - s_0| < \delta\right\} \tag{3.39}$$

is contained in Ω (cf. the proof of Theorem 3.6). It follows from (3.38) that, for all $t \in (\alpha, \beta)$ and $s \in (s_0 - \delta, s_0 + \delta)$, the function $x(t,s)$ is defined and $(t, x(t,s), s) \in \widetilde{\Omega}$ (see the diagram below).



In what follows, we restrict the domain of the variables (t, x, s) to $\widetilde{\Omega}$. Note that $\widetilde{\Omega}$ is an open subset of $\mathbb{R}^{n+m+1}$, and it is is convex with respect to the variable (x, s), for any fixed t. Indeed, by the definition (3.39), $(t, x, s) \in \widetilde{\Omega}$ if and only if

$$(x, s) \in B(x(t), \varepsilon) \times (s_0 - \delta, s_0 + \delta)$$

so that $\widetilde{\Omega}_t = B(x(t), \varepsilon) \times (s_0 - \delta, s_0 + \delta)$. Since both the ball $B(x(t), \varepsilon)$ and the interval $(s_0 - \delta, s_0 + \delta)$ are convex sets, $\widetilde{\Omega}_t$ is also convex.

Applying the Hadamard lemma to the function $f(t, x, s)$ in the domain $\widetilde{\Omega}$ and using the fact that f is continuously differentiable with respect to (x, s), we obtain the identity

$$f(t, y, s) - f(t, x, s_0) = \varphi(t, x, s_0, y, s)(y - x) + \psi(t, x, s_0, y, s)(s - s_0),$$

where φ and ψ are continuous functions on the appropriate domains. In particular, substituting $x = x(t)$ and $y = x(t, s)$, we obtain

$$\begin{aligned} f(t, x(t,s), s) - f(t, x(t), s_0) &= \varphi(t, x(t), s_0, x(t,s), s)(x(t,s) - x(t)) \\ &\quad + \psi(t, x(t), s_0, x(t,s), s)(s - s_0) \\ &= a(t,s)(x(t,s) - x(t)) + b(t,s)(s - s_0), \end{aligned}$$

where the functions

$$a(t,s) = \varphi(t, x(t), s_0, x(t,s), s) \text{ and } b(t,s) = \psi(t, x(t), s_0, x(t,s), s) \tag{3.40}$$

are continuous in $(t, s) \in (\alpha, \beta) \times (s_0 - \delta, s_0 + \delta)$ (the dependence on s_0 is suppressed because s_0 is fixed).

For any $s \in (s_0 - \delta, s_0 + \delta) \setminus \{s_0\}$ and $t \in (\alpha, \beta)$, set

$$z(t,s) = \frac{x(t,s) - x(t)}{s - s_0} \tag{3.41}$$

and observe that, by (3.28) and (3.40),

$$\begin{aligned} z' &= \frac{x'(t,s) - x'(t)}{s - s_0} = \frac{f(t, x(t,s), s) - f(t, x(t), s_0)}{s - s_0} \\ &= a(t,s)\, z + b(t,s). \end{aligned}$$

Note also that $z(t_0, s) = 0$ because both $x(t,s)$ and $x(t, s_0)$ satisfy the same initial condition. Hence, function $z(t,s)$ solves for any fixed $s \in (s_0 - \delta, s_0 + \delta) \setminus \{s_0\}$ the IVP

$$\begin{cases} z' = a(t,s)\, z + b(t,s) \\ z(t_0, s) = 0. \end{cases} \tag{3.42}$$

Since this ODE is linear and the functions a and b are continuous in $(t,s) \in (\alpha, \beta) \times (s_0 - \delta, s_0 + \delta)$, we conclude by Theorem 2.1 that the solution to this IVP exists for all $s \in (s_0 - \delta, s_0 + \delta)$ and $t \in (\alpha, \beta)$ and, by Theorem 3.6, the solution is continuous in $(t,s) \in (\alpha, \beta) \times (s_0 - \delta, s_0 + \delta)$. By the uniqueness theorem, the solution of (3.42) for $s \neq s_0$ coincides with the function $z(t,s)$ defined by (3.41). Although (3.41) does not allow to define $z(t,s)$ for $s = s_0$, we can define $z(t, s_0)$ as the solution of the IVP (3.42) with $s = s_0$. Using the continuity of $z(t,s)$ in s, we obtain

$$\lim_{s \to s_0} z(t,s) = z(t, s_0),$$

that is,

$$\partial_s x(t,s)\,|_{s=s_0} = \lim_{s \to s_0} \frac{x(t,s) - x(t,s_0)}{s - s_0} = \lim_{s \to s_0} z(t,s) = z(t, s_0).$$

Hence, the derivative $y(t) = \partial_s x(t,s)\,|_{s=s_0}$ exists and is equal to $z(t, s_0)$, that is, $y(t)$ satisfies the IVP

$$\begin{cases} y' = a(t, s_0)\, y + b(t, s_0), \\ y(t_0) = 0. \end{cases}$$

Note that by (3.40) and Lemma 3.8

$$a(t, s_0) = \varphi(t, x(t), s_0, x(t), s_0) = f_x(t, x(t), s_0)$$

and

$$b(t, s_0) = \psi(t, x(t), s_0, x(t), s_0) = f_s(t, x(t), s_0)$$

Hence, we obtain that $y(t)$ satisfies the variational equation (3.29).

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 (3.29) and, hence, is continuous in (t,s) by Theorem 3.6. Finally, for the derivative $\partial_t x$ we have the identity

$$\partial_t x = f(t, x(t,s), s), \tag{3.43}$$

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

Remark. It follows from (3.43) that $\partial_t x$ is differentiable in s and, by the chain rule,

$$\partial_s \left(\partial_t x\right) = \partial_s \left[f\left(t, x\left(t, s\right), s\right)\right] = f_x\left(t, x\left(t, s\right), s\right) \partial_s x + f_s\left(t, x\left(t, s\right), s\right). \tag{3.44}$$

On the other hand, it follows from (3.29) that

$$\partial_t \left(\partial_s x\right) = \partial_t y = f_x\left(t, x\left(t, s\right), s\right) \partial_s x + f_s\left(t, x\left(t, s\right), s\right), \tag{3.45}$$

whence we conclude that

$$\partial_s \partial_t x = \partial_t \partial_s x. \tag{3.46}$$

Hence, the derivatives ∂_s and ∂_t commute[59] on x. As we have seen above, if one knew the identity (3.46) a priori then the derivation of the variational equation (3.29) would have been easy. However, in the present proof the identity (3.46) comes *after* the variational equation.

Theorem 3.10 *Under the conditions of* Theorem 3.7, *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* $\alpha = \left(\alpha_1, ..., \alpha_m\right)$ *of the order* $|\alpha| \leq k$, *we have*

$$\partial_t \partial_s^\alpha x = \partial_s^\alpha \partial_t x. \tag{3.47}$$

A multiindex α is a sequence $\left(\alpha_1, ..., \alpha_m\right)$ of m non-negative integers α_i, its order $|\alpha|$ is defined by $|\alpha| = \alpha_1 + ... + \alpha_n$, and the derivative ∂_s^α is defined by

$$\partial_s^\alpha = \frac{\partial^{|\alpha|}}{\partial s_1^{\alpha_1} ... \partial s_m^{\alpha_m}}.$$

Proof. Induction in k. If $k = 1$ then the fact that $x \in C^1\left(s\right)$ is the claim of Theorem 3.7, and the equation (3.47) with $|\alpha| = 1$ was verified in the above Remark. Let us make the inductive step from $k - 1$ to k, for any $k \geq 2$. 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 3.7

$$\begin{cases} y' = f_x\left(t, x, s\right) y + f_s\left(t, x, s\right), \\ y\left(t_0\right) = 0, \end{cases} \tag{3.48}$$

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 (3.48) 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 $x \in C^k\left(s\right)$.

[59] 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 (3.44) and (3.45), which prove at the same time also their equality.

$\boxed{\text{Lecture 21}}$ 29.06.2009 Prof. A. Grigorian, ODE, SS 2009

Let us now prove (3.47). Choose some index i so that $\alpha_i \geq 1$, and set $\beta = (\alpha_1, ..., \alpha_i - 1, ...\alpha_n)$, where 1 is subtracted only at the position i. It follows that $\partial_s^\alpha = \partial_s^\beta \partial_{s_i}$. Set $y_i = \partial_{s_i} x$ so that y_i is the i-th column of the matrix $y = \partial_s x$. It follows from (3.48) that the vector-function y_i satisfies the ODE

$$y_i' = f_x(t, x, s)\, y_i + f_{s_i}(t, x, s). \tag{3.49}$$

Using the identity $\partial_s^\alpha = \partial_s^\beta \partial_{s_i}$ that holds on $C^k(s)$ functions, the fact that $f(t, x(t,s), s) \in C^k(s)$, which follows from the first part of the proof, and the chain rule, we obtain

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

Observe that the function on the right hand side of (3.49) belongs to the class $C^{k-1}(y, s)$. Applying the inductive hypothesis to the ODE (3.49) and noticing that $|\beta| = k - 1$, we obtain

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

whence it follows that

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

∎

4 Qualitative analysis of ODEs

4.1 Autonomous systems

Consider a vector ODE

$$x' = f(x) \tag{4.1}$$

where the right hand side does not depend on t. Such equations are called *autonomous.* Here f is defined on an open set $\Omega \subset \mathbb{R}^n$ (or $\Omega \subset \mathbb{C}^n$) and takes values in $\mathbb{R}^n$ (resp., $\mathbb{C}^n$), so that the domain of the ODE is $\mathbb{R} \times \Omega$.

Definition. The set Ω is called the *phase space* of the ODE. Any path $x : I \to \Omega$, where $x(t)$ is a solution of the ODE on an interval I, is called a *phase trajectory.* A plot of all (or typical) phase trajectories is called the *phase diagram* or the *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.

Assume in the sequel that f is continuously differentiable in Ω. For any $s \in \Omega$, denote by $x(t, s)$ the maximal solution to the IVP

$$\begin{cases} x' = f(x) \\ x(0) = s. \end{cases}$$

Recall that, by Theorem 3.7, the domain of function $x(t, s)$ is an open subset of $\mathbb{R}^{n+1}$ and $x(t, s)$ is continuously differentiable in this domain.

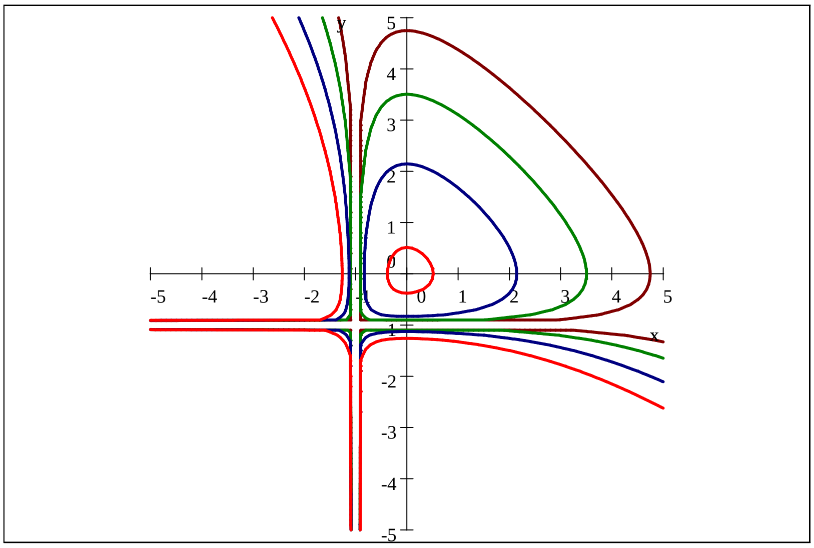
The fact that f does not depend on t, implies easily the following two consequences.

1. If $x(t)$ is a solution of (4.1) then also $x(t-a)$ is a solution of (4.1), for any $a \in \mathbb{R}$. In particular, the function $x(t-t_0, s)$ solves the following IVP
$$\begin{cases} x' = f(x) \\ x(t_0) = s. \end{cases}$$
2. If $f(x_0) = 0$ for some $x_0 \in \Omega$ then the constant function $x(t) \equiv x_0$ is a solution of $x' = f(x)$. Conversely, if $x(t) \equiv x_0$ is a constant solution then $f(x_0) = 0$.

Definition. If $f(x_0) = 0$ at some point $x_0 \in \Omega$ then x_0 is called a *stationary point*[60] or a *stationary solution* of the ODE $x' = f(x)$.

It follows from the above observation that $x_0 \in \Omega$ is a stationary point if and only if $x(t, x_0) \equiv x_0$. It is frequently the case that the stationary points determine the shape of the phase diagram.

Example. Consider the following system
$$\begin{cases} x' = y + xy \\ y' = -x - xy \end{cases} \tag{4.2}$$
that can be partially solved as follows. Dividing one equation by the other, we obtain a separable ODE for y as a function of x:
$$\frac{dy}{dx} = -\frac{x(1+y)}{y(1+x)},$$
whence it follows that
$$\int \frac{ydy}{1+y} = -\int \frac{xdx}{1+x}$$
and
$$y - \ln|y+1| + x - \ln|x+1| = C. \tag{4.3}$$
Although the equation (4.3) does not give the dependence of x and y of t, it does show the dependence between x and y and allows to plot the phase diagram using different values of the constant C:



One can see that there are two points of "attraction" on this plot: $(0,0)$ and $(-1,-1)$, that are exactly the stationary points of (4.2).

[60] In the literature one can find the following synonyms for the term "stationary point": rest point, singular point, equilibrium point, fixed point.

Definition. A stationary point x_0 is called *(Lyapunov) stable* for the system $x' = f(x)$ (or the system is called stable at x_0) if, for any $\varepsilon > 0$, there exists $\delta > 0$ with the following property: for all $s \in \Omega$ such that $\|s - x_0\| < \delta$, the solution $x(t, s)$ is defined for all $t > 0$ and

$$\sup_{t\in[0,+\infty)} \|x(t, s) - x_0\| < \varepsilon. \tag{4.4}$$

In other words, the Lyapunov stability means that if $x(0)$ is close enough to x_0 then the solution $x(t)$ is defined for all $t > 0$ and

$$x(0) \in B(x_0, \delta) \implies x(t) \in B(x_0, \varepsilon) \text{ for all } t > 0.$$

If we replace in (4.4) the interval $[0, +\infty)$ by any bounded interval $[0, T]$ then, by the continuity of $x(t, s)$,

$$\sup_{t\in[0,T]} \|x(t, s) - x_0\| = \sup_{t\in[0,T]} \|x(t, s) - x(t, x_0)\| \to 0 \text{ as } s \to x_0.$$

Hence, the main issue for the stability is the behavior of the solutions as $t \to +\infty$.

Definition. A stationary point x_0 is called *asymptotically stable* for the system $x' = f(x)$ (or the system is called asymptotically stable at x_0), if it is Lyapunov stable and, in addition,

$$\|x(t, s) - x_0\| \to 0 \text{ as } t \to +\infty,$$

provided $\|s - x_0\|$ 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 all complex eigenvalues λ of A, we have* $\operatorname{Re}\lambda < 0$ *then* 0 *is asymptotically stable for the system $x' = Ax$. If, for some eigenvalue λ of A,* $\operatorname{Re}\lambda \geq 0$ *then* 0 *is not asymptotically stable. If, for some eigenvalue λ of A,* $\operatorname{Re}\lambda > 0$ *then* 0 *is unstable.*

Proof. By Corollary to Theorem 2.20, the general complex solution of the system $x' = Ax$ has the form

$$x(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t} P_k(t), \tag{4.5}$$

where C_k are arbitrary complex constants, $\lambda_1, ..., \lambda_n$ are all the eigenvalues of A listed with the algebraic multiplicity, and $P_k(t)$ are some vector valued polynomials of t. The latter means that $P_k(t) = u_1 + u_2 t + ... + u_l t^{l-1}$ for some $l \in \mathbb{N}$ and for some vectors $u_1, ..., u_l$.

It follows from (4.5) that, for all $t \geq 0$,

$$\begin{aligned}\|x(t)\| &\leq \sum_{k=1}^{n} \left|C_k e^{\lambda_k t}\right| \|P_k(t)\| \\ &\leq \max_k |C_k| e^{(\operatorname{Re}\lambda_k)t} \sum_{k=1}^{n} \|P_k(t)\| \\ &\leq \max_k |C_k| e^{\alpha t} \sum_{k=1}^{n} \|P_k(t)\|\end{aligned} \tag{4.6}$$

where

$$\alpha = \max_k \operatorname{Re}\lambda_k < 0.$$

Observe that the polynomials admits the estimates of the type

$$\|P_k(t)\| \leq c\left(1 + t^N\right)$$

for all $t > 0$ and for some large enough constants c and N. On the other hand, since

$$x(0) = \sum_{k=1}^{n} C_k P_k(0),$$

we see that the coefficients C_k are the components of the initial value $x(0)$ in the basis $\{P_k(0)\}_{k=1}^{n}$. Therefore, $\max |C_k|$ is nothing other but the ∞-norm $\|x(0)\|_\infty$ in that basis. Replacing this norm by the current norm in $\mathbb{R}^n$ at expense of an additional constant multiple, we obtain from (4.6) that

$$\|x(t)\| \leq C \|x(0)\| e^{\alpha t}\left(1 + t^N\right) \tag{4.7}$$

for some constant C and all $t \geq 0$.

Since the function $e^{\alpha t}\left(1 + t^N\right)$ is bounded on $[0, +\infty)$, we obtain that, for all $t \geq 0$,

$$\|x(t)\| \leq K \|x(0)\|,$$

where $K = C \sup_{t\geq 0} e^{\alpha t}\left(1 + t^N\right)$, whence it follows that the stationary point 0 is Lyapunov stable. Moreover, since

$$e^{\alpha t}\left(1 + t^N\right) \to 0 \text{ as } t \to +\infty,$$

we conclude from (4.7) that $\|x(t)\| \to 0$ as $t \to \infty$, that is, the stationary point 0 is asymptotically stable.

Assume now that $\operatorname{Re}\lambda \geq 0$ for some eigenvalue λ and prove that the stationary point 0 is not asymptotically stable. For that it suffices to construct a real solution $x(t)$ of the system $x' = Ax$ such that $\|x(t)\| \not\to 0$ as $t \to +\infty$. Indeed, for any real $c \neq 0$, the solution $cx(t)$ has the initial value $cx(0)$, which can be made arbitrarily small provided c is small enough, whereas $cx(t)$ does not go to 0 as $t \to +\infty$, which implies that there is no asymptotic stability. To construct such a solution $x(t)$, fix an eigenvector v of the eigenvalue λ with $\operatorname{Re}\lambda \geq 0$. Then we have a particular solution

$$x(t) = e^{\lambda t} v \tag{4.8}$$

for which

$$\|x(t)\| = \left|e^{\lambda t}\right| \|v\| = e^{t \operatorname{Re}\lambda} \|v\| \geq \|v\| . \tag{4.9}$$

Hence, $x(t)$ does not tend to 0 as $t \to \infty$. If $x(t)$ is a real solution then this completes the proof of the asymptotic instability. If $x(t)$ is a complex solution then both $\operatorname{Re} x(t)$ and $\operatorname{Im} x(t)$ are real solutions and one of them must not go to 0 as $t \to \infty$.

Finally, assume that $\operatorname{Re}\lambda > 0$ for some eigenvalue λ and prove that 0 is unstable. It suffices to prove the existence of a real solution $x(t)$ such that $\|x(t)\|$ is unbounded. For the same solution (4.8), we see from (4.9) that $\|x(t)\| \to \infty$ as $t \to +\infty$, which settles the claim in the case when $x(t)$ is real-valued. For a complex-valued $x(t)$, one of the solutions $\operatorname{Re} x(t)$ or $\operatorname{Im} x(t)$ is unbounded, which finishes the proof. ∎

Note that Theorem 4.1 does not answer the question whether 0 is Lyapunov stable or not in the case when $\operatorname{Re}\lambda \leq 0$ for all eigenvalues λ but there is an eigenvalue λ with $\operatorname{Re}\lambda = 0$. Although we know that 0 is not asymptotically stable in his case, it can actually be either Lyapunov stable or unstable, as we will see below.

Consider in details the case $n = 2$ where we give a full classification of the stability cases and a detailed description of the phase diagrams. Consider a linear system $x' = Ax$ in $\mathbb{R}^2$ where A is a constant linear 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

$$x(t) = C_1 e^{\lambda_1 t} b_1 + C_2 e^{\lambda_2 t} b_2.$$

In other words, in the basis b,

$$x(t) = \left(C_1 e^{\lambda_1 t}, C_2 e^{\lambda_2 t}\right)$$

and $x(0) = (C_1, C_2)$. It follows that

$$\|x(t)\|_\infty = \max\left(\left|C_1 e^{\lambda_1 t}\right|, \left|C_2 e^{\lambda_2 t}\right|\right) = \max\left(|C_1| e^{\operatorname{Re}\lambda_1 t}, |C_2| e^{\operatorname{Re}\lambda_2 t}\right) \le \|x(0)\|_\infty e^{\alpha t}$$

where

$$\alpha = \max(\operatorname{Re}\lambda_1, \operatorname{Re}\lambda_2).$$

If $\alpha \le 0$ then

$$\|x(t)\|_\infty \le \|x(0)\|$$

which implies the Lyapunov stability. As we know from Theorem 4.1, if $\alpha > 0$ then the stationary point 0 is unstable. Hence, in this particular situation, the Lyapunov stability is equivalent to $\alpha \le 0$.

Let us construct the phase diagrams of the system $x' = Ax$ under the above assumptions.

Case λ_1, λ_2 are real.

Let $x_1(t)$ and $x_2(t)$ be the components of the solution $x(t)$ in the basis $\{b_1, b_2\}$. Then

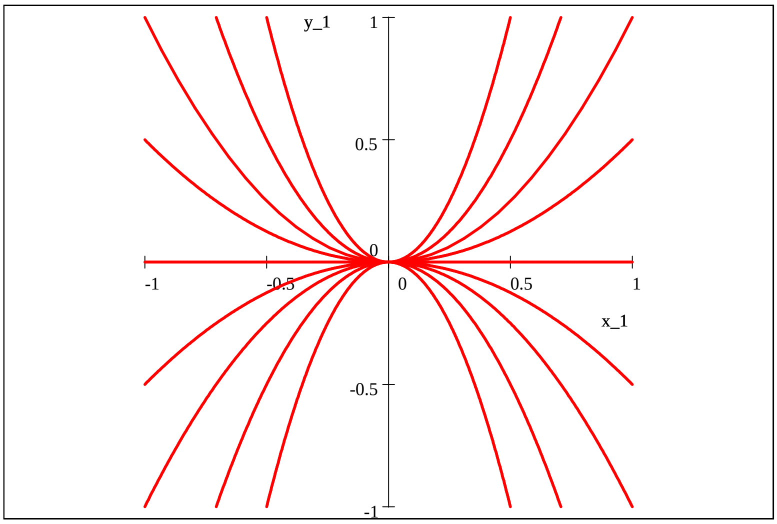
$$x_1 = C_1 e^{\lambda_1 t} \quad \text{and} \quad x_2 = C_2 e^{\lambda_2 t}.$$

Assuming that $\lambda_1, \lambda_2 \neq 0$, we obtain the relation between x_1 and x_2 as follows:

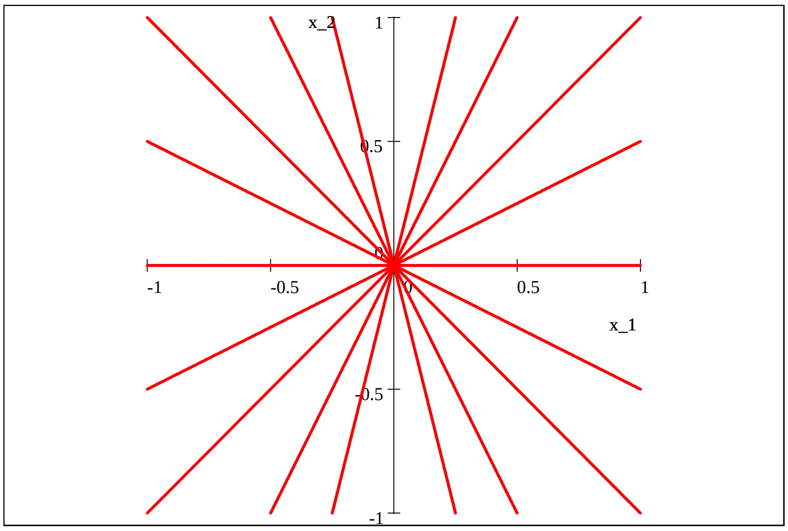
$$x_2 = C|x_1|^\gamma,$$

where $\gamma = \lambda_2/\lambda_1$. Hence, the phase diagram consists of all curves of this type as well as of the half-axis $x_1 > 0, x_1 < 0, x_2 > 0, x_2 < 0$.

If $\gamma > 0$ (that is, λ_1 and λ_2 are of the same sign) then the phase diagram (or the 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 $\gamma > 1$:

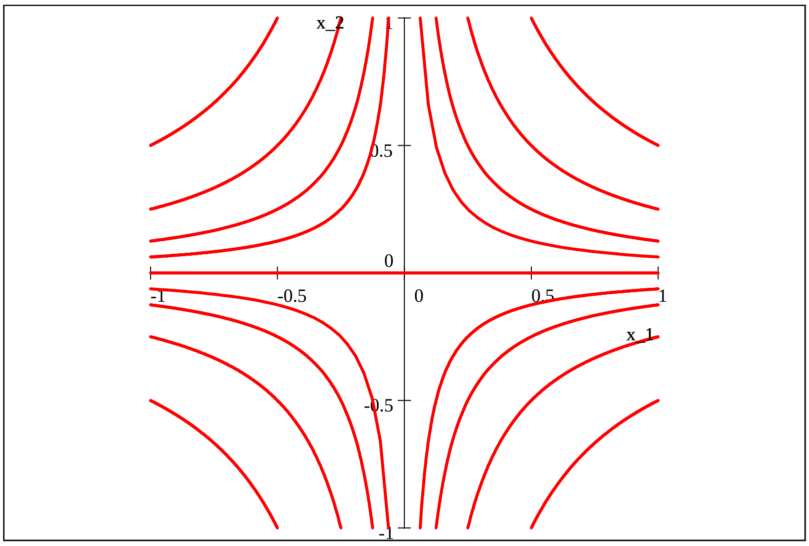


and here is a node with $\gamma = 1$:



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 $\beta \neq 0$.

Note that b_1 is the eigenvector of λ_1 and $b_2 = \overline{b_1}$ is the eigenvector of λ_2. Let $u = \operatorname{Re} b_1$ and $v = \operatorname{Im} b_1$ so that $b_1 = u + iv$ and $b_2 = u - iv$. Since the eigenvectors b_1 and b_2 form a basis in $\mathbb{C}^2$, it follows that also the vectors $u = \frac{b_1+b_2}{2}$ and $v = \frac{b_1-b_2}{2i}$ form a basis in $\mathbb{C}^2$; consequently, the couple u, v is a basis in $\mathbb{R}^2$. The general real solution is

$$\begin{aligned} x(t) &= C_1 \operatorname{Re} e^{(\alpha-i\beta)t} b_1 + C_2 \operatorname{Im} e^{(\alpha-i\beta)t} b_1 \\ &= C_1 e^{\alpha t} \operatorname{Re}(\cos\beta t - i\sin\beta t)(u+iv) + C_2 e^{\alpha t} \operatorname{Im}(\cos\beta t - i\sin\beta t)(u+iv) \\ &= e^{\alpha t}(C_1\cos\beta t - C_2\sin\beta t)\, u + e^{\alpha t}(C_1 \sin\beta t + C_2\cos\beta t)\, v \\ &= Ce^{\alpha t}\cos(\beta t + \psi)\, u + Ce^{\alpha t}\sin(\beta t+\psi)\, v \end{aligned}$$

where $C = \sqrt{C_1^2 + C_2^2}$ and

$$\cos\psi = \frac{C_1}{C}, \ \sin\psi = \frac{C_2}{C}.$$

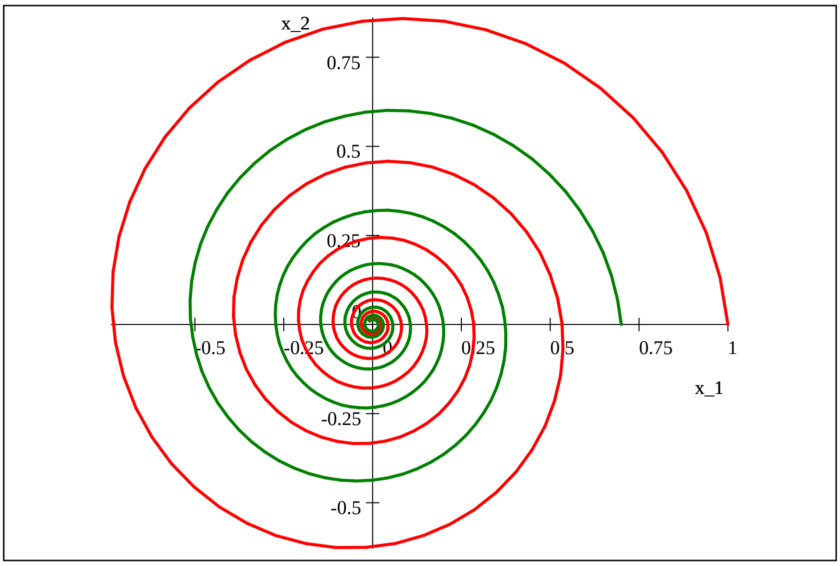
Hence, in the basis (u, v), the solution $x(t)$ is as follows:

$$x(t) = C \begin{pmatrix} e^{\alpha t}\cos(\beta t+\psi) \\ e^{\alpha t}\sin(\beta t + \psi) \end{pmatrix}.$$

If (r, θ) are the polar coordinates on the plane in the basis (u, v), then the polar coordinates of the solution $x(t)$ are

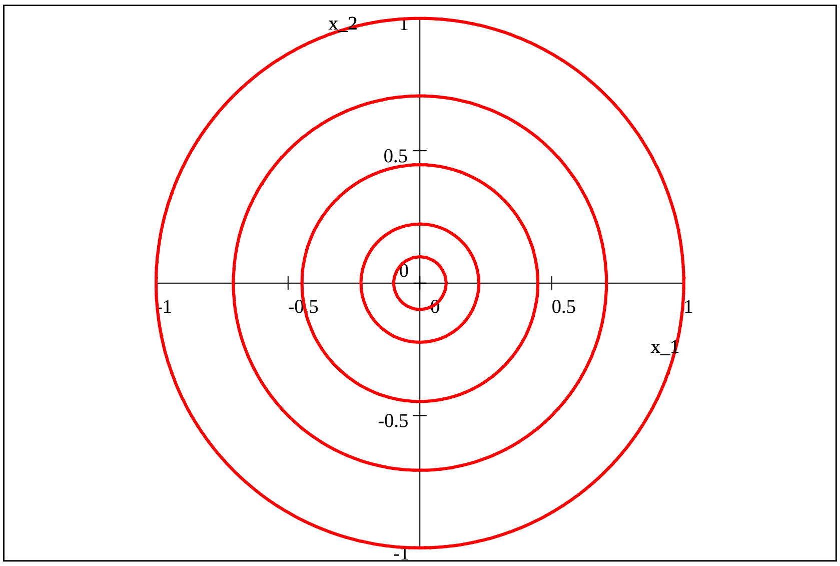
$$r(t) = Ce^{\alpha t} \ \text{ and } \ \theta(t) = \beta t + \psi.$$

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 if $\alpha < 0$ 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. By Theorem 2.20, the general solution is

$$\begin{aligned} x(t) &= C_1 e^{\lambda t} b_1 + C_2 e^{\lambda t} (b_1 t + b_2) \\ &= (C_1 + C_2 t) e^{\lambda t} b_1 + C_2 e^{\lambda t} b_2 \end{aligned}$$

whence $x(0) = C_1 b_1 + C_2 b_2$. That is, in the basis b, we can write $x(0) = (C_1, C_2)$ and

$$x(t) = \left(e^{\lambda t}(C_1 + C_2 t), e^{\lambda t} C_2\right) \tag{4.10}$$

whence

$$\|x(t)\|_1 = e^{\lambda t} |C_1 + C_2 t| + e^{\lambda t} |C_2|.$$

If $\lambda < 0$ then we obtain again the asymptotic stability (which follows also from Theorem 4.1), whereas in the case $\lambda \geq 0$ the stationary point 0 is unstable. Indeed, taking $C_1 = 0$ and $C_2 = 1$, we obtain a particular solution with the norm

$$\|x(t)\|_1 = e^{\lambda t}(1 + t),$$

which is unbounded, whenever $\lambda \geq 0$.

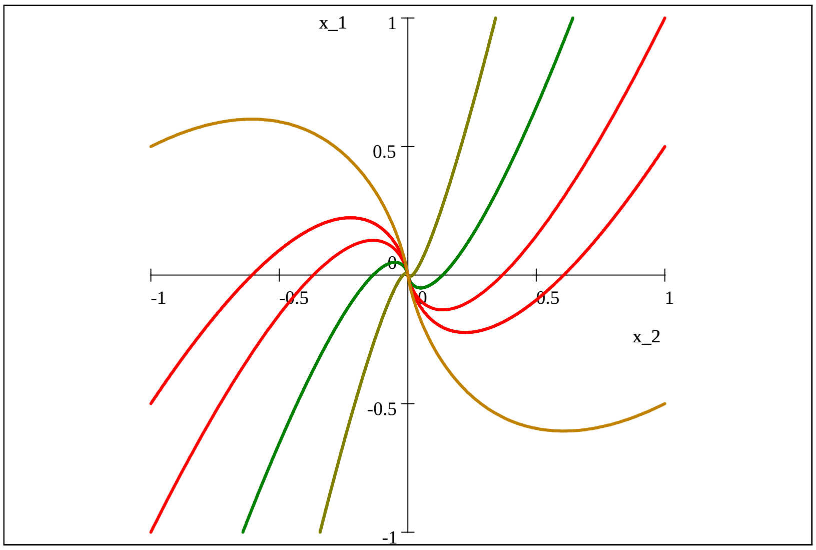
If $\lambda \neq 0$ then it follows from (4.10) that the components x_1, x_2 of x are related as follows:

$$\frac{x_1}{x_2} = \frac{C_1}{C_2} + t \quad \text{and} \quad t = \frac{1}{\lambda} \ln \frac{x_2}{C_2}$$

whence

$$x_1 = C x_2 + \frac{x_2 \ln |x_2|}{\lambda},$$

where $C = \frac{C_1}{C_2} - \ln |C_2|$. 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:

- a *node* (λ_1, λ_2 are real, non-zero and of the same sign);
- a *saddle* (λ_1, λ_2 are real, non-zero and of opposite signs);
- a *focus or spiral* (λ_1, λ_2 are imaginary and $\operatorname{Re} \lambda \neq 0$);
- a *center* (λ_1, λ_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 $\max \operatorname{Re} \lambda = 0$ both stability and instability can happen, depending on the structure of the Jordan normal form.

4.3 Lyapunov's theorems

Consider again an autonomous ODE $x' = f(x)$ where $f : \Omega \to \mathbb{R}^n$ is continuously differentiable and Ω is an open set in $\mathbb{R}^n$. Let x_0 be a stationary point of the system $x' = f(x)$, that is, $f(x_0) = 0$. We investigate the stability of the stationary point x_0.

Theorem 4.2 (Lyapunov's theorem) *Assume that $f \in C^2(\Omega)$ and set $A = f_x(x_0)$ (that is, A is the Jacobian matrix of f at x_0).*

(a) If $\operatorname{Re} \lambda < 0$ for all eigenvalues λ of A then the stationary point x_0 is asymptotically stable for the system $x' = f(x)$.

(b) If $\operatorname{Re} \lambda > 0$ for some eigenvalue λ of A then the stationary point x_0 is unstable for the system $x' = f(x)$.

The proof of part (b) is somewhat lengthy and will not be presented here.

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

$$A = f_x(0,0) = \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ \partial_x f_2 & \partial_y f_2 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ 3 & -4 \end{pmatrix}.$$

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.

Coming back to the setting of Theorem 4.2, represent the function $f(x)$ in the form

$$f(x) = A(x - x_0) + \varphi(x)$$

where $\varphi(x) = o(\|x - x_0\|)$ as $x \to x_0$. Consider a new unknown function $X(t) = x(t) - x_0$ that obviously satisfies then the ODE

$$X' = AX + o(\|X\|) \,.$$

By dropping out the term $o(\|X\|)$, we obtain the *linearized* equation $X' = AX$. Frequently, the matrix A can be found directly by the linearization rather than by evaluating $f_x(x_0)$.

The stability of the stationary point 0 of the linear ODE $X' = AX$ is closely related (but not identical) to the stability of the stationary point x_0 of the non-linear ODE $x' = f(x)$. The hypotheses that $\operatorname{Re}\lambda < 0$ for all the eigenvalues λ of A yields by Theorem 4.2 that x_0 is asymptotically stable for the non-linear system $x' = f(x)$, and by Theorem 4.1 that 0 is asymptotically stable for the linear system

$X' = AX$. Similarly, under the hypothesis $\operatorname{Re}\lambda > 0$, x_0 is unstable for $x' = f(x)$ and 0 is unstable for $X' = AX$. However, in the case $\max \operatorname{Re}\lambda = 0$, the stability types for ODEs $x' = f(x)$ and $X' = AX$ are not necessarily identical.

Example. Consider the system

$$\begin{cases} x' = y + xy, \\ y' = -x - xy. \end{cases} \tag{4.11}$$

Solving the equations

$$\begin{cases} y + xy = 0 \\ x + xy = 0 \end{cases}$$

we obtain the two stationary points $(0, 0)$ and $(-1, -1)$. Let us linearize the system at the stationary point $(-1, -1)$. Setting $X = x + 1$ and $Y = y + 1$, we obtain the system

$$\begin{cases} X' = (Y - 1) X = -X + XY = -X + o(\| (X, Y) \|) \\ Y' = -(X - 1) Y = Y - XY = Y + o(\| (X, Y) \|) \end{cases} \tag{4.12}$$

whose linearization is

$$\begin{cases} X' = -X \\ Y' = Y. \end{cases}$$

Hence, the matrix is

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the eigenvalues are -1 and $+1$ so that the type of the stationary point is a saddle. Hence, the system (4.11) is unstable at $(-1, -1)$ because one of the eigenvalues is positive.

Consider now the stationary point $(0, 0)$. Near this point, the system (4.11) can be written in the form

$$\begin{cases} x' = y + o(\| (x, y) \|) \\ y' = -x + o(\| (x, y) \|) \end{cases}$$

so that the linearized system is

$$\begin{cases} x' = y, \\ y' = -x. \end{cases}$$

Hence, the matrix is

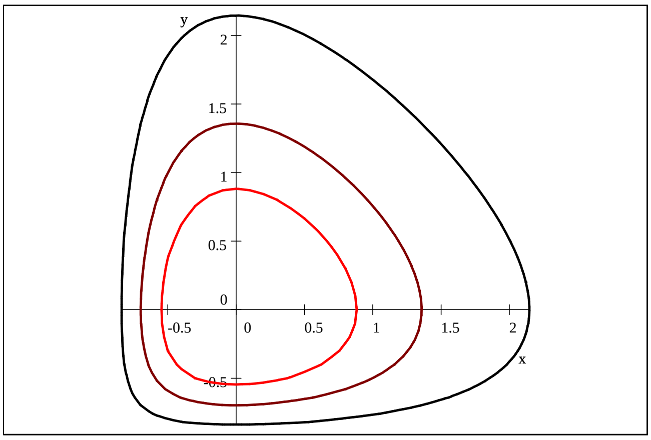
$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the eigenvalues are $\pm i$. Since they are purely imaginary, the type of the stationary point $(0, 0)$ for the linearized system is a center. Hence, the linearized system is stable at $(0, 0)$ but not asymptotically stable. For the non-linear system (4.11), no conclusion can be drawn from the eigenvalues.

In this case, one can use the fact that the phase trajectories for the system (4.11) can be explicitly described by the equation:

$$x - \ln|x + 1| + y - \ln|y + 1| = C.$$

(cf. (4.3)). It follows that the phase trajectories near $(0, 0)$ are closed curves (see the plot below) and, hence, the stationary point $(0, 0)$ of the system (4.11) is Lyapunov stable but not asymptotically stable.


y
2
1.5
1
0.5
0
-0.5
-0.5
0
0.5
1
1.5
2
x

The main tool for the proof of Theorem 4.2 is the second Lyapunov theorem, that is of its own interest. Recall that for any vector $u \in \mathbb{R}^n$ and a differentiable function V in a domain in $\mathbb{R}^n$, the directional derivative $\partial_u V$ can be determined by

$$\partial_u V(x) = V_x(x)\, u = \sum_{k=1}^{n} \frac{\partial V}{\partial x_k}(x)\, u_k.$$

Theorem 4.3 (Lyapunov's second theorem) *Consider the system $x' = f(x)$ where $f \in C^1(\Omega)$ and let x_0 be a stationary point of it. Let U be an open subset of Ω containing x_0, and $V(x)$ be a C^1 function on U such that $V(x_0) = 0$ and $V(x) > 0$ for any $x \in U \setminus \{x_0\}$.*

(*a*) *If, for all $x \in U$,*
$$\partial_{f(x)} V(x) \leq 0, \tag{4.13}$$
then the stationary point x_0 is stable.

(*b*) *Let $W(x)$ be a continuous function on U such that $W(x) > 0$ for $x \in U \setminus \{x_0\}$. If, for all $x \in U$,*
$$\partial_{f(x)} V(x) \leq -W(x), \tag{4.14}$$
then the stationary point x_0 is asymptotically stable.

(*c*) *If, for all $x \in U$,*
$$\partial_{f(x)} V(x) \geq W(x), \tag{4.15}$$
where W is as above then the stationary point x_0 is unstable.

A function V from the statement is called the *Lyapunov function.* Note that the vector field $f(x)$ in the expression $\partial_{f(x)} V(x)$ depends on x. By definition, we have

$$\partial_{f(x)} V(x) = \sum_{k=1}^{n} \frac{\partial V}{\partial x_k}(x)\, f_k(x).$$

In this context, $\partial_f V$ is also called the *orbital derivative* of V with respect to the ODE $x' = f(x)$.

There are no general methods for constructing Lyapunov functions. Let us consider some examples.

Example. Consider the system $x' = Ax$ where $A \in \mathcal{L}(\mathbb{R}^n)$. In order to investigate the stability of the stationary point 0, consider the function

$$V(x) = \|x\|_2^2 = \sum_{k=1}^{n} x_k^2,$$

which is positive in $\mathbb{R}^n \setminus \{0\}$ and vanishes at 0. Setting $f(x) = Ax$ and $A = (A_{kj})$, we obtain for the components

$$f_k(x) = \sum_{j=1}^{n} A_{kj} x_j.$$

Since $\frac{\partial V}{\partial x_k} = 2x_k$, it follows that

$$\partial_f V = \sum_{k=1}^{n} \frac{\partial V}{\partial x_k} f_k = 2 \sum_{j,k=1}^{n} A_{kj} x_j x_k.$$

Recall that the matrix A is called a *non-positive definite* if

$$\sum_{j,k=1}^{n} A_{kj} x_j x_k \le 0 \text{ for all } x \in \mathbb{R}^n.$$

Hence, if A is non-positive definite, then $\partial_f V \le 0$; by Theorem 4.3(a), we conclude that 0 is Lyapunov stable. The matrix A is called *negative definite* if

$$\sum_{j,k=1}^{n} A_{kj} x_j x_k < 0 \text{ for all } x \in \mathbb{R}^n \setminus \{0\}.$$

In this case, set $W(x) = -2\sum_{j,k=1}^{n} A_{kj} x_j x_k$ so that $\partial_f V = -W$. Then we conclude by Theorem 4.3(b), that 0 is asymptotically stable. Similarly, if the matrix A is *positive definite* then 0 is unstable by Theorem 4.3(c).

For example, if $A = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ where λ_k are real, then A is non-positive definite if all $\lambda_k \le 0$, negative definite if all $\lambda_k < 0$, and positive definite if all $\lambda_k > 0$.

Example. Consider the second order scalar ODE $x'' + kx' = F(x)$ that describes the one-dimensional movement of a particle under the external potential force $F(x)$ and friction with the coefficient k. This ODE can be written as a normal 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 for all $x \in \mathbb{R}$) 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 decide 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 $k \ge 0$ and

$$F(x) < 0 \text{ for } x > 0, \qquad F(x) > 0 \text{ for } x < 0, \tag{4.16}$$

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 $V(x, y)$ vanishes at $(0, 0)$ and is positive away from $(0, 0)$.

Setting

$$f(x,y) = (y, -ky + F(x)),$$

compute the orbital derivative $\partial_f V$:

$$\begin{aligned}\partial_f V &= V_x y + V_y(-ky + F(x)) \\ &= U'(x)\, y + y(-ky + F(x)) \\ &= -F(x)\, y - ky^2 + yF(x) \\ &= -ky^2 \leq 0.\end{aligned}$$

Hence, V is indeed the Lyapunov function, and by Theorem 4.3 the stationary point $(0,0)$ is Lyapunov stable.

Physically the condition (4.16) means that the force always acts in the direction of the origin thus forcing the displaced particle to the origin, which causes the stability.

For example, the following functions satisfy (4.16):

$$F(x) = -x \quad \text{and} \quad F(x) = -x^3.$$

The corresponding Lyapunov functions are

$$V(x,y) = \frac{x^2}{2} + \frac{y^2}{2} \quad \text{and} \quad V(x,y) = \frac{x4}{4} + \frac{y^2}{2},$$

respectively.

Example. Consider a system

$$\begin{cases} x' = y - x \\ y' = -x^3. \end{cases} \tag{4.17}$$

The function $V(x,y) = \frac{x^4}{4} + \frac{y^2}{2}$ is positive everywhere except for the origin. The orbital derivative of this function with respect to the given ODE is

$$\begin{aligned}\partial_f V &= V_x f_1 + V_y f_2 \\ &= x^3(y - x) - yx^3 \\ &= -x^4 \leq 0.\end{aligned}$$

Hence, by Theorem 4.3(a), $(0,0)$ is Lyapunov stable for the system (4.17).

Using a more subtle Lyapunov function $V(x,y) = (x-y)^2 + y^2$, one can show that $(0,0)$ is, in fact, asymptotically stable for the system (4.17). Since the matrix $\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$ of the linearized system (4.17) has the eigenvalues 0 and -1, the stability of $(0,0)$ for the system (4.17) cannot be deduced from Theorem 4.2. Observe that, by Theorem 4.1, the stationary point $(0,0)$ is not asymptotically stable for the linearized system. However, it is still stable since the matrix is diagonalizable (cf. Section 4.2).

Example. Consider again the system (4.11), that is,

$$\begin{cases} x' = y + xy \\ y' = -x - xy \end{cases}$$

and the stationary point $(0,0)$. As it was shown above, the phase trajectories of this system satisfy the following equation:

$$x - \ln|x+1| + y - \ln|y+1| = C.$$

This motivates us to consider the function

$$V(x,y) = x - \ln|x+1| + y - \ln|y+1|$$

in a neighborhood of $(0,0)$. It vanishes at $(0,0)$ and is positive away from $(0,0)$. Its orbital derivative is

$$\begin{aligned}
\partial_f V &= V_x f_1 + V_y f_2 \\
&= \left(1 - \frac{1}{x+1}\right)(y+xy) + \left(1 - \frac{1}{y+1}\right)(-x-xy) \\
&= xy - xy = 0.
\end{aligned}$$

By Theorem 4.3(a), the point $(0,0)$ is Lyapunov stable.

Proof of Theorem 4.3. (a) The main idea is that, as long as a solution $x(t)$ remains in the domain U of V, we have by the chain rule

$$\frac{d}{dt}V(x(t)) = V'(x)\,x'(t) = V'(x)\,f(x) = \partial_{f(x)}V(x) \le 0. \tag{4.18}$$

It follows that the function V is decreasing along any solution $x(t)$. Hence, if $x(0)$ is close to x_0 then $V(x(0))$ must be small, whence it follows that $V(x(t))$ is also small for all $t > 0$, so that $x(t)$ is close to x_0.

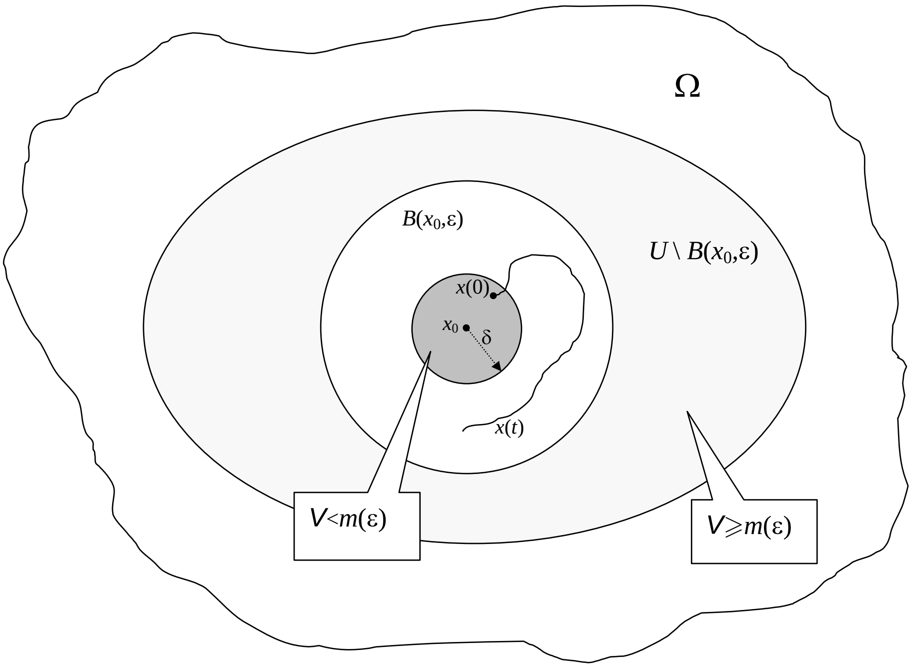
To implement this idea, we first shrink U so that U is bounded and $V(x)$ is defined on the closure $\overline{U}$. Set

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

Let $\varepsilon > 0$ be so small that $\overline{B_\varepsilon} \subset U$. For any such ε, 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

$$V(x) \geq m(\varepsilon) \quad \text{for all} \quad x \in \overline{U} \setminus B_\varepsilon. \tag{4.19}$$

Since $V(x_0) = 0$, for any given $\varepsilon > 0$ there is $\delta > 0$ so small that

$$V(x) < m(\varepsilon) \quad \text{for all} \quad x \in B_\delta. \tag{4.20}$$

Let $x(t)$ be a maximal solution to the ODE $x' = f(x)$ in the domain $\mathbb{R} \times U$, such that $x(0) \in B_\delta$. We prove that $x(t)$ is defined for all $t > 0$ and that $x(t) \in B_\varepsilon$ for all $t > 0$, which will prove the Lyapunov stability of the stationary point x_0. Since $x(0) \in B_\delta$, we have by (4.20) that

$$V(x(0)) < m(\varepsilon).$$

Since the function $V(x(t))$ decreases in t, we obtain

$$V(x(t)) < m(\varepsilon) \quad \text{for all } t > 0,$$

as long as $x(t)$ is defined[61]. It follows from (4.19) that $x(t) \in B_\varepsilon$.

We are left to verify that $x(t)$ is defined for all $t > 0$. Indeed, assume that $x(t)$ is defined only for $t < T$ where T is finite. By Theorem 3.3, when $t \to T-$, then the graph of the solution $x(t)$ must leave any compact subset of $\mathbb{R} \times U$, whereas the graph is contained in the compact set $[0, T] \times \overline{B_\varepsilon}$. This contradiction shows that $T = +\infty$, which finishes the proof.

(b) It follows from (4.14) and (4.18) that

$$\frac{d}{dt} V(x(t)) \leq -W(x(t)).$$

It suffices to show that

$$V(x(t)) \to 0 \text{ as } t \to \infty$$

since this will imply that $x(t) \to x_0$ (recall that x_0 is the only point where V vanishes). Since $V(x(t))$ is decreasing in t, the limit

$$c = \lim_{t \to +\infty} V(x(t))$$

exists. Assume from the contrary that $c > 0$. By the continuity of V, there is $r > 0$ such that

$$V(x) < c \text{ for all } x \in B_r.$$

Since $V(x(t)) \geq c$ for $t > 0$, it follows that $x(t) \notin B_r$ for all $t > 0$. Set

$$m = \inf_{z \in \overline{U} \setminus B_r} W(z) > 0.$$

It follows that $W(x(t)) \geq m$ for all $t > 0$ whence

$$\frac{d}{dt} V(x(t)) \leq -W(x(t)) \leq -m$$

[61] Since $x(t)$ has been defined as the maximal solution in the domain $\mathbb{R} \times U$, the point $x(t)$ is always contained in U as long as it is defined.

for all $t > 0$. However, this implies upon integration in t that

$$V(x(t)) \leq V(x(0)) - mt,$$

whence it follows that $V(x(t)) < 0$ for large enough t. This contradiction finishes the proof.

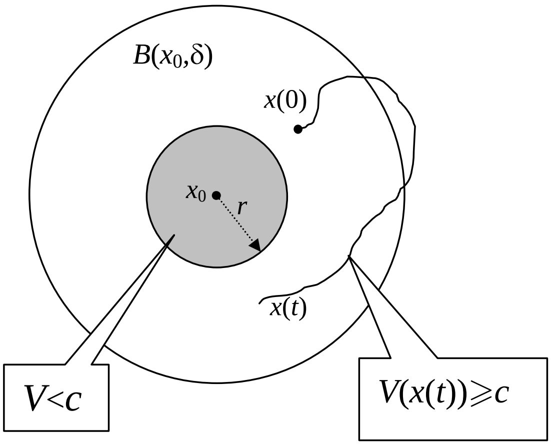
(c) Assume from the contrary that the stationary point x_0 is stable, that is, for any $\varepsilon > 0$ there is $\delta > 0$ such that $x(0) \in B_\delta$ implies that $x(t)$ is defined for all $t > 0$ and $x(t) \in B_\varepsilon$. Let ε be so small that $\overline{B_\varepsilon} \subset U$. Chose $x(0)$ to be any point in $B_\delta \setminus \{x_0\}$. Since $x(t) \in B_\varepsilon$ for all $t > 0$, we have also $x(t) \in U$ for all $t > 0$. It follows from the hypothesis (4.15) that

$$\frac{d}{dt}V(x(t)) \geq W(x(t)) \geq 0$$

so that the function $V(x(t))$ is monotone increasing. Then, for all $t > 0$,

$$V(x(t)) \geq V(x(0)) =: c > 0.$$

Let $r > 0$ be so small that $V(x) < c$ for all $x \in B_r$.



It follows that $x(t) \notin B_r$ for all $t > 0$. Setting as before

$$m = \inf_{z \in \overline{U} \setminus B_r} W(z) > 0,$$

we obtain that $W(x(t)) \geq m$ whence

$$\frac{d}{dt}V(x(t)) \geq m \text{ for all } t > 0.$$

It follows that $V(x(t)) \geq mt \to +\infty$ as $t \to +\infty$, which is impossible since V is bounded in $\overline{U}$. ■

Proof of Theorem 4.2. Without loss of generality, set $x_0 = 0$ so that $f(0) = 0$. By the differentiability of $f(x)$, we have

$$f(x) = Ax + h(x) \tag{4.21}$$

where $h(x) = o(\|x\|)$ as $x \to \infty$. Using that $f \in C^2$, let us prove that, in fact,

$$\|h(x)\| \leq C\|x\|^2 \tag{4.22}$$

provided $\|x\|$ is small enough. Applying the Taylor formula to any component f_k of f, we write

$$f_k(x) = \sum_{i=1}^{n} \partial_i f_k(0)\, x_i + \frac{1}{2}\sum_{i,j=1}^{n} \partial_{ij} f_k(0)\, x_i x_j + o\left(\|x\|^2\right) \text{ as } x \to 0.$$

The first term on the right hand side is the k-th component of Ax, whereas the rest is the k-th component of $h(x)$, that is,

$$h_k(x) = \frac{1}{2}\sum_{i,j=1}^{n} \partial_{ij} f_k(0)\, x_i x_j + o\left(\|x\|^2\right).$$

Setting $B = \max_{i,j,k}|\partial_{ij} f_k(0)|$, we obtain

$$|h_k(x)| \leq B\sum_{i,j=1}^{n}|x_i x_j| + o\left(\|x\|^2\right) = B\|x\|_1^2 + o\left(\|x\|^2\right).$$

Replacing $\|x\|_1$ by $\|x\|$ at expense of an additional constant factor and passing from the components h_k to the vector-function h, we obtain (4.22).

Consider the following function

$$V(x) = \int_0^\infty \left\|e^{sA}x\right\|_2^2 ds, \tag{4.23}$$

and prove that $V(x)$ is the Lyapunov function. Firstly, let us first verify that $V(x)$ is finite for any $x \in \mathbb{R}^n$, that is, the integral in (4.23) converges. Indeed, in the proof of Theorem 4.1 we have established the inequality

$$\left\|e^{tA}x\right\| \leq Ce^{\alpha t}\left(t^N + 1\right)\|x\|, \tag{4.24}$$

where C, N are some positive numbers (depending on A) and

$$\alpha = \max \operatorname{Re}\lambda,$$

where max is taken over all eigenvalues λ of A. Since by hypothesis $\alpha < 0$, (4.24) implies that $\left\|e^{sA}x\right\|$ decays exponentially as $s \to +\infty$, whence the convergence of the integral in (4.23) follows.

Secondly, let us show that $V(x)$ is of the class $C^1(\mathbb{R}^n)$ (in fact, $C^\infty(\mathbb{R}^n)$). For that, represent x in the canonical basis $v_1, ..., v_n$ of $\mathbb{R}^n$ as

$$x = \sum_{i=1}^{n} x_i v_i$$

and notice that

$$\|x\|_2^2 = \sum_{i=1}^{n}|x_i|^2 = x \cdot x.$$

Therefore,

$$\begin{aligned}\left\|e^{sA}x\right\|_2^2 &= e^{sA}x\cdot e^{sA}x=\left(\sum_i x_i\left(e^{sA}v_i\right)\right)\cdot\left(\sum_j x_j\left(e^{sA}v_j\right)\right)\\ &= \sum_{i,j}x_ix_j\left(e^{sA}v_i\cdot e^{sA}v_j\right).\end{aligned}$$

Integrating in s, we obtain

$$V\left(x\right)=\sum_{i,j}b_{ij}x_ix_j,$$

where $b_{ij}=\int_0^\infty\left(e^{sA}v_i\cdot e^{sA}v_j\right)ds$ are constants. This means that $V\left(x\right)$ is a quadratic form that is obviously infinitely many times differentiable in x.

Remark. Usually we work with any norm in $\mathbb{R}^n$. In the definition (4.23) of $V\left(x\right)$, we have specifically chosen the 2-norm to ensure the smoothness of $V\left(x\right)$.

Function $V\left(x\right)$ is obviously non-negative and $V\left(x\right)=0$ if and only if $x=0$. In order to complete the proof of the fact that $V\left(x\right)$ is the Lyapunov function, we need to estimate $\partial_{f(x)}V\left(x\right)$. Noticing that by (4.21)

$$\partial_{f(x)}V\left(x\right)=\partial_{Ax}V\left(x\right)+\partial_{h(x)}V\left(x\right),\tag{4.25}$$

let us first evaluate $\partial_{Ax}V\left(x\right)$ for any $x\in\mathbb{R}^n$. We claim that

$$\partial_{Ax}V\left(x\right)=\frac{d}{dt}V\left(e^{tA}x\right)\bigg|_{t=0}.\tag{4.26}$$

Indeed, using the chain rule, we obtain

$$\frac{d}{dt}V\left(e^{tA}x\right)=V_x\left(e^{tA}x\right)\frac{d}{dt}\left(e^{tA}x\right)$$

whence by Theorem 2.16

$$\frac{d}{dt}V\left(e^{tA}x\right)=V_x\left(e^{tA}x\right)Ae^{tA}x.$$

Setting here $t=0$ and using that $e^{0A}=\mathrm{id}$, we obtain

$$\frac{d}{dt}V\left(e^{tA}x\right)\bigg|_{t=0}=V_x\left(x\right)Ax=\partial_{Ax}V\left(x\right),$$

which proves (4.26).

To evaluate the right hand side of (4.26), observe first that

$$V\left(e^{tA}x\right)=\int_0^\infty\left\|e^{sA}\left(e^{tA}x\right)\right\|_2^2ds=\int_0^\infty\left\|e^{(s+t)A}x\right\|_2^2ds=\int_t^\infty\left\|e^{\tau A}x\right\|_2^2d\tau,$$

where we have made the change $\tau=s+t$. Therefore, differentiating this identity in t, we obtain

$$\frac{d}{dt}V\left(e^{tA}x\right)=-\left\|e^{tA}x\right\|_2^2.$$

Setting $t = 0$ and combining with (4.26), we obtain

$$\partial_{Ax} V(x) = -\|x\|_2^2 .$$

The second term in (4.25) can be estimated as follows:

$$\partial_{h(x)} V(x) = V_x(x)\, h(x) \leq \|V_x(x)\| \, \|h(x)\|$$

(where $\|V_x(x)\|$ is the operator norm of the linear operator $V_x(x) : \mathbb{R}^n \to \mathbb{R}$; the latter is, in fact, an $1 \times n$ matrix that can also be identified with a vector in $\mathbb{R}^n$). It follows that

$$\partial_{f(x)} V(x) = \partial_{Ax} V(x) + \partial_{h(x)} V(x) \leq -\|x\|_2^2 + \|V_x(x)\| \, \|h(x)\| .$$

Using (4.22) and switching there to the 2-norm of x, we obtain that, in a small neighborhood of 0,

$$\partial_{f(x)} V(x) \leq -\|x\|_2^2 + C \|V_x(x)\| \, \|x\|_2 .$$

Observe next that the function $V(x)$ has minimum at 0, which implies that $V_x(0) = 0$. Hence, for any $\varepsilon > 0$,

$$\|V_x(x)\| \leq \varepsilon$$

provided $\|x\|$ is small enough. Choosing $\varepsilon = \frac{1}{2}C$ and combining together the above two lines, we obtain that, in a small neighborhood U of 0,

$$\partial_{f(x)} V(x) \leq -\|x\|_2^2 + \frac{1}{2} \|x\|_2^2 = -\frac{1}{2} \|x\|_2^2 .$$

Setting $W(x) = \frac{1}{2}\|x\|_2^2$, we conclude by Theorem 4.3, that the ODE $x' = f(x)$ is asymptotically stable at 0. ■

4.4 Zeros of solutions

In this section, we consider a scalar linear second order ODE

$$x'' + p(t)\,x' + q(t)\,x = 0, \tag{4.27}$$

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 $x(t)$.

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 at most 1 zero. One of the questions to be discussed is how to determine or to estimate the number of zeros of solutions of (4.27).

Let us start with the following simple observation.

Lemma 4.4 *If $x(t)$ is a solution to* (4.27) *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 $I \subset \mathbb{R}$*

$$x'' + p(t)\,x' + q_1(t)\,x = 0 \ \textit{and} \ \ y'' + p(t)\,y' + q_2(t)\,y = 0,$$

where $p \in C^1(I)$, $q_1, q_2 \in C(I)$, and, for all $t \in I$,

$$q_1(t) \leq q_2(t).$$

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 that 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 $\frac{\pi}{\sqrt{q_2}} \leq \frac{\pi}{\sqrt{q_1}}$. 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 $x'' + p(t)x' + q(t)x = 0$. Then we claim that if $a < b$ are two 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 a 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 case is given by the couple $x(t) = \sin t$ and $y(t) = \cos t$.

Proof of Theorem 4.5. Making in the ODE

$$x'' + p(t)x' + q(t)x = 0$$

the change of unknown function $u(t) = x(t)\exp\left(\frac{1}{2}\int p(t)\,dt\right)$, we transform it to the form

$$u'' + Q(t)u = 0$$

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 $p \equiv 0$, which will be assumed in the sequel.

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

$$x'(b)\,y(b) - x'(a)\,y(a) = [x'y - y'x]_a^b = \int_a^b (q_2(t) - q_1(t))\,x(t)\,y(t)\,dt. \tag{4.28}$$

Since $q_2 \geq q_1$ on $[a,b]$ and $x(t)$ and $y(t)$ are non-negative on $[a,b]$, the integral in (4.28) is non-negative. On the other hand, the left hand side of (4.28) is negative because $y(a)$ and $y(b)$ are positive whereas $x'(b)$ and $-x'(a)$ are negative. This contradiction finishes the proof. ■

Corollary. *Under the conditions of Theorem 4.5, let $q_1(t) < q_2(t)$ for all $t \in I$. If $a, b \in I$ are two distinct zeros of $x(t)$ then there is a zero of $y(t)$ in the open interval (a,b).*

Proof. Indeed, as in the proof of Theorem 4.5, we can assume that a, b are consecutive zeros of x and $x(t) > 0$ in (a,b). If y has no zeros in (a,b) then we can assume that $y(t) > 0$ on (a,b) whence $y(a) \geq 0$ and $y(b) \geq 0$. The integral in the right hand side of (4.28) is positive because $q_2(t) > q_1(t)$ and $x(t), y(t)$ are positive on (a,b), while the left hand side is non-positive because $x'(b) \leq 0$ and $x'(a) \geq 0$. ■

Consider the differential operator

$$L = \frac{d^2}{dt^2} + p(t)\,\frac{d}{dt} + q(t) \tag{4.29}$$

so that the ODE (4.27) 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 (4.29):

$$\begin{cases} Lx = f(t) \\ x(a) = \alpha \\ x(b) = \beta \end{cases}$$

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

$$\lim_{k \to \infty} (t_{k+1} - t_k) = \frac{\pi}{\sqrt{b}}. \tag{4.30}$$

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

$$t_{k+1} - t_k \leq \frac{\pi}{\sqrt{m}}. \tag{4.31}$$

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 (4.31).

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 (4.30) 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

$$t^2 x'' + t x' + \left(t^2 - \alpha^2\right) x = 0 \tag{4.32}$$

where $t>0$ is an independent variable, $x=x(t)$ is the unknown function, $\alpha\in\mathbb{R}$ is a given parameter[62]. The *Bessel functions*[63] 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

$$x''+\frac{1}{t}x'+\left(1-\frac{\alpha^2}{t^2}\right)x=0, \tag{4.33}$$

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

$$Q(t)=q-\frac{p^2}{4}-\frac{p'}{2}=1-\frac{\alpha^2}{t^2}+\frac{1}{4t^2}. \tag{4.34}$$

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 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 $(0,\tau)$

$$z(t)=\ln\frac{1}{t}-\sin t$$

[62] In general, one can let α to be a complex number as well but here we restrict ourselves to the real case.

[63] 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 (4.32). If α is non-integer then J_α and $J_{-\alpha}$ are linearly independent solutions to (4.32). 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.

which is positive in $(0,\tau)$ provided τ is small enough (clearly, $z(t) \to +\infty$ as $t \to 0$). For this function we have

$$z' = -\frac{1}{t} - \cos t \quad \text{and} \quad z'' = \frac{1}{t^2} + \sin t$$

whence

$$z'' + \frac{1}{t}z' + z = \ln\frac{1}{t} - \frac{\cos t}{t}.$$

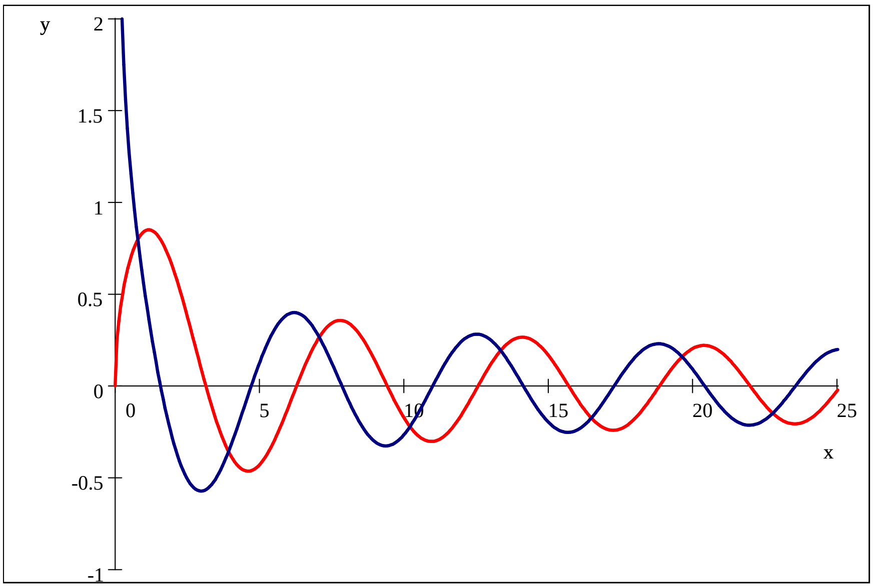
Since $\frac{\cos t}{t} \sim \frac{1}{t}$ and $\ln\frac{1}{t} = o\left(\frac{1}{t}\right)$ as $t \to 0$, we see that the right hand side here is negative in $(0,\tau)$ provided τ is small enough. It follows that

$$z'' + \frac{1}{t}z' + \left(1 - \frac{\alpha^2}{t^2}\right)z < 0, \tag{4.35}$$

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 (4.34) $Q(t) \equiv 1$ 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 (4.35) 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 accumulating at 0 (see Exercise 64).