Ordinary Differential Equation

Alexander Grigorian
University of Bielefeld

Lecture Notes, April - July 2008

Contents

1 Introduction: the notion of ODEs and examples

A differential equation (*Differentialgleichung*) is an equation for an unknown function that contains not only the function but also its derivatives (*Ableitung*). In general, the unknown function may depend on several variables and the equation may include various partial derivatives. However, in this course we consider only the differential equations for a function of a *single* real variable. Such equations are called *ordinary differential equations*[1] – shortly ODE (*die gewöhnliche Differentialgleichungen*).

A most general ODE has the form

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

where F is a given function of $n+2$ variables and $y = y(x)$ is an unknown function of a real variable x. The maximal order n of the derivative $y^{(n)}$ in (1.1) is called the *order* of the ODE.

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 when one can find all the solutions.

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* (*Definitionsbereich*) 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[2] $I \subset \mathbb{R}$, is called a (*particular*) solution of (1.2) if $y(x)$ is differentiable at any $x \in I$, the point $(x, y(x))$ belongs to D for any $x \in I$ and the identity $y'(x) = f(x, y(x))$ holds for all $x \in I$.

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

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.

[1] The theory of *partial* differential equations, that is, the equations containing partial derivatives, is a topic of another lecture course.

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

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

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

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*[3] of f. Assuming that f is a continuous (*stetig*) 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

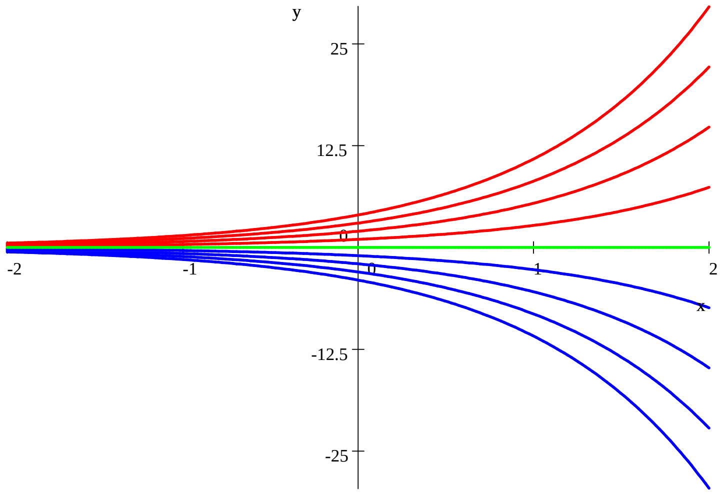
$$y = -C_1 e^x,$$

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:

[3] By definition, a primitive function of f is any function whose derivative is equal to f.



Let us show that the family of solutions $y = Ce^x$, $C \in \mathbb{R}$, is the general solution. Indeed, if $y(x)$ is a solution that takes positive value somewhere then it is positive in some open interval, say I. By the above argument, $y(x) = Ce^x$ in I, where $C > 0$. Since $e^x \neq 0$, this solution does not vanish also at the endpoints of I. This implies that the solution must be positive on the whole interval where it is defined. It follows that $y(x) = Ce^x$ in the domain of $y(x)$. The same applies if $y(x) < 0$ for some x.

Hence, the general solution of the ODE $y' = y$ is $y(x) = Ce^x$ where $C \in \mathbb{R}$. The constant C is referred to as a *parameter*. It is clear that the particular solutions are distinguished by the values of the parameter.

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*

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

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 $f(x) = 1$ and $g(y) = y \neq 0$ so that Theorem 1.1 applies. We have

$$F(x) = \int f(x)\, dx = \int dx = x$$

and

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

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 $y = C_1 e^x$ as in the previous example. Note that Theorem 1.1 does not cover the case when $g(y)$ may vanish, which must be analyzed separately when needed.

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

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

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

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

Hence, the equation (1.3) is equivalent to

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

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 $y(x)$ is differentiable. Since the function $g(y)$ does not vanish, it is either positive or negative in the whole domain. Then the function $G(y)$, 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

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

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* $y(x)$ *defined by (1.7) satisfies the condition* $y(x_0) = y_0$.

The condition $y(x_0) = y_0$ is called the *initial condition* ($Anfangsbedingung$).

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, $C = G(y_0) - F(x_0)$. Conversely, assume that C is given by this formula and prove that it determines by (1.7) a solution $y(x)$. If the right hand side of (1.7) is defined on an interval containing x_0, then by Theorem 1.1 it defines a solution $y(x)$, and this solution satisfies $y(x_0) = y_0$ by the choice of C. We only have to make sure that the domain of the right hand side of (1.7) contains an interval around 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 $y(x)$ 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. However, the meaning of the uniqueness claim in this form is a bit ambiguous because out of any solution $y(x)$, one can make another solution just by slightly reducing the domain, and if the reduced domain still contains x_0 then the initial condition will be satisfied also by the new solution. The precise uniqueness claim means that any two solutions satisfying the same initial condition, coincide on the intersection of their domains; also, such solutions correspond to the same value of the parameter C.

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

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

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

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

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

If in the 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, then we have a family of constant solutions $y(x) = y_k$. The method of separation of variables provides solutions in any domain $y_k < y < y_{k+1}$. The integral curves in the domains $y_k < y < y_{k+1}$ can in general touch the constant solutions, as will be shown in the next example.

Example. Consider the equation

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

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

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

whence

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

and

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

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

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

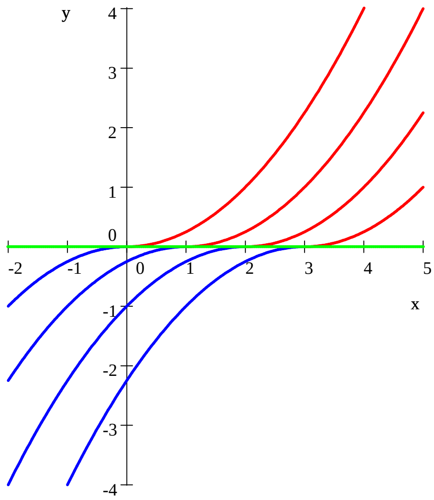
whence

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

and

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

We obtain the following integrals curves:



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

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

Note that such solutions are not obtained automatically by the method of separation of variables. It follows that 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

$$y'+a(x)y=b(x) \tag{1.9}$$

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* (1.9) *has the form*

$$y(x) = e^{-A(x)} \int b(x) e^{A(x)} dx, \tag{1.10}$$

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

Note that the function $y(x)$ given by (1.10) 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,

$$y(x) = u(x) e^{-A(x)}. \tag{1.11}$$

Substituting this to the equation (1.9) 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 (1.11), we finish the proof. ■

One may wonder how one can guess to make the change (1.11). Here is the motivation. Consider first the case when $b(x) \equiv 0$. In this case, the equation (1.9) 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 (1.9) 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

$$y' + \frac{1}{x} y = e^{x^2} \tag{1.12}$$

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 (1.12):

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

Hence,

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

where C_0 is an arbitrary constant.

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

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

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 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,

$$dF = adx + bdy. \tag{1.13}$$

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 (1.13) 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

$$dF = adx + bdy, \tag{1.14}$$

and *inexact* otherwise. If the form is exact then the function F from (1.14) 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*

$$a_y = b_x.$$

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

Example. The form $ydx - xdy$ is inexact because $a_y = 1$ while $b_x = -1$.

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

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

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

$$a(x, y) + b(x, y)\, y' = 0. \tag{1.15}$$

This ODE 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 (1.15) in the form

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

which explains why the equation (1.15) is related to the differential form $adx + bdy$. We say that the equation (1.15) is exact if the form $adx + bdy$ is exact.

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. Consider a differentiable function $y(x)$ defined on an interval $I \subset \mathbb{R}$ such that the graph of y is contained in Ω. Then y solves the equation* (1.15) *if and only if*

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

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

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

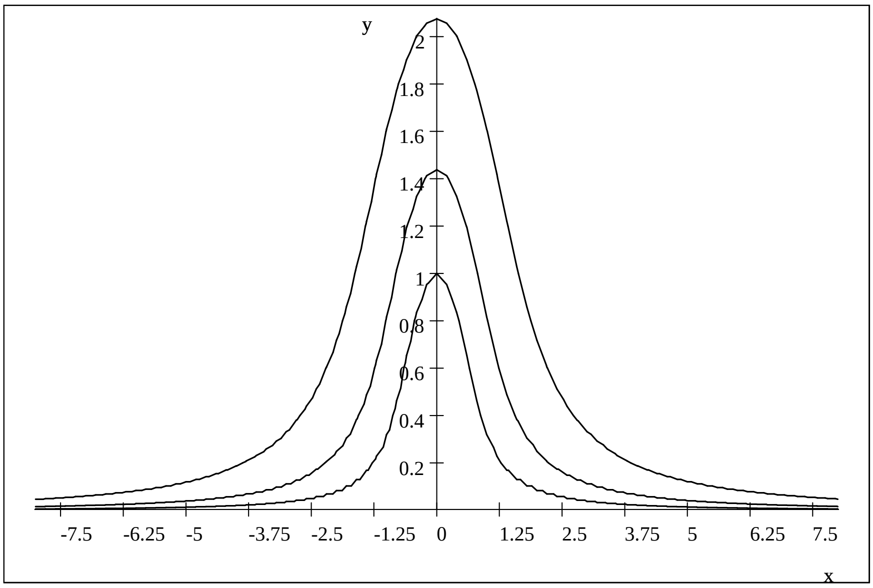
Hence, the equation $a + by' = 0$ is equivalent to $\frac{d}{dx} F(x, y(x)) = 0$, and the latter is equivalent to $F(x, y(x)) = \text{const}$. ■

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

The equation $2xy + (x^2 + y^2)\, y' = 0$ is exact and is equivalent to

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

Below are some integral curves of this equation:



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

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

Theorem 1.5 (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.*

Proof of Theorem 1.5. Assume first that the integral F exists and $F(x_0, y_0) = 0$ for some point $(x_0, y_0) \in \Omega$ (the latter can always be achieved by adding a constant to F). For any point $(x, y) \in \Omega$, also the point $(x, y_0) \in \Omega$; moreover, the intervals $[(x_0, y_0), (x, y_0)]$ and $[(x, y_0), (x, y)]$ are contained in Ω because Ω is a rectangle. Since $F_x = a$ and $F_y = b$, we obtain by the fundamental theorem of calculus that

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

and

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

whence

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

Now use the formula (1.16) to *define* function $F(x, y)$. Let us show that F is indeed the integral of the form $adx + bdy$. Since a and b are continuous, it suffices to verify that

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

It is easy to see from (1.16) that

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

Next, we have

$$\begin{aligned} F_x &= \frac{\partial}{\partial x} \int_{x_0}^{x} a(s,y_0)\,ds + \frac{\partial}{\partial x} \int_{y_0}^{y} b(x,t)\,dt \\ &= a(x,y_0) + \int_{y_0}^{y} \frac{\partial}{\partial x} b(x,t)\,dt. \end{aligned} \tag{1.17}$$

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 (1.17)

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

which finishes the proof. ■

Now we prove the lemma, which is needed to justify (1.17).

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 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 $x \in I$,

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

which amounts to

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

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

$$\frac{g(x',t) - g(x,t)}{x' - x} \to g_x(x,t) \text{ as } x' \to x. \tag{1.18}$$

Consider all parts of (1.18) as functions of t, with fixed x and with x' as a parameter. Then we have a convergence of a sequence of functions, and we would like to deduce that their integrals converge as well. By a result from Analysis II, this is the case, if the convergence is *uniform* (*gleichmässig*) in the whole interval $[\alpha,\beta]$, that is, if

$$\sup_{t\in[\alpha,\beta]}\left|\frac{g(x',t)-g(x,t)}{x'-x}-g_x(x,t)\right|\to 0 \quad \text{as } x'\to x. \tag{1.19}$$

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

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

Hence, the difference quotient in (1.19) 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

$$x,\xi\in I,|x-\xi|<\delta \text{ and } t,s\in J,|t-s|<\delta \ \Rightarrow\ |g_x(x,t)-g_x(\xi,s)|<\varepsilon. \tag{1.20}$$

If $|x-x'|<\delta$ then also $|x-\xi|<\delta$ and, by (1.20) with $s=t$,

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

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

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

whence (1.19) 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 (1.16) taking $x_0=y_0=0$:

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

as it was observed above.

Example. Consider the differential form

$$\frac{-ydx+xdy}{x^2+y^2} \tag{1.21}$$

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 domain in Ω. However, let us show that the form is inexact in Ω.

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

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

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'

$$d\theta = \frac{-ydx + xdy}{x^2+y^2}. \tag{1.22}$$

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 (1.22) 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 (1.22) 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 (1.22) follows again. Since Ω' is the union of the three half-planes $\{x > 0\}$, $\{y > 0\}$, $\{y < 0\}$, we conclude that (1.22) holds in Ω' and, hence, the form (1.21) is exact in Ω'.

Why the form (1.21) is inexact in Ω? Assume from the contrary that the form (1.21) 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[4] $F = \theta + \text{const}$ in Ω'. It follows from this identity that function θ can be extended from Ω' to a continuous

[4] 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.

function on Ω, which however is not true, because the limits of θ when approaching the point $(-1,0)$ (or any other point $(x,0)$ with $x<0$) from above and below are different.

The moral of this example is that the statement of Theorem 1.5 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.

1.4 Integrating factor

Consider again the quasilinear equation

$$a\,(x,y)+b\,(x,y)\,y'=0 \tag{1.23}$$

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 (1.23) 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 (1.23) 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 (1.16) $x_0=y_0=1$, 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

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

1.5 Second order ODE

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

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

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

1.5.1 Newtons' second law

Consider movement of a point particle along a straight line and let its coordinate at time t be $x(t)$. The velocity (*Geschwindigkeit*) of the particle is $v(t) = x'(t)$ and the acceleration (*Beschleunigung*) is $a(t) = x''(t)$. The Newton's second law says that at any time

$$mx'' = F, \tag{1.24}$$

where m is the mass of the particle and F is the force ($Kraft$) acting on the particle. In general, F is a function of t, x, x' so that (1.24) 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, conservative are gravitation force, spring force, electrostatic force, while friction and the air resistance are non-conservative as they depend in the velocity v. Assuming $F = F(x)$, denote by $U(x)$ a primitive function of $-F(x)$. The function U is called the *potential* of the force F. Multiplying the equation (1.24) by x' and integrating in t, we obtain

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

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

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

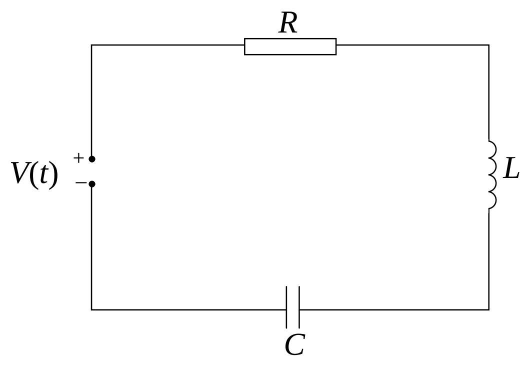
and

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

The sum $\frac{mv^2}{2} + U(x)$ is called the *total 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*: the total energy of the particle in a conservative field remains constant.

1.5.2 Electrical circuit

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



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

$$v_R = RI$$

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

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

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

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

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

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

whence

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

Differentiating in t, we obtain

$$LI'' + RI' + \frac{I}{C} = V', \tag{1.25}$$

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

2 Existence and uniqueness theorems

2.1 1st order ODE

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

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

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

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

$$\begin{cases} x' = f(t, x), \\ x(t_0) = x_0. \end{cases} \tag{2.1}$$

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

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

Definition. We say that a function $f : \Omega \to \mathbb{R}$ is *Lipschitz* in x if there is a constant L such that

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

for all t, x, y such that $(t, x) \in \Omega$ and $(t, y) \in \Omega$. The constant L is called the *Lipschitz constant* of f in Ω.

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

$$R = [t_0 - \delta, t_0 + \delta] \times [x_0 - \varepsilon, x_0 + \varepsilon] \tag{2.2}$$

is contained in Ω and the function f is Lipschitz in R; that is, there is a constant L such that for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in [x_0 - \varepsilon, x_0 + \varepsilon]$,

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

Note that in the latter case the constant L may be different for different rectangles.

Lemma 2.1 (*a*) *If the partial derivative f_x exists and is bounded in a rectangle $R \subset \mathbb{R}^2$ then f is Lipschitz in x in R.*

(*b*) *If the partial derivative f_x exists and is continuous in an open set $\Omega \subset \mathbb{R}^2$ then f is locally Lipschitz in x in Ω.*

Proof. (*a*) If (t, x) and (t, y) belong to R then the whole interval between these points is also in R, and we have by the mean value theorem

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

for some $\xi \in [x, y]$. By hypothesis, f_x is bounded in R, that is,

$$L := \sup_R |f_x| < \infty, \tag{2.3}$$

whence we obtain

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

Hence, f is Lipschitz in R with the Lipschitz constant (2.3).

(*b*) Fix a point $(t_0, x_0) \in \Omega$ and choose positive ε, δ so small that the rectangle R defined by (2.2) is contained in Ω (which is possible because Ω is an open set). Since R is a bounded closed set, the continuous function f_x is bounded on R. By part (*a*) we conclude that f is Lipschitz in R, which means that f is locally Lipschitz in Ω. ∎

Example. The function $f(t,x) = |x|$ is Lipschitz in x in $\mathbb{R}^2$ because

$$||x| - |y|| \le |x - y|,$$

by the triangle inequality for $|x|$. Clearly, f is not differentiable in x at $x = 0$. Hence, the continuous differentiability of f is sufficient for f to be Lipschitz in x but not necessary.

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

Theorem 2.2 (The Picard - Lindelöf theorem) *Let Ω be an open set in $\mathbb{R}^2$ and $f(t,x)$ be a continuous function in Ω that is locally Lipschitz in x.*

(Existence) Then, for any point $(t_0, x_0) \in \Omega$, the initial value problem IVP (2.1) *has a solution.*

(Uniqueness) If $x_1(t)$ and $x_2(t)$ are two solutions of the same IVP then $x_1(t) = x_2(t)$ in their common domain.

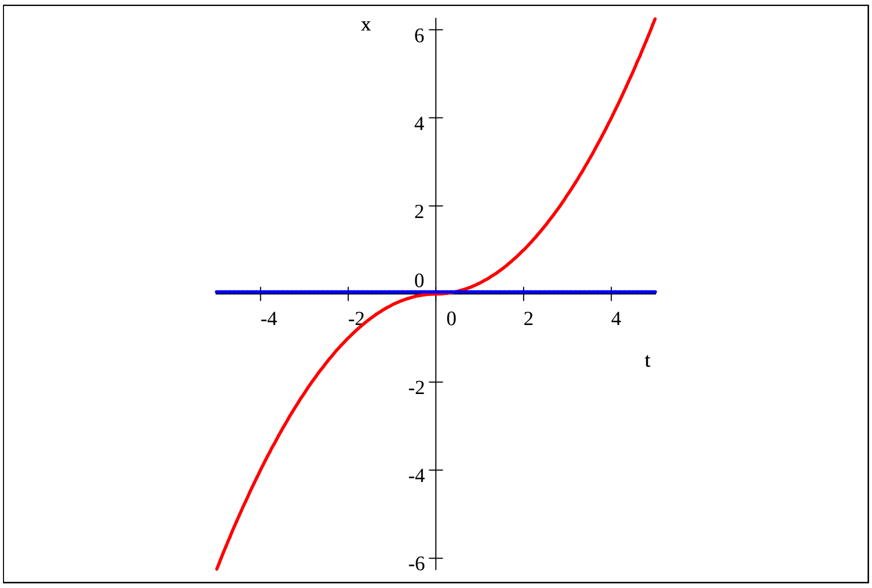
Remark. By Lemma 2.1, the hypothesis of Theorem 2.2 that f is locally Lipschitz in x could be replaced by a simpler hypotheses that f_x is continuous. However, as we have seen above, there are examples of functions that are Lipschitz but not differentiable, and Theorem 2.2 applies for such functions.

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

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

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

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



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

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

Claim. *Let $x(t)$ be a function defined on an open interval $I \subset \mathbb{R}$. A function $x(t)$ solves IVP if and only if $x(t)$ is continuous, $(t, x(t)) \in \Omega$ for all $t \in I$, $t_0 \in I$, and*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{2.4}$$

Indeed, if x solves IVP then (2.4) follows from $x' = f(t, x(t))$ just by integration:

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

whence

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

Conversely, if x is a continuous function that satisfies (2.4) then the right hand side of (2.4) is differentiable in t whence it follows that $x(t)$ is differentiable. It is trivial that $x(t_0) = x_0$, and after differentiation (2.4) we obtain the ODE $x' = f(t, x)$.

This claim reduces the problem of solving IVP to the integral equation (2.4). Fix a point $(t_0, x_0) \in \Omega$ and let ε, δ be the parameter from the the local Lipschitz condition at this point; that is, there is a constant L such that

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

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

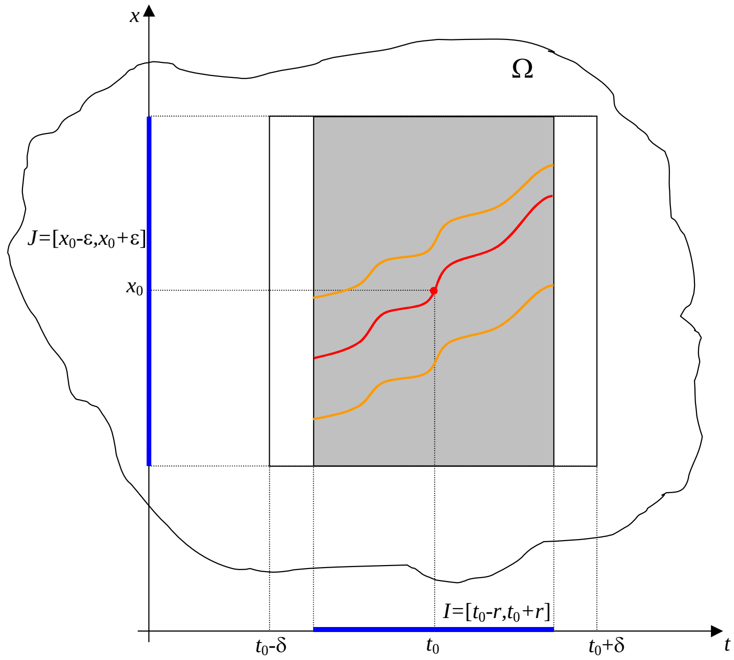
$$J = [x_0 - \varepsilon, x_0 + \varepsilon] \text{ and } I = [t_0 - r, t_0 + r],$$

were $0 < r \leq \delta$ is a new parameter, whose value will be specified later on. By construction, $I \times J \subset \Omega$.

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

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

(see the diagram below).



Consider the integral operator A defined on functions $x \in X$ by

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

which is obviously motivated by (2.4). To be more precise, we would like to ensure that $x \in X$ implies $Ax \in X$. Note that, for any $x \in X$, the point $(s, x(s))$ belongs to Ω so that the above integral makes sense and the function Ax is defined on I. This function is obviously continuous. We are left to verify that the image of Ax is contained in J. Indeed, the latter condition means that

$$|Ax(t) - x_0| \leq \varepsilon \text{ for all } t \in I. \tag{2.5}$$

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

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

where

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

Hence, if r is so small that $Mr \leq \varepsilon$ then (2.5) is satisfied and, hence, $Ax \in X$.

To summarize the above argument, we have defined a function family X and a mapping $A : X \to X$. By the above Claim, a function $x \in X$ will solve the IVP if function x is a *fixed point* of the mapping A, that is, if $x = Ax$.

The existence of a fixed point will be obtained using the Banach fixed point theorem: If (X, d) is a complete metric space (*Vollständiger metrische Raum*) and $A : X \to X$ is a contraction mapping (*Kontraktionsabbildung*), that is,

$$d(Ax, Ay) \leq q d(x, y)$$

for *some* $q \in (0, 1)$ and all $x, y \in X$, then A has a fixed point. By the proof of this theorem, one starts with any element $x_0 \in X$, constructs a sequence of iteration $\{x_n\}_{n=1}^{\infty}$ using the rule $x_{n+1} = Ax_n$, $n = 0, 1, ...$, and shows that the sequence $\{x_n\}_{n=1}^{\infty}$ converges in X to a fixed point.

In order to be able to apply this theorem, we must introduce a distance function d (*Abstand*) on X so that (X, d) is a complete metric space and A is a contraction mapping with respect to this distance.

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

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

Using the fact that the convergence in (X, d) is the uniform convergence of functions and the uniform limits of continuous functions is continuous, one can show that the metric space (X, d) is complete (see Exercise 16).

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

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

Therefore,

$$\sup_{t \in I} |Ax(t) - Ay(t)| \leq \sup_{s \in I} |x(s) - y(s)| L |t - t_0|$$

whence

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

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

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

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

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

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

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

$$r\le\delta,\ r\le\frac{\varepsilon}{M},\ r<\frac{1}{L}. \tag{2.6}$$

For some applications, it is important that r can be determined as a function of ε,δ,M,L.

Example. The method of the proof of the existence in Theorem 2.2 suggests the following procedure of computation of the solution of IVP. We start with any function $x_0\in X$ (using 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 $x_0(t)\equiv1$, we obtain

$$x_1(t)=1+\int_0^t x_0ds=1+t,$$

$$x_2(t)=1+\int_0^t x_1ds=1+t+\frac{t^2}{2}$$

$$x_3(t)=1+\int_0^t x_2dt=1+t+\frac{t^2}{2!}+\frac{t^3}{3!}$$

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.

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

Lemma 2.3 (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 $C,L\ge0$ such that*

$$z(t)\le C+L\int_{t_0}^t z(s)\,ds \tag{2.7}$$

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

$$z(t)\le C\exp(L(t-t_0)) \tag{2.8}$$

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

Proof. We can assume that C is strictly positive. Indeed, if (2.7) holds with $C=0$ then it holds with any $C>0$. Therefore, (2.8) 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 (2.7) 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 (2.7) that $z\le F$ whence

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

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

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

whence by integration

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

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

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

Using again (2.7), that is, $z\le F$, we obtain (2.8). ■

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

Any set U that satisfies the conclusion of Lemma 2.4 is called *connected* (*zusammenhängend*). Hence, Lemma 2.4 says that any interval is a connected set.

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

Proof of the uniqueness in Theorem 2.2. Assume that $x_1(t)$ and $x_2(t)$ are two solutions of the same IVP both defined on an open interval $U\subset\mathbb{R}$ and prove that they coincide on U.

We first prove that the two solution coincide in some interval around t_0. Let ε and δ be the parameters from the Lipschitz condition at the point (t_0,x_0) as above. Choose

$0 < r < \delta$ so small that the both functions $x_1(t)$ and $x_2(t)$ restricted to $I = [t_0 - r, t_0 + r]$ take values in $J = [x_0 - \varepsilon, x_0 + \varepsilon]$ (which is possible because both $x_1(t)$ and $x_2(t)$ are continuous functions). As in the proof of the existence, the both solutions satisfies the integral identity

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

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

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

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

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

whence

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

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

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

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

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

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

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

2.2 Dependence on the initial value

Consider the IVP

$$\begin{cases} x' = f(t, x) \\ x(t_0) = s \end{cases} \tag{2.9}$$

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

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

$$R = [t_0 - \delta, t_0 + \delta] \times [x_0 - \varepsilon, x_0 + \varepsilon]$$

is contained in Ω and, for all $(t,x),(t,y)\in R$,

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

Let M be the supremum of $|f(t,x)|$ in R. By the proof of Theorem 2.2, the solution $x(t)$ with the initial condition $x(t_0)=x_0$ is defined in the interval $[t_0-r,t_0+r]$ where r is any positive number that satisfies (2.6), and $x(t)$ takes values in $[x_0-\varepsilon,x_0+\varepsilon]$ for all $t\in[t_0-r,t_0+r]$. Let us choose r as follows

$$r=\min\left(\delta,\frac{\varepsilon}{M},\frac{1}{2L}\right). \tag{2.10}$$

For what follows, it is only important that r can be determined as a function of ε,δ,L,M.

Now consider the IVP with the condition $x(t_0)=s$ where s is close enough to x_0, say

$$s\in[x_0-\varepsilon/2,x_0+\varepsilon/2]. \tag{2.11}$$

Then the rectangle

$$R'=[t_0-\delta,t_0+\delta]\times[s-\varepsilon/2,s+\varepsilon/2]$$

is contained in R. Therefore, the Lipschitz condition holds in R' also with constant L and $\sup_{R'}|f|\le M$. Hence, the solution $x(t,s)$ with the initial condition $x(t_0)=s$ is defined in $[t_0-r(s),t_0+r(s)]$ and takes values in $[s-\varepsilon/2,s+\varepsilon/2]\subset[x_0-\varepsilon,x_0+\varepsilon]$ provided

$$r(s)\le\min\left(\delta,\frac{\varepsilon}{2M},\frac{1}{2L}\right) \tag{2.12}$$

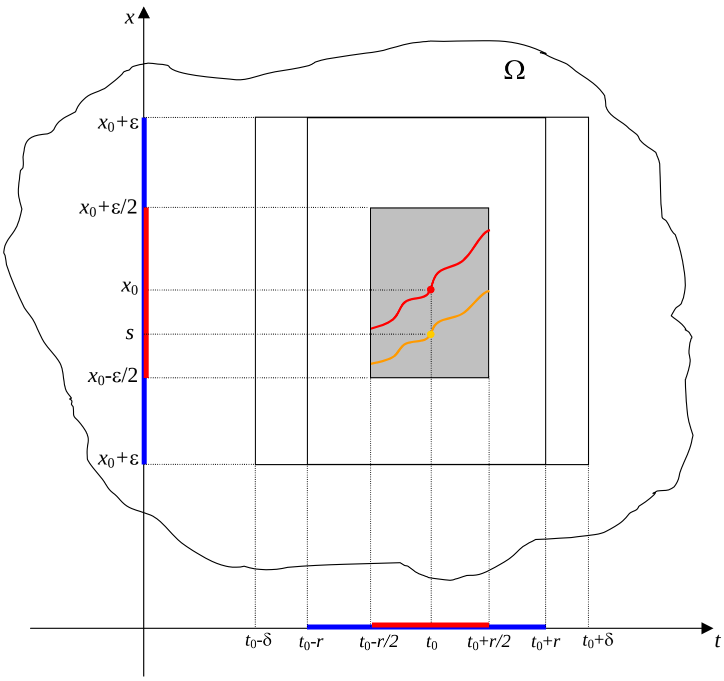
(in comparison with (2.10), here ε is replaced by $\varepsilon/2$ in accordance with the definition of R'). Clearly, if r satisfies (2.10) then the value

$$r(s)=\frac{r}{2}$$

satisfies (2.12). Let us state the result of this argument as follows.

Claim. *Fix a point $(t_0,x_0)\in\Omega$ and choose $\varepsilon,\delta>0$ from the local Lipschitz condition at (t_0,x_0). Let L be the Lipschitz constant in $R=[t_0-\delta,t_0+\delta]\times[x_0-\varepsilon,x_0+\varepsilon]$, $M=\sup\limits_R|f|$, and define $r=r(\varepsilon,\delta,L,M)$ by (2.10). Then, for any $s\in[x_0-\varepsilon/2,x_0+\varepsilon/2]$, the solution $x(t,s)$ of (2.9) is defined in $[t_0-r/2,t_0+r/2]$ and takes values in $[x_0-\varepsilon,x_0+\varepsilon]$.*

In particular, we can compare solutions with different initial value s since they have the common domain $[t_0-r/2,t_0+r/2]$ (see the diagram below).



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

$$|x(t,s') - x(t,s'')| \leq 2|s'-s''|. \tag{2.13}$$

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

Proof. Consider again the integral equations

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

and

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

It follows that, for all $t \in [t_0, t_0+r/2]$,

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

where we have used the Lipschitz condition because by the above Claim $(\tau, x(\tau,s)) \in [t_0-\delta, t_0+\delta] \times [x_0-\varepsilon, x_0+\varepsilon]$ for all $s \in [x_0-\varepsilon/2, x_0+\varepsilon/2]$.

Setting $z(t) = |x(t,s') - x(t,s'')|$ we obtain

$$z(t) \leq |s'-s''| + L\int_{t_0}^{t} z(\tau)\,d\tau,$$

which implies by the Lemma 2.3

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

Since $t - t_0 \leq r/2$ and by (2.10) $L \leq \frac{1}{2r}$ we see that $L(t - t_0) \leq \frac{1}{4}$ and

$$\exp(L(t - t_0)) \leq e^{1/4} < 2,$$

which proves (2.13) for $t \geq t_0$. Similarly one obtains the same for $t \leq t_0$.

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

$$x(t_n, s_n) \to x(t, s)$$

if $(t_n, s_n) \to (t, s)$ as $n \to \infty$. Then by (2.13)

$$\begin{aligned} |x(t_n, s_n) - x(t, s)| &\leq |x(t_n, s_n) - x(t_n, s)| + |x(t_n, s) - x(t, s)| \\ &\leq 2|s_n - s| + |x(t_n, s) - x(t, s)|, \end{aligned}$$

and this goes to 0 as $n \to \infty$ by the continuity of $x(t, s)$ in t for a fixed s. ∎

Remark. The same argument shows that if a function $x(t, s)$ is continuous in t for any fixed s and *uniformly* continuous in s, then $x(t, s)$ is jointly continuous in (t, s).

2.3 Higher order ODE and reduction to the first order system

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

$$y^{(n)} = F\left(t, y, ..., y^{(n-1)}\right), \tag{2.14}$$

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

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

$$x' = (x'_1, x'_2, ..., x'_n).$$

Consider now a *vector ODE of the first order*

$$x' = f(t, x) \tag{2.15}$$

where f is a given function of $n+1$ variables, which takes values in $\mathbb{R}^n$, that is, $f : \Omega \to \mathbb{R}^n$ where Ω is an open subset of $\mathbb{R}^{n+1}$ (so that the couple (t, x) is considered as a point in Ω). Denoting by f_k the components of f, we can rewrite the vector equation (2.15) as a system of n scalar equations

$$\begin{cases} x'_1 = f_1(t, x_1, ..., x_n) \\ ... \\ x'_k = f_k(t, x_1, ..., x_n) \\ ... \\ x'_n = f_n(t, x_1, ..., x_n) \end{cases} \tag{2.16}$$

A system of ODEs of the form (2.15) is called the *normal system*.

Let us show how the equation (2.14) can be reduced to the normal system (2.16). Indeed, with any function $y(t)$ let us associate the vector-function

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

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

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

Obviously,

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

and using (2.14) we obtain a system of equations

$$\begin{cases} x_1' = x_2 \\ x_2' = x_3 \\ ... \\ x_{n-1}' = x_n \\ x_n' = F(t, x_1, ...x_n) \end{cases} \tag{2.17}$$

Obviously, we can rewrite this system as a vector equation (2.15) where

$$f(t, x) = (x_2, x_3, ..., x_n, F(t, x_1, ..., x_n)). \tag{2.18}$$

Conversely, the system (2.17) implies

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

so that we obtain equation (2.14) with respect to $y = x_1$. Hence, the equation (2.14) is equivalent to the vector equation (2.15) with function f defined by (2.18).

Example. For example, consider the second order equation

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

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

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

whence

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

Hence, we obtain the normal system (2.15) with

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

What initial value problem is associated with the vector equation (2.15) and the scalar higher order equation (2.14)? Motivated by the study of the 1st order ODE, one can presume that it makes sense to consider the following IVP for the vector 1st order ODE

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

where $x_0 \in \mathbb{R}^n$ is a given initial value of $x(t)$. For the equation (2.14), 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

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

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

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

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

Recall that a *norm* in $\mathbb{R}^n$ is a function $N : \mathbb{R}^n \to \mathbb{R}$ with the following properties:

1. $N(x) \geq 0$ for all $x \in \mathbb{R}^n$ and $N(x) = 0$ 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 $x, y \in \mathbb{R}^n$.

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

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

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

$$C'' \leq \frac{N_1(x)}{N_2(x)} \leq C' \text{ for all } x \neq 0. \tag{2.19}$$

For example, it follows from the definitions of $\|x\|_1$ and $\|x\|_\infty$ that

$$1 \leq \frac{\|x\|_1}{\|x\|_\infty} \leq n.$$

For most applications, the relation (2.19) means that the choice of a specific norm is not important.

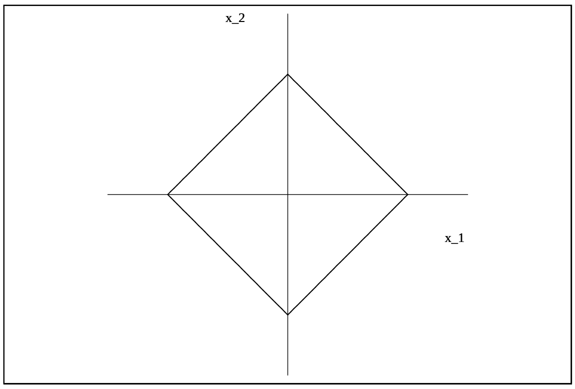
The notion of a norm is used in order to define the Lipschitz condition for functions in $\mathbb{R}^n$. Let us fix some norm $\|x\|$ in $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$ and $r > 0$, and define a *closed ball* $\overline{B}(x,r)$ by

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

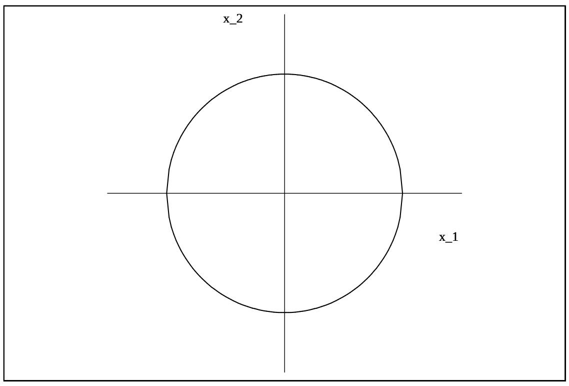
For example, in $\mathbb{R}$ with $\|x\| = |x|$ we have $\overline{B}(x,r) = [x-r, x+r]$. Similarly, one defines an *open ball*

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

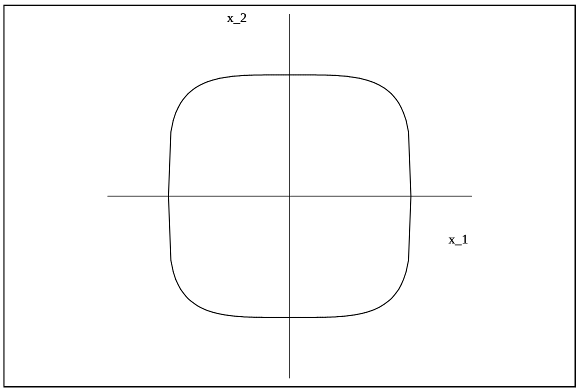
Below are sketches of the ball $B(0,1)$ in $\mathbb{R}^2$ for different norms:
the 1-norm:



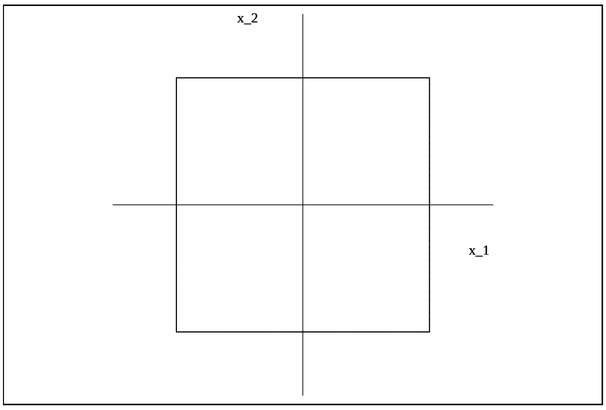
the 2-norm (a round ball):



the 4-norm:



the ∞-norm (a box):



2.5 Existence and uniqueness for a system of ODEs

Let Ω be an open subset of $\mathbb{R}^{n+1}$ and $f = f(t,x)$ be a mapping from Ω to $\mathbb{R}^n$. Fix a norm $\|x\|$ in $\mathbb{R}^n$.

Definition. Function $f(t,x)$ is called Lipschitz in x in Ω if there is a constant L such that for all $(t,x),(t,y) \in \Omega$

$$\|f(t,x) - f(t,y)\| \le L\|x-y\|. \tag{2.20}$$

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 K of $\mathbb{R}^{n+1}$ will be called a *cylinder* if it has the form $K = I \times B$ 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

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

is contained in Ω and f is Lipschitz in x in K.

Lemma 2.6 (a) *If all components f_k of f are differentiable functions in a cylinder K and all the partial derivatives $\frac{\partial f_k}{\partial x_i}$ are bounded in K then the function $f(t,x)$ is Lipschitz in x in K.*

(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

$$g(y)-g(x)=\sum_{j=1}^{n}\frac{\partial g}{\partial x_j}(\xi)(y_j-x_j) \tag{2.21}$$

(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 (2.21).

(a) Let $K=I\times B$ where I is an interval in $\mathbb{R}$ and B is a ball in $\mathbb{R}^n$. If $(t,x),(t,y)\in K$ 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

$$f_k(t,x)-f_k(t,y)=\sum_{j=1}^{n}\frac{\partial f_k}{\partial x_j}(t,\xi)(x_j-y_j), \tag{2.22}$$

where ξ is a point in the interval $[x,y]\subset B$. Set

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

and note that by the hypothesis $C<\infty$. Hence, by (2.22)

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

Taking max in k, we obtain

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

Switching 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 K.

(b) Given a point $(t_0, x_0) \in \Omega$, choose positive ε and δ so that the cylinder

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

is contained in Ω, which is possible by the openness of Ω. Since the components f_k are continuously differentiable, they are differentiable. Since K is a closed bounded set and the partial derivatives $\frac{\partial f_k}{\partial x_j}$ are continuous, they are bounded on K. By part (a) we conclude that f is Lipschitz in x in K, which finishes the proof. ■

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

$$\begin{cases} x' = f(t, x), \\ x(t_0) = x_0, \end{cases} \tag{2.23}$$

where (t_0, x_0) is a given point in Ω. A function $x(t) : I \to \mathbb{R}^n$ is called a solution (2.23) 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 ODE $x' = f(t, x)$ in I and the initial condition $x(t_0) = x_0$.

The graph of function $x(t)$, that is, the set of points $(t, x(t))$, is hence a curve in Ω that goes through the point (t_0, x_0). It is also called the integral curve of the ODE $x' = f(t, x)$.

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

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

where $f : \Omega \to \mathbb{R}^n$ is a mapping from an open set $\Omega \subset \mathbb{R}^{n+1}$ to $\mathbb{R}^n$. Assume that f is continuous on Ω and locally Lipschitz in x. Then, for any point $(t_0, x_0) \in \Omega$, the initial value problem IVP (2.23) *has a solution.*

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

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

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

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds. \tag{2.24}$$

Here the integral of the vector valued function is understood component-wise. If x solves IVP then (2.24) follows from $x_k' = f_k(t, x(t))$ just by integration:

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

whence

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

and (2.24) follows. Conversely, if x is a continuous function that satisfies (2.24) then

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

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

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

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

for all $t \in [t_0 - \delta, t_0 + \delta]$ and $x, y \in \overline{B}(x_0, \varepsilon)$. Choose some $r \in (0, \delta]$ to be specified later on, and set

$$I = [t_0 - r, t_0 + r] \quad \text{and } J = \overline{B}(x_0, \varepsilon).$$

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

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

Consider the integral operator A defined on functions $x(t)$ by

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

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

$$\|Ax(t) - x_0\| \le \varepsilon \text{ for all } t \in I. \tag{2.25}$$

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

$$\begin{aligned} \|Ax(t) - x_0\| &= \left\| \int_{t_0}^{t} f(s, x(s))\, ds \right\| \\ &\le \int_{t_0}^{t} \|f(s, x(s))\|\, ds \quad \text{(see Exercise 15)} \\ &\le \sup_{s \in I, x \in J} \|f(s, x)\| \, |t - t_0| \le Mr, \end{aligned}$$

where

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

Hence, if r is so small that $Mr \le \varepsilon$ then (2.5) is satisfied and, hence, $Ax \in X$.

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

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

Then (X, d) is a complete metric space (see Exercise 16).

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 around t_0. Let ε and δ be the parameters from the Lipschitz condition at the point (t_0, x_0) as above. Choose $0 < r < \delta$ so small that the both functions $x(t)$ and $y(t)$ restricted to $I = [t_0 - r, t_0 + r]$ take values in $J = \overline{B}(x_0, \varepsilon)$ (which is possible because both $x(t)$ and $y(t)$ are continuous functions). As in the proof of the existence, the both solutions satisfies the integral identity

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

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

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

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

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

whence

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

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

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

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

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

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

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

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

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

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

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

Let

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

and choose any positive r to satisfy

$$r \leq \delta, \; r \leq \frac{\varepsilon}{M}, \; r < \frac{1}{L}. \tag{2.26}$$

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

The fact that the domain of the solution admits the explicit estimates (2.26) can be used as follows.

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

$$\begin{cases} x' = f(t, x), \\ x(t_1) = x_1, \end{cases} \tag{2.27}$$

has a solution $x(t)$ which is defined for all $t \in [t_0 - r/2, t_0 + r/2]$ and takes values in $\overline{B}(x_0, \varepsilon)$.

Proof. Let $\varepsilon, \delta, L, M$ be as in the proof of Theorem 2.7. 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, the IVP (2.27) has solution $x(t)$ in the interval $[t_1 - r, t_1 + r]$, and $x(t)$ takes values in $\overline{B}(x_1, \varepsilon/2) \subset \overline{B}(x, \varepsilon)$ provided

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

For example, take

$$r = \min\left(\frac{\delta}{2}, \frac{\varepsilon}{2M}, \frac{1}{2L}\right).$$

If $t_1 \in [t_0 - r/2, t_0 + r/2]$ then $[t_0 - r/2, t_0 + r/2] \subset [t_1 - r, t_1 + r]$ so that the solution $x(t)$ of (2.27) is defined on $[t_0 - r/2, t_0 + r/2]$ and takes value in $\overline{B}(x, \varepsilon)$, which was to be proved. ■

2.6 Maximal solutions

Consider again the ODE

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

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

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

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

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

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

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

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

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

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

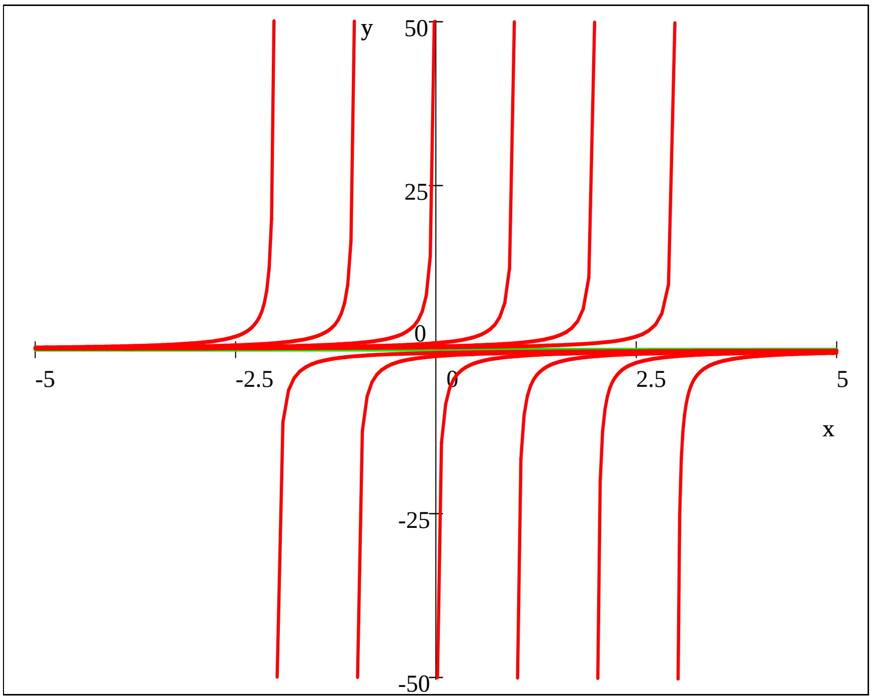
Example. 1. Consider the ODE $x' = x^2$ in the domain $\Omega = \mathbb{R}^2$. This is separable equation and can be solved as follows. Obviously, $x \equiv 0$ is a constant solution. In the domains where $x \neq 0$ we have

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

whence

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

and $x(t) = -\frac{1}{t-C}$ (where we have replaced C by $-C$). Hence, the family of all solutions consists of a straight line $x(t) = 0$ and hyperbolas $x(t) = \frac{1}{C-t}$ 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 $x(t) \to \pm\infty$.

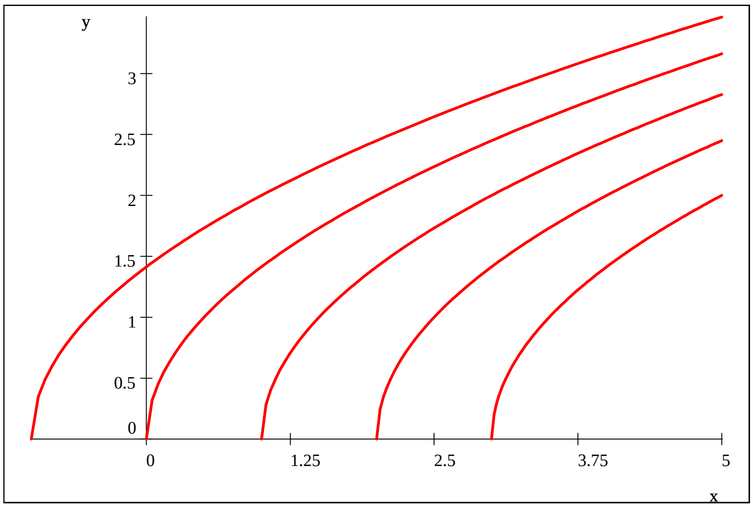
2. Consider the ODE $x' = \frac{1}{x}$ in the domain $\Omega = \mathbb{R} \times (0, +\infty)$ (that is, $t \in \mathbb{R}$ and $x > 0$). By the separation of variables, we obtain

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

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 2.8 will be preceded by a lemma.

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

$$x(t) = x_\alpha(t) \text{ if } t \in I_\alpha. \tag{2.28}$$

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

The function $x(t)$ defined by (2.28) is referred to as the *union* of the family $\{x_\alpha(t)\}$.

Proof. First of all, let us verify that the identity (2.28) defines $x(t)$ correctly, that is, the right hand side does not depend on the choice of α. Indeed, if also $t \in I_\beta$ then t belongs to the intersection $I_\alpha \cap I_\beta$ and by the uniqueness theorem, $x_\alpha(t) = x_\beta(t)$. Hence, the value of $x(t)$ is independent of the choice of the index α. Note that the graph of $x(t)$ is the union of the graphs of all functions $x_\alpha(t)$.

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

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

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

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

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

Proof of Theorem 2.8. (a) Consider the IVP

$$\begin{cases} x' = f(t,x), \\ x(t_0) = x_0 \end{cases} \tag{2.29}$$

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

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

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

(c) Let $x(t)$ be a maximal solution defined on (a,b) 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.

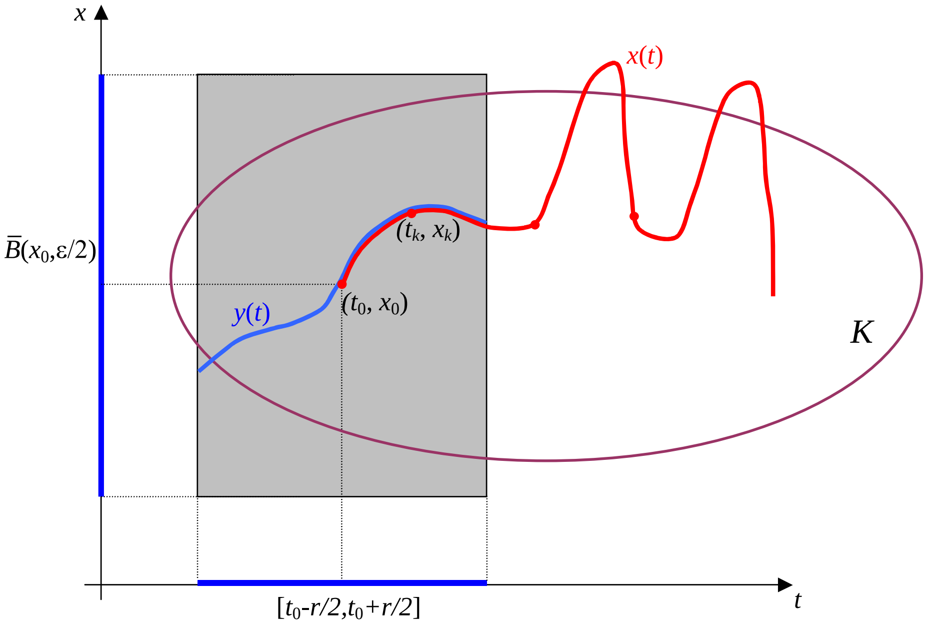
By Corollary to Theorem 2.7, for the point (t_0, x_0), there exist $r, \varepsilon > 0$ such that the IVP with the initial point inside the cylinder

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

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

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

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



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

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

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

Consider the IVP

$$\begin{cases} x' = f(t, x) \\ x(t_0) = x_0 \end{cases} \tag{2.30}$$

In Section 2.2, we have investigated in the one dimensional case the dependence of the solution $x(t)$ upon the initial value x_0. A more general question, which will be treated here, is how the solution $x(t)$ depends on the right hand side $f(t, x)$. The dependence on the initial condition can be reduced to the dependence of the right hand side as follows. Consider the function $y(t) = x(t) - x_0$, which obviously solves the IVP

$$\begin{cases} y' = f(t, y + x_0), \\ y(t_0) = 0. \end{cases} \tag{2.31}$$

Hence, if we know that the solution $y(t)$ of (2.31) depends continuously on the right hand side, then it will follow that $y(t)$ is continuous in x_0, which implies that also the solution $x(t)$ of (2.30) is continuous in x_0.

Let Ω be an open set in $\mathbb{R}^{n+1}$ and f, g be two functions 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

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} \tag{2.32}$$

and

$$\begin{cases} y' = g(t,y) \\ y(t_0) = x_0 \end{cases} \tag{2.33}$$

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

Assume that the function f as fixed and $x(t)$ is a fixed solution of (2.32). 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 (2.33) is close enough to $x(t)$. Apart from the theoretical interest, this question has significant practical consequences. For example, if one knows the function $f(t,x)$ only approximately (which is always the case in applications in Sciences and Engineering) then solving (2.32) approximately means solving another problem (2.33) where g is an approximation to f. Hence, it is important to know that the solution $y(t)$ of (2.33) is actually an approximation of $x(t)$.

Theorem 2.10 *Let $x(t)$ be a solution to the IVP (2.32) defined on an interval (a,b). Then, for all real α, β such that $a < \alpha < t_0 < \beta < b$ and for any $\varepsilon > 0$, there is $\eta > 0$ such that, for any function $g : \Omega \to \mathbb{R}^n$ such that*

$$\sup_{\Omega} \|f - g\| \leq \eta, \tag{2.34}$$

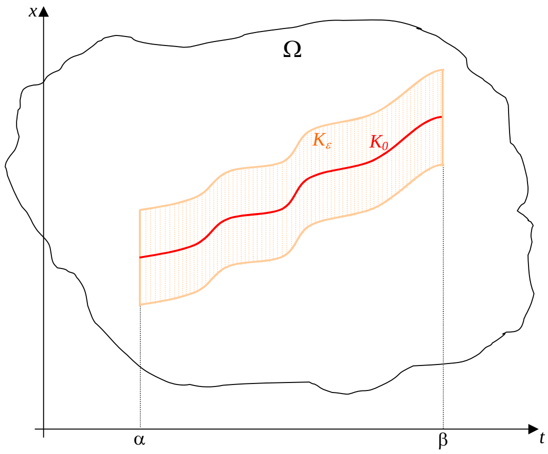
there is a solution $y(t)$ of the IVP (2.33) defined in $[\alpha, \beta]$, and this solution satisfies the inequality

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

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

$$K_\varepsilon = \left\{(t,x) \in \mathbb{R}^{n+1} : \alpha \leq t \leq \beta, \ \ \|x - x(t)\| \leq \varepsilon\right\} \tag{2.35}$$

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



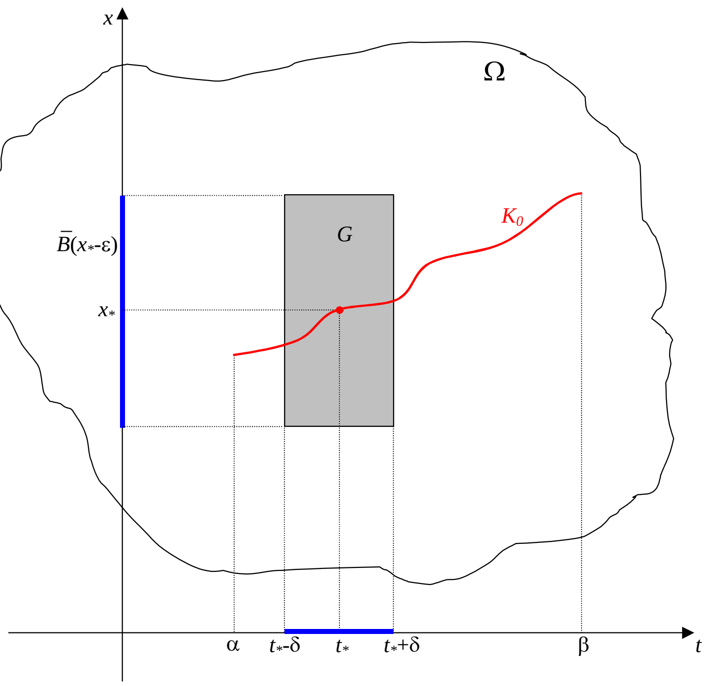
The set K_0 is compact because it is the image of the compact interval $[\alpha, \beta]$ under the continuous mapping $t \mapsto (t, x(t))$. Hence, K_0 is bounded and closed, which implies that also K_ε for any $\varepsilon > 0$ is also bounded and closed. Thus, K_ε is a compact subset of $\mathbb{R}^{n+1}$ for any $\varepsilon \geq 0$.

Claim 1. *There is $\varepsilon > 0$ such that $K_\varepsilon \subset \Omega$ and f is Lipschitz in x in K_ε.*

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).



Varying the point (t_*, x_*) in K_0, we obtain a cover of K_0 by the family of the open cylinders $H = (t_* - \delta, t_* + \delta) \times B(x_*, \varepsilon/2)$ where ε, δ depend on (t_*, x_*). Since K_0 is compact, there is a finite subcover, that is, a finite number of points $\{(t_i, x_i)\}_{i=1}^m$ on K_0 and the corresponding numbers $\varepsilon_i, \delta_i > 0$, such that the cylinders

$$H_i = (t_i - \delta_i, t_i + \delta_i) \times B(x_i, \varepsilon_i/2)$$

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. Set

$$\varepsilon = \frac{1}{2} \min_{1 \le i \le m} \varepsilon_i \text{ and } \ L = \max_{1 \le i \le m} L_i \tag{2.36}$$

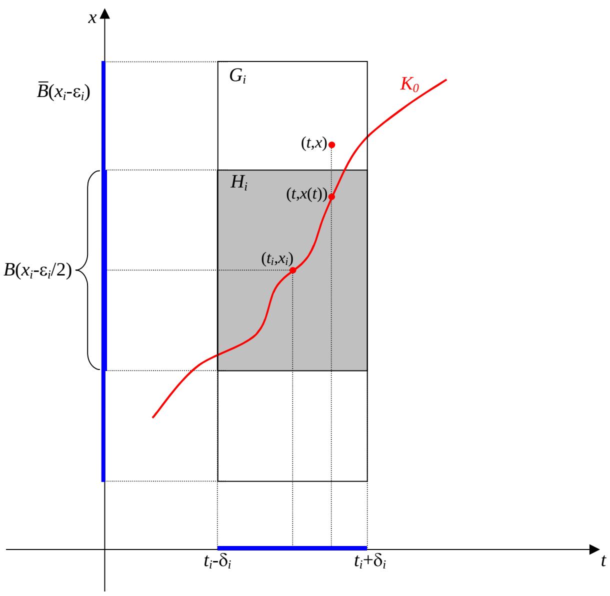
and prove that $K_\varepsilon \subset \Omega$ and that function f is Lipschitz in K_ε with the constant L. For any point $(t, x) \in K_\varepsilon$, we have by the definition of K_ε that $t \in [\alpha, \beta]$, $(t, x(t)) \in K_0$ and

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

The point $(t, x(t))$ belongs to one of the cylinders H_i so that

$$t \in (t_i - \delta_i, t_i + \delta_i) \quad \text{and} \quad \|x(t) - x_i\| < \varepsilon_i/2$$

(see the diagram below).



By the triangle inequality, we have

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

where we have used that by (2.36) $\varepsilon \le \varepsilon_i/2$. Therefore, $x \in B(x_i, \varepsilon_i)$ whence it follows that $(t, x) \in G_i$ and, hence, $(t, x) \in \Omega$. 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

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

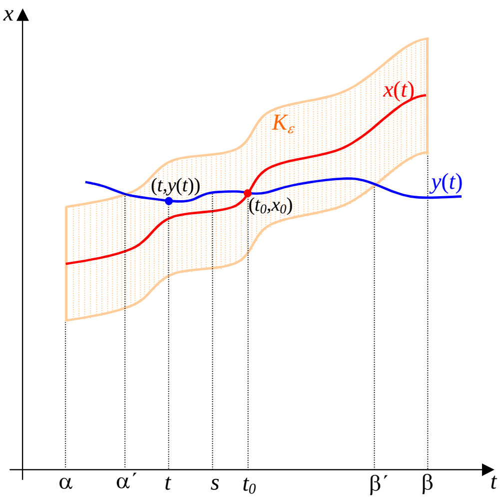
where we have used the definition (2.36) of L. This shows that f is Lipschitz in x in K_ε and finishes the proof of Claim 1.

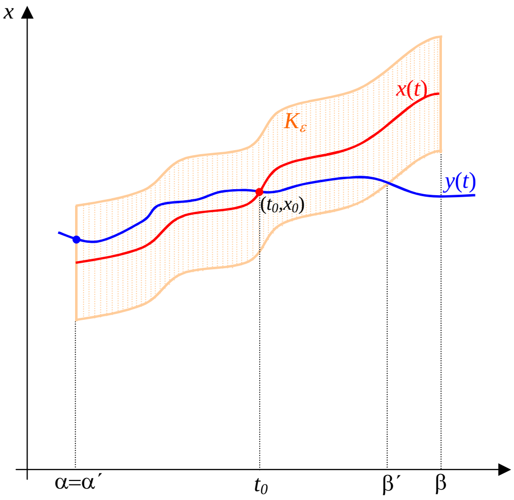
Observe that if the statement of Claim 1 holds for some value of ε then it holds for all smaller values of ε as well, with the same L. Hence, we can assume that the value of ε from Theorem 2.10 is small enough so that it satisfies Claim 1.

Let now $y(t)$ be the maximal solution to the IVP (2.33), and let (a',b') be its domain. By Theorem 2.8, the graph of $y(t)$ leaves K_ε when $t \to a'$ and when $t \to b'$. Let (α',β') be the maximal interval such that the graph of $y(t)$ on this interval is contained in K_ε; that is,

$$\alpha' = \inf\{t \in (\alpha,\beta)\cap(a',b') : (s,y(s)) \in K_\varepsilon \text{ for all } s \in [t,t_0]\} \tag{2.37}$$

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





In particular, (α',β') is contained in $(a',b')\cap(\alpha,\beta)$, function $y(t)$ is defined on (α',β') and

$$(t,y(t)) \in K_\varepsilon \text{ for all } t \in (\alpha',\beta'). \tag{2.38}$$

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

$$\alpha' > \alpha \quad \text{and} \quad \|x(t)-y(t)\| = \varepsilon \text{ for } t = \alpha'. \tag{2.39}$$

Similarly, either $\beta' = \beta$ or

$$\beta' < \beta \quad \text{and} \quad \|x(t)-y(t)\| = \varepsilon \text{ for } t = \beta'.$$

By Theorem 2.8, $y(t)$ leaves K_ε as $t \to a'$. Hence, for all values of t close enough to a' we have $(t,y(t)) \notin K_\varepsilon$. For any such t we have by (2.37) $t \le \alpha'$ whence $a' < t \le \alpha$ and $a' < \alpha'$. Similarly, one shows that $b' > \beta'$, whence the inclusion $[\alpha',\beta'] \subset [a',b']$ follows.

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

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

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

$$(\alpha' - \tau, \alpha' + \tau) \subset (\alpha, \beta) \cap (a', b'). \tag{2.40}$$

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

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

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

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

then, by the continuity of $x(t)$ and $y(t)$, that the same inequality holds for all $t \in (\alpha' - \tau, \alpha' + \tau)$ provided $\tau > 0$ is small enough. Choosing τ to satisfy also (2.40), we obtain that $(t, y(t)) \in K_\varepsilon$ for all $t \in (\alpha' - \tau, \alpha']$, which contradicts the definition of α'.

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

$$\sup_{K_\varepsilon} \|f - g\| \leq \eta, \tag{2.41}$$

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

In fact, Claim 3 will finish the proof of Theorem 2.10. Indeed, Claims 2 and 3 imply that $y(t)$ is defined on $[\alpha, \beta]$; by the definition of α' and β' (see (2.38)), we obtain $(t, y(t)) \in K_\varepsilon$ for all $t \in (\alpha, \beta)$, and by continuity, the same holds for $t \in [\alpha, \beta]$. By the definition (2.35) of K_ε, this means

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

which was the claim of Theorem 2.10.

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

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

and

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

Assuming for simplicity that $t \geq t_0$ and using the triangle inequality, we obtain

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

Since the points $(s, x(s))$ and $(s, y(s))$ are in K_ε, we obtain by the Lipschitz condition in K_ε (Claim 1) that

$$\|x(t) - y(t)\| \leq L \int_{t_0}^{t} \|x(s) - y(s)\| \, ds + \sup_{K_{\varepsilon s}} \|f - g\| \, (\beta - \alpha) . \tag{2.42}$$

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

$$\begin{aligned} \|x(t) - y(t)\| &\leq (\beta - \alpha) \exp L (t - t_0) \sup_{K_{\varepsilon s}} \|f - g\| \\ &\leq (\beta - \alpha) \exp L (\beta - \alpha) \sup_{K_{\varepsilon s}} \|f - g\| . \end{aligned} \tag{2.43}$$

In the same way, (2.43) holds for $t \leq t_0$ so that it is true for all $t \in [\alpha', \beta']$.

Now choose η in (2.41) as follows

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

Then it follows from (2.43) that

$$\|x(t) - y(t)\| \leq \varepsilon/2 < \varepsilon \text{ for all } t \in [\alpha', \beta'] . \tag{2.44}$$

By Claim 2, we conclude that $\alpha' = \alpha$ and $\beta' = \beta$, which finishes the proof. ∎

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

Corollary *Under the hypotheses of Theorem* 2.10, *let* $x(t)$ *be a solution to the IVP* (2.32) *defined on an interval* (a, b), *and let* α, β *be such that* $a < \alpha < t_0 < \beta < b$. *Let* $\varepsilon > 0$ *be sufficiently small so that* $f(t, x)$ *is Lipschitz in* x *in* K_ε *with the Lipschitz constant* L. *If* $\sup_{K_\varepsilon} \|f - g\|$ *is sufficiently small, then the IVP* (2.33) *has a solution* $y(t)$ *defined on* $[\alpha, \beta]$, *and the following estimate holds*

$$\sup_{[\alpha,\beta]} \|x(t) - y(t)\| \leq (\beta - \alpha) e^{L(\beta-\alpha)} \sup_{K_\varepsilon} \|f - g\| . \tag{2.45}$$

Proof. By Claim 2 of the above proof, the maximal solution $y(t)$ of (2.33) is defined on $[\alpha', \beta']$. Also, the difference $\|x(t) - y(t)\|$ satisfies (2.43) for all $t \in [\alpha', \beta']$. If $\sup_{K_\varepsilon} \|f - g\|$ is small enough then by Claim 3 $[\alpha', \beta'] = [\alpha, \beta]$. It follows that $y(t)$ is defined on $[\alpha, \beta]$ and satisfies (2.45). ∎

2.8 Continuity of solutions with respect to a parameter

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

$$\begin{cases} x' = f(t, x, s) \\ x(t_0) = x_0 \end{cases} \tag{2.46}$$

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

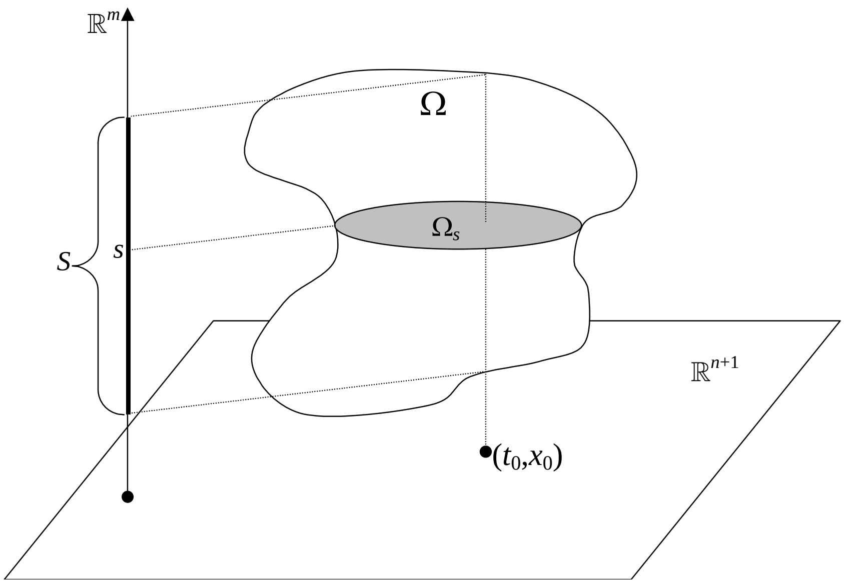
$$(t, x, s) = (t, x_1, .., x_n, s_1, ..., s_m) .$$

How do we understand (2.46)? 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 (2.46) can be considered in the domain Ω_s for any $s \in S$. We always assume that the set S is non-empty. Assume also in the sequel that $f(t, x, s)$ is a continuous function in $(t, x, s) \in \Omega$ and is locally Lipschitz in x for any $s \in S$. For any $s \in S$, denote by $x(t, s)$ the maximal solution of (2.46) and let I_s be its domain (that is, I_s is an open interval on the axis t). Hence, $x(t, s)$ as a function of (t, s) is defined in the set

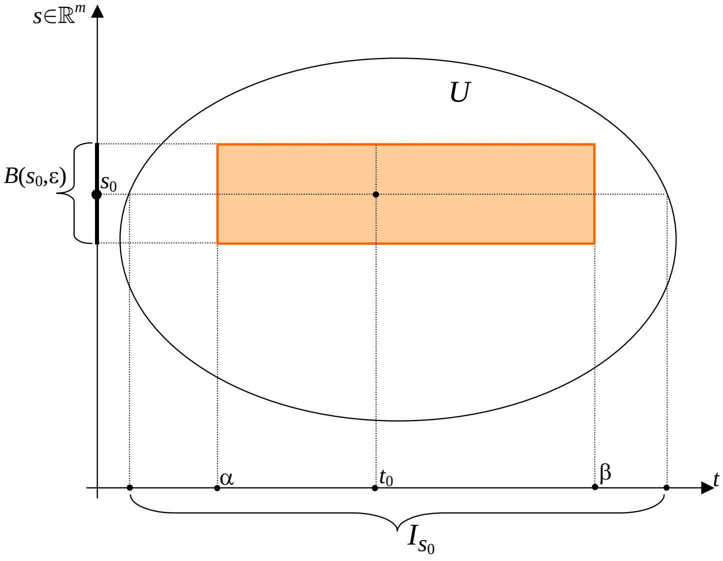
$$U = \left\{(t, s) \in \mathbb{R}^{m+1} : s \in S, t \in I_s\right\}.$$

Theorem 2.11 *Under the above assumptions, the set U is an open subset of $\mathbb{R}^{n+1}$ and the function $x(t, s) : U \to \mathbb{R}^n$ is continuous 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 $\varepsilon > 0$ such that

$$[\alpha, \beta] \times \overline{B}(s_0, \varepsilon) \subset U, \tag{2.47}$$

which will imply that U is open. Here $\overline{B}(s_0, \varepsilon)$ 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).



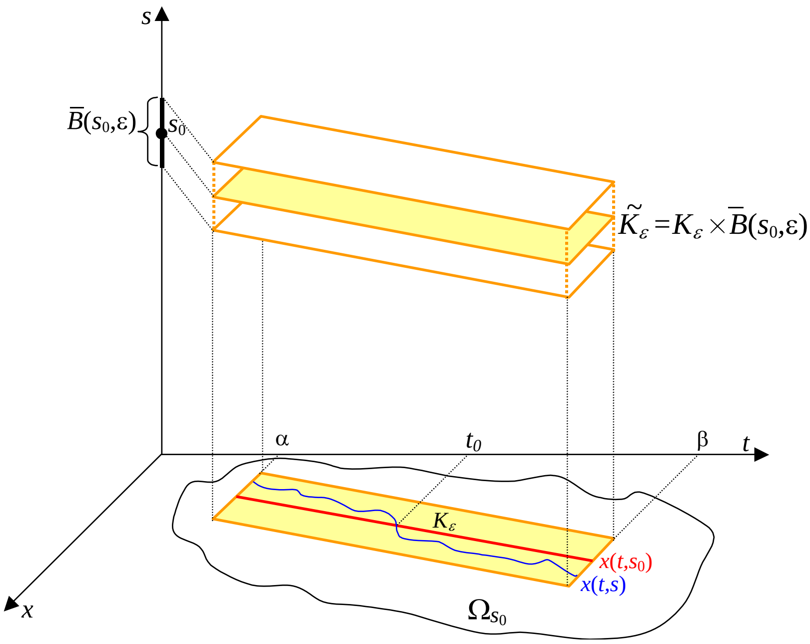
As in the proof of Theorem 2.10, consider a set

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

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

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

(see the diagram below).



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

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

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

Using Corollary to Theorem 2.10 with[5] $f(t,x) = f(t,x,s_0)$ and $g(t,x) = f(t,x,s)$ where $s \in \overline{B}(s_0,\varepsilon)$, we obtain that if

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

is small enough then then the solution $y(t) = x(t,s)$ is defined on $[\alpha,\beta]$. In particular, this implies (2.47) for small enough ε. Furthermore, by Corollary to Theorem 2.10 we also obtain that

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

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

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

so that $x(t,s)$ is continuous in s at s_0 uniformly in $t \in [\alpha,\beta]$. Since $x(t,s)$ is continuous in t for any fixed s, we conclude that x is continuous in (t,s) (see Exercise 28), which finishes the proof. ■

2.9 Global existence

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

$$\|f(t,x)\| \le a(t)\|x\| + b(t), \tag{2.48}$$

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

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} \tag{2.49}$$

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

In other words, under the specified conditions, the maximal solution of (2.49) is defined on I.

Proof. Let $x(t)$ be the maximal solution to the problem (2.49), and let $J = (\alpha,\beta)$ be the open interval where $x(t)$ is defined. We will show that $J = I$. Assume from the contrary that this is not the case. Then one of the points α,β is contained in I, say $\beta \in I$.

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

Let us investigate the behavior of $\|x(t)\|$ as $t \to \beta$. By Theorem 2.8, $(t, x(t))$ leaves any compact $K \subset \Omega := I \times \mathbb{R}^n$. Consider a compact set

$$K = [\beta - \varepsilon, \beta] \times \overline{B}(0, r)$$

where $\varepsilon > 0$ is so small that $[\beta - \varepsilon, \beta] \subset I$. Clearly, $K \subset \Omega$. If t is close enough to β then $t \in [\beta - \varepsilon, \beta]$. Since $(t, x(t))$ must be outside K, we conclude that $x \notin \overline{B}(0, r)$, that is, $\|x(t)\| > r$. Since r is arbitrary, we have proved that $\|x(t)\| \to \infty$ as $t \to \beta$.

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

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

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

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

where

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

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

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

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

Example. We have considered above the ODE $x' = x^2$ defined in $\mathbb{R} \times \mathbb{R}$ and have seen that the solution $x(t) = \frac{1}{C-t}$ cannot be defined on full $\mathbb{R}$. The same occurs for the equation $x' = x^\alpha$ for $\alpha > 1$. The reason is that the function $f(t, x) = x^\alpha$ does not admit the estimate (2.48) for large x, due to $\alpha > 1$. This example also shows that the condition (2.48) is rather sharp.

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

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

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

$$x_k' = \sum_{l=1}^{n} A_{kl}(t)\, x_l + B_k(t),$$

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

Theorem 2.13 *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*

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

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

Proof. It suffices to check that the function $f(t,x) = A(t)\,x + B(t)$ satisfies the conditions of Theorem 2.12. This function is obviously continuous in (t,x) and continuously differentiable in x, which implies by Lemma 2.6 that $f(t,x)$ is locally Lipschitz in x.

We are left to verify (2.48). By the triangle inequality, we have

$$\|f(t,x)\| \le \|A(t)\,x\| + \|B(t)\|. \tag{2.50}$$

Let all the norms be the ∞-norm. Then

$$b(t) := \|B(t)\|_\infty = \max_k |B_k(t)|$$

is a continuous function of t. Next,

$$\|A(t)\,x\|_\infty = \max_k |(A(t)\,x)_k| = \max_k \left|\sum_{l=1}^{\infty} A_{kl}(t)\,x_l\right| \le \left(\max_k \sum_{l=1}^{\infty} |A_{kl}(t)|\right) \max_l |x_l| = a(t)\,\|x\|_\infty\,,$$

where

$$a(t) = \max_k \sum_{l=1}^{\infty} |A_{kl}(t)|$$

is a continuous function. Hence, we obtain from (2.50)

$$\|f(t,x)\| \le a(t)\,\|x\| + b(t)\,,$$

which finishes the proof. ■

2.10 Differentiability of solutions in parameter

Consider again the initial value problem with parameter

$$\begin{cases} x' = f(t,x,s), \\ x(t_0) = x_0, \end{cases} \tag{2.51}$$

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 Jacobian matrices of f with respect to x and s. Set

$$f_x = \partial_x f = \frac{\partial f}{\partial x} := \left(\frac{\partial f_i}{\partial x_k}\right),$$

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. Similarly, set

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

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.

If f_x is continuous in Ω then by Lemma 2.6 f is locally Lipschitz in x so that all the previous results apply.

Let $x(t, s)$ be the maximal solution to (2.51). Recall that, by Theorem 2.11, the domain U of $x(t, s)$ is an open subset of $\mathbb{R}^{m+1}$ and $x : U \to \mathbb{R}^n$ is continuous.

Theorem 2.14 *Assume that function $f(t, x, s)$ is continuous and f_x and f_s exist and are also continuous in Ω. Then $x(t, s)$ is continuously differentiable in $(t, s) \in U$ and the Jacobian matrix $y = \partial_s x$ solves the initial value problem*

$$\begin{cases} y' = f_x(t, x(t, s), s)\, y + f_s(t, x(t, s), s), \\ y(t_0) = 0. \end{cases} \tag{2.52}$$

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 (2.52) are also $n \times m$ matrices so that (2.52) 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 ODE in (2.52) is called the *variational equation* for (2.51) along the solution $x(t, s)$ (or the *equation in variations*).

Note that the variational equation is linear. Indeed, for any fixed s, its right hand side can be written in the form

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

where $A(t) = f_x(t, x(t, s), s)$ and $B(t) = f_s(t, x(t, s), s)$. Since f is continuous and $x(t, s)$ is continuous by Theorem 2.11, the functions $A(t)$ and $B(t)$ are continuous in t. If the domain in t of the solution $x(t, s)$ is I_s then the domain of the variational equation is $I_s \times \mathbb{R}^{n\times m}$. By Theorem 2.13, the solution $y(t)$ of (2.52) exists in the full interval I_s. Hence, Theorem 2.14 can be stated as follows: if $x(t, s)$ is the solution of (2.51) on I_s and $y(t)$ is the solution of (2.52) on I_s then we have the identity $y(t) = \partial_s x(t, s)$ for all $t \in I_s$. This provides a method of evaluating $\partial_s x(t, s)$ for a fixed s without finding $x(t, s)$ for all s.

Example. Consider the IVP with parameter

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

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

For $s = 0$ we have the IVP

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

whence we obtain $x(t, 0) = -\frac{1}{t}$. Noticing that $f_x = 2x$ and $f_s = 2/t$ we obtain the variational equation along this solution

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

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

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

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

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

The initial condition $y(1) = 0$ is satisfied for $C = -1$ so that $y(t) = 1 - t^{-2}$. By Theorem 2.14, we conclude that $\partial_s x(t, 0) = 1 - t^{-2}$.

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

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

that is,

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

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 $x(t, s)$.

Remark. It is easy to deduce the variational equation (2.52) provided we know that the function $x(t, s)$ is sufficiently many times differentiable. Assume that the mixed partial derivatives $\partial_s \partial_t x$ and $\partial_t \partial_s x$ exist and are the equal (for example, this is the case when $x(t, s) \in C^2(U)$). Then differentiating (2.51) 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 (2.52) after substitution $\partial_s x = y$. Although this argument is not a proof of Theorem 2.14, it allows to memorize the variational equation. The main technical difficulty in the proof of Theorem 2.14 is verifying the differentiability of x in s.

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 (2.52), which we write in the form

$$\begin{cases} y' = g(t, y, s) \\ y(t_0) = 0 \end{cases} \tag{2.53}$$

where

$$g(t, y, s) = f_x(t, x(t, s), s)\, y + f_s(t, x(t, s), s). \tag{2.54}$$

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 2.14 that f_x and f_s are continuous, can be shortly written as $f \in C^1(x, s)$, and the claim of Theorem 2.14 is that $x(t, s) \in C^1(t, s)$.

Assume now that $f \in C^2(x,s)$. Then by (2.54) we obtain that g is continuous and $g \in C^1(y,s)$, whence by Theorem 2.14 $y \in C^1(s)$. In particular, the function $z = \partial_s y = \partial_{ss} x$ is defined. Applying the variational equation to the problem (2.53), 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

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

Note that here x must be substituted by $x(t,s)$ and y – by $y(t,s)$.

The equation (2.55) 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 finds the variational equations of the higher orders.

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

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

where we have computed that

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

Let us now evaluate $z = \partial_{ss} x(t,0)$. 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

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

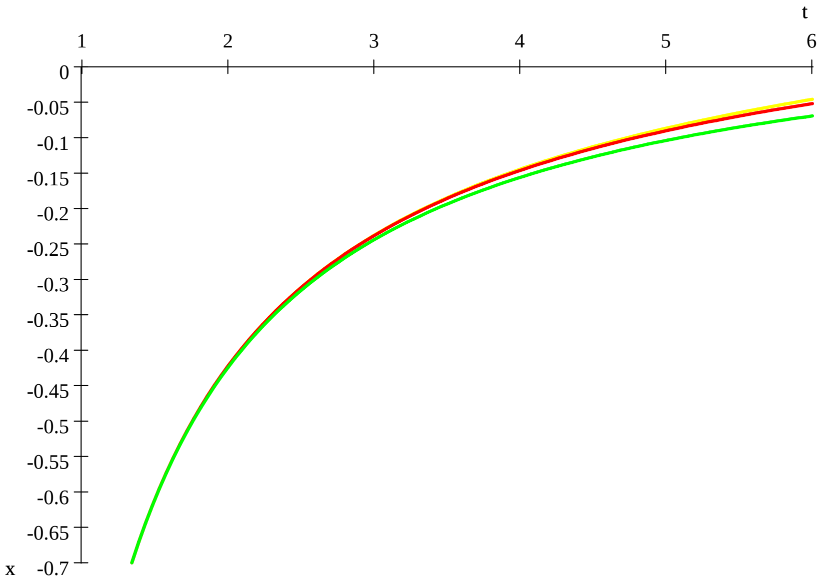
The initial condition $z(1) = 0$ yields $C = \frac{16}{3}$ whence

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

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

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

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



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

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

Let us write a similar expansion for $x'=\partial_t x$, 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,

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

Substituting this into the equation

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

we obtain

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

whence

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

Equating the terms with the same powers of s (which can be done by the uniqueness of the Taylor expansion), we obtain the equations

$$\begin{aligned} x'(t) &= x^2(t) \\ y'(t) &= 2x(t)\,y(t) + 2s/t \\ z'(t) &= 2x(t)\,z(t) + 2y^2(t). \end{aligned}$$

From the initial condition $x(1,s) = -1$ we obtain

$$-1 = x(1) + sy(1) + \frac{s^2}{2}z(1) + o\left(s^2\right),$$

whence $x(t) = -1$, $y(1) = z(1) = 0$. Solving successively the above equations with these initial conditions, we obtain the same result as above.

Before we prove Theorem 2.14, 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 $\lambda \in [0,1]$

$$\|(1-\lambda)\,x + \lambda y\| \le (1-\lambda)\,\|x\| + \lambda\,\|y\| < r.$$

It follows that $(1-\lambda)\,x + \lambda y \in B(0,r)$, which was to be proved.

Lemma 2.15 (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

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

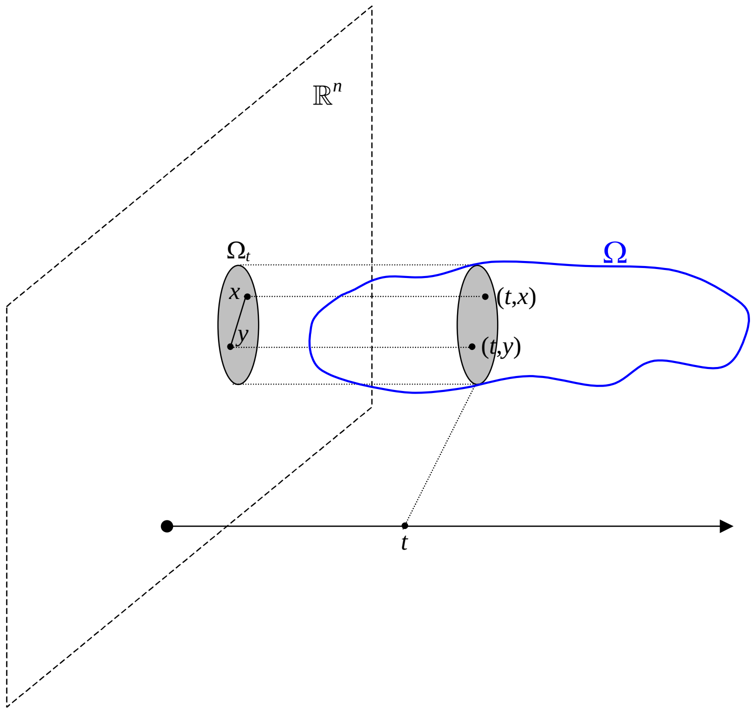
Then there exists a continuous mapping $\varphi(t,x,y) : \Omega' \to \mathbb{R}^{l \times n}$ such that the following identity holds:

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

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

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

$$\varphi(t,x,x) = f_x(t,x). \tag{2.57}$$



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

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

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

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

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

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

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

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

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

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

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

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

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

Clearly, this argument 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 2.15. It suffices to prove this lemma for each component f_i separately. Hence, we can assume that $l = 1$ so that φ is a row $(\varphi_1, ..., \varphi_n)$. Hence, we need to prove the existence of n real valued continuous functions $\varphi_1, ..., \varphi_n$ of (t, x, y) such that the following identity holds:

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

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

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

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

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

By the fundamental theorem of calculus, we obtain

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

where

$$\varphi_i(t, x, y) = \int_0^1 f_{x_i}(t, x + \lambda(y - x))\, d\lambda. \tag{2.58}$$

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 2.16 *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 2.16. 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 2.15 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 2.16 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 (2.58) that

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

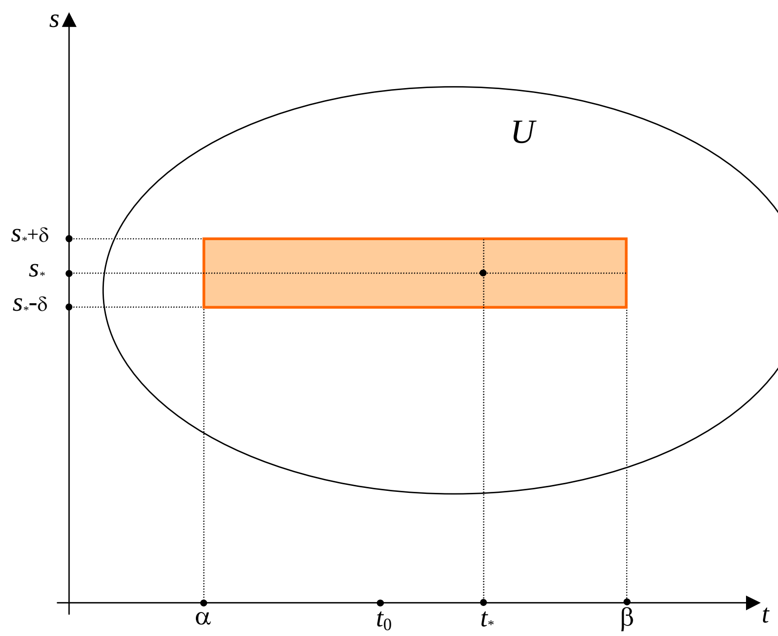
and, hence, $\varphi(t, x, x) = f_x(t, x)$, that is, (2.57). ■

Now we are in position to prove Theorem 2.14.

Proof of Theorem 2.14. In the main part of the proof, we show that the partial derivative $\partial_{s_i} x$ exists. Since this can be done separately for any component s_i, in this part we can and will assume that s is one-dimensional (that is, $m = 1$).

Fix some $(t_*, s_*) \in U$ and prove that $\partial_s x$ exists at this point. Since the differentiability is a local property, we can restrict the domain of the variables (t, s) as follows. Choose $[\alpha, \beta]$ to be any interval in I_{s_*}containing both t_0 and t_*. By Theorem 2.11, for any $\varepsilon > 0$ there is $\delta > 0$ such that the rectangle $(a, \beta) \times (s_* - \delta, s_* + \delta)$ is contained in U and, for all $s \in (s_* - \delta, s_* + \delta)$,

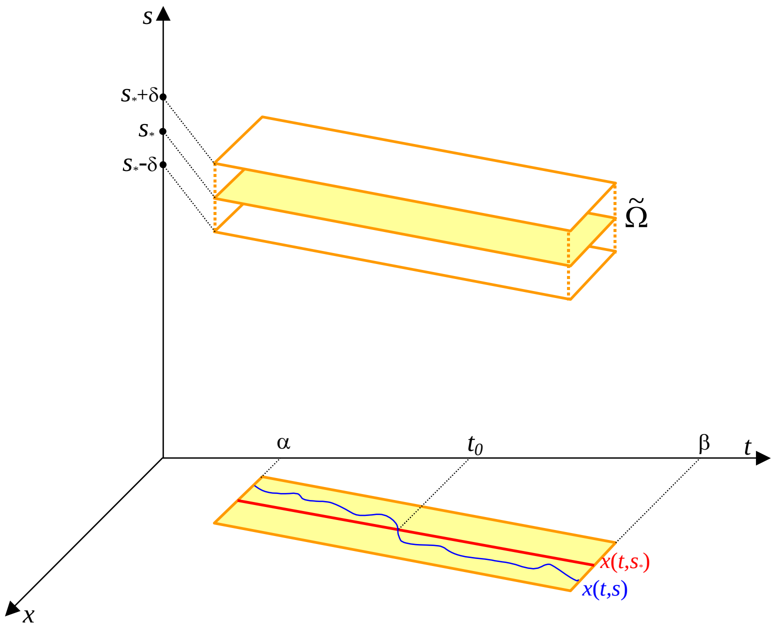
$$\sup_{t \in (\alpha, \beta)} \|x(t, s) - x(t, s_*)\| < \varepsilon.$$



Besides, by the openness of Ω, ε and δ can be chosen so small that the following condition is satisfied:

$$\widetilde{\Omega} := \left\{(t, x, s) \in \mathbb{R}^{n+m+1} : \alpha < t < \beta,\ \|x - x(t, s_*)\| < \varepsilon,\quad |s - s_*| < \delta\right\} \subset \Omega$$

(cf. the proof of Theorem 2.11). In particular, for all $t \in (\alpha, \beta)$ and $s \in (s_* - \delta, s_* + \delta)$, the solution $x(t, s)$ is defined and $(t, x(t, s), s) \in \widetilde{\Omega}$.



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

Applying the Hadamard lemma to the function $f(t, x, s)$ in this domain and using the fact that f is continuously differentiable with respect to (x, s), we obtain the identity

$$f(t, y, s) - f(t, x, \sigma) = \varphi(t, x, \sigma, y, s)(y - x) + \psi(t, x, \sigma, y, s)(s - \sigma),$$

where φ and ψ are continuous functions on the appropriate domains. In particular, substituting $\sigma = s_*$, $x = x(t, s_*)$ and $y = x(t, s)$, we obtain

$$\begin{aligned} f(t, x(t, s), s) - f(t, x(t, s_*), s_*) &= \varphi(t, x(t, s_*), s_*, x(t, s), s)(x(t, s) - x(t, s_*)) \\ &\quad + \psi(t, x(t, s_*), s_*, x(t, s), s)(s - s_*) \\ &= a(t, s)(x(t, s) - x(t, s_*)) + b(t, s)(s - s_*), \end{aligned}$$

where the functions

$$a(t, s) = \varphi(t, x(t, s_*), s_*, x(t, s), s) \text{ and } b(t, s) = \psi(t, x(t, s_*), s_*, x(t, s), s) \tag{2.59}$$

are continuous in $(t, s) \in (\alpha, \beta) \times (s_* - \delta, s_* + \delta)$ (the dependence on s_* is suppressed because s_* is fixed).

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

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

and observe that

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

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

$$\begin{cases} z' = a(t,s) z + b(t,s) \\ z(t_0, s) = 0. \end{cases} \tag{2.60}$$

Since this ODE is linear and the functions a and b are continuous in $(t,s) \in (\alpha, \beta) \times (s_* - \delta, s_* + \delta)$, we conclude by Theorem 2.13 that the solution to this IVP exists for all $s \in (s_* - \delta, s_* + \delta)$ and $t \in (\alpha, \beta)$ and, by Theorem 2.11, the solution is continuous in $(t,s) \in (\alpha, \beta) \times (s_* - \delta, s_* + \delta)$. Hence, we can define $z(t,s)$ also at $s = s_*$ as the solution of the IVP (2.60). In particular, using the continuity of $z(t,s)$ in s, we obtain

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

that is,

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

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

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

Note that by (2.59) and Lemma 2.15

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

and

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

Hence, we obtain that $y(t)$ satisfies the variational equation (2.52).

To finish the proof, we have to verify that $x(t,s)$ is continuously differentiable in (t,s). Here we come back to the general case $s \in \mathbb{R}^m$. The derivative $\partial_s x = y$ satisfies the IVP (2.52) and, hence, is continuous in (t,s) by Theorem 2.11. Finally, for the derivative $\partial_t x$ we have the identity

$$\partial_t x = f(t, x(t,s), s), \tag{2.61}$$

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

Remark. It follows from (2.61) that $\partial_t x$ is differentiable in s and, by the chain rule,

$$\partial_s(\partial_t x) = \partial_s [f(t, x(t,s), s)] = f_x(t, x(t,s), s) \partial_s x + f_s(t, x(t,s), s). \tag{2.62}$$

On the other hand, it follows from (2.52) that

$$\partial_t (\partial_s x) = \partial_t y = f_x (t, x(t,s), s) \partial_s x + f_s (t, x(t,s), s), \tag{2.63}$$

whence we conclude that

$$\partial_s \partial_t x = \partial_t \partial_s x. \tag{2.64}$$

Hence, the derivatives ∂_s and ∂_t commute[6] on x. As we have seen above, if one knew the identity (2.64) a priori then the derivation of the variational equation (2.52) would have been easy. However, in the present proof the identity (2.64) comes *after* the variational equation.

Theorem 2.17 *Under the conditions of* Theorem 2.14, *assume that, for some $k \in \mathbb{N}$, $f(t,x,s) \in C^k(x,s)$. Then the maximal solution $x(t,s)$ belongs to $C^k(s)$. Moreover, for any multiindex α of the order $|\alpha| \leq k$ and of the dimension m (the same as that of s), we have*

$$\partial_t \partial_s^\alpha x = \partial_s^\alpha \partial_t x. \tag{2.65}$$

Here $\alpha = (\alpha_1, ..., \alpha_m)$ where α_i are non-negative integers, $|\alpha| = \alpha_1 + ... + \alpha_n$, and

$$\partial_s^\alpha = \frac{\partial^{|\alpha|}}{\partial s_1^{\alpha_1} ... \partial s_m^{\alpha_m}}.$$

3 Linear equations and systems

A linear (system of) ODE of the first order is a (vector) ODE of the form

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

where $A(t) : I \to \mathbb{R}^{n \times n}$, $B : I \to \mathbb{R}^n$, and I being an open interval in $\mathbb{R}$. If $A(t)$ and $B(t)$ are continuous in t then, for any $t_0 \in I$ and $x_0 \in \mathbb{R}^n$, the IVP

$$\begin{cases} x' = A(t)x + B(t) \\ x(t_0) = x_0 \end{cases} \tag{3.1}$$

has a unique solution defined on the full interval I (cf. Theorem 2.13). In the sequel, we always assume that $A(t)$ and $B(t)$ are continuous on I and consider only solutions defined on the entire interval I.

3.1 Space of solutions of homogeneous systems

The linear ODE is called *homogeneous* if $B(t) \equiv 0$, and *inhomogeneous* otherwise. In this Section, we consider a homogeneous equation, that is, the equation $x' = A(t)x$. Denote by $\mathcal{A}$ the set of all solutions of this ODE.

[6] The equality of the mixed derivatives can be concluded by a theorem from Analysis II if one knows that both $\partial_s \partial_t x$ and $\partial_t \partial_s x$ are continuous. Their continuity follows from the identities (2.62) and (2.63), which prove at the same time also their equality.

Theorem 3.1 *$\mathcal{A}$ is a linear space and* $\dim \mathcal{A} = n$. *Consequently, if $x_1, ..., x_n$ are n linearly independent solutions to $x' = A(t)x$ then the general solution has the form*

$$x(t) = C_1 x_1(t) + ... + C_n x_n(t), \tag{3.2}$$

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

Proof. 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.

If x and $y \in \mathcal{A}$ then also $x + y \in \mathcal{A}$ because

$$(x+y)' = x' + y' = Ax + Ax = A(x+y)$$

and similarly $\lambda x \in \mathcal{A}$ for any $\lambda \in \mathbb{R}$. Hence, $\mathcal{A}$ is a linear space.

Fix $t_0 \in I$ and consider the mapping $\Phi : \mathcal{A} \to \mathbb{R}^n$ given by $\Phi(x) = x(t_0)$. This mapping is obviously linear. It is surjective since 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 (3.2). ∎

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

$$x^{(n)} + a_1(t)x^{(n-1)} + ... + a_n(t)x = 0, \tag{3.3}$$

where all functions $a_k(t)$ are defined on an open interval $I \subset \mathbb{R}$ and are continuous on I. As we know, such an ODE can be reduced to the vector ODE of the 1st order as follows. Consider the vector function

$$\mathbf{x}(t) = \left(x(t), x'(t), ..., x^{(n-1)}(t)\right) \tag{3.4}$$

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 (3.3) is equivalent to the system

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

that is,

$$\mathbf{x}' = A(t)\mathbf{x} \tag{3.5}$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \ldots & -a_1 \end{pmatrix}.$$

Since $A(t)$ is continuous in t on I, we can assume that any solution $\mathbf{x}(t)$ of (3.5) is defined on the entire interval I and, hence, the same is true for any solution $x(t)$ of (3.3).

Denote now by $\widetilde{\mathcal{A}}$ the set of all solutions of (3.3) defined on I.

Corollary. *$\widetilde{\mathcal{A}}$ is a linear space and* $\dim \widetilde{\mathcal{A}} = n$*. Consequently, if $x_1, ..., x_n$ are n linearly independent solutions to $x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = 0$ then the general solution has the form*

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

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

Proof. The fact that $\widetilde{\mathcal{A}}$ is a linear space is obvious (cf. the proof of Theorem 3.1). The relation (3.4) 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 (3.3) gives back a solution $x(t)$ of (3.5). Hence, $\widetilde{\mathcal{A}}$ and $\mathcal{A}$ are linearly isomorphic, whence $\dim \widetilde{\mathcal{A}} = \dim \mathcal{A} = n$. ■

3.2 Linear homogeneous ODEs with constant coefficients

Consider the methods of finding n independent solutions to the ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = 0, \tag{3.6}$$

where $a_1, ..., a_n$ are real constants.

It will be convenient to obtain the complex valued general solution $x(t)$ and then to extract the real valued general solution. The idea is very simple. Let us look for a solution in the form $x(t) = e^{\lambda t}$ where λ is a complex number to be determined. Substituting this function into (3.6) and noticing that $x^{(k)} = \lambda^k e^{\lambda t}$, we obtain the equation for λ (after cancellation by $e^{\lambda t}$):

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

This equation is called the *characteristic equation* of (3.6) and the polynomial $P(\lambda) = \lambda^n + a_1 \lambda^{n-1} + ... + a_n$ is called the *characteristic polynomial* of (3.6). Hence, if λ is the root of the characteristic polynomial then the function $e^{\lambda t}$ solves (3.6). We try to obtain in this way n independent solutions.

Theorem 3.2 *If the characteristic polynomial $P(\lambda)$ of* (3.6) *has n distinct complex roots $\lambda_1, ..., \lambda_n$, then the following n functions $e^{\lambda_1 t}, ..., e^{\lambda_n t}$ are linearly independent solutions of* (3.6)*. Consequently, the general complex solution of* (3.6) *is given by*

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

where C_j are arbitrary complex numbers.

If $\lambda = \alpha + i\beta$ is a non-real root of $P(\lambda)$ then $\overline{\lambda} = \alpha - i\beta$ is also a root, and the functions $e^{\lambda t}$, $e^{\overline{\lambda} t}$ in the above sequence can be replaced by the real-valued functions $e^{\alpha t} \cos \beta t$, $e^{\alpha t} \sin \beta t$.

Proof. Let us prove this by induction in n that the functions $e^{\lambda_1 t}, ..., e^{\lambda_n t}$ are linearly independent provided $\lambda_1, ..., \lambda_n$ 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 $t \in \mathbb{R}$,

$$C_1 e^{\lambda_1 t} + ... + C_n e^{\lambda_n t} = 0, \tag{3.7}$$

and prove that $C_1 = ... = C_n = 0$. Dividing (3.7) 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 (3.7), we obtain also $C_n = 0$.

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

$$e^{\lambda t} = e^{\alpha t}(\cos \beta t + i \sin \beta t) \quad \text{and} \quad e^{\overline{\lambda} t} = e^{\alpha t}(\cos \beta t - \sin \beta t) \tag{3.8}$$

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

$$e^{\alpha t} \cos \beta t = \frac{1}{2}\left(e^{\lambda t} + e^{\overline{\lambda} t}\right) \quad \text{and} \quad e^{\alpha t} \sin \beta t = \frac{1}{2i}\left(e^{\lambda t} - e^{\overline{\lambda} t}\right). \tag{3.9}$$

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. ∎

Example. Consider the ODE

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

The characteristic polynomial is $P(\lambda) = \lambda^2 - 3\lambda + 2$, which has the roots $\lambda_1 = 2$ and $\lambda_2 = 1$. Hence, the linearly independent solutions are e^{2t} and e^t, and the general solution is $C_1 e^{2t} + C_2 e^t$.

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 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)

$$\frac{e^{it} + e^{-it}}{2} = \cos t \quad \text{and} \quad \frac{e^{it} - e^{-it}}{2i} = \sin t.$$

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 $P(\lambda) = \lambda^3 - 1 = (\lambda - 1)\left(\lambda^2 + \lambda + 1\right)$ that has the roots $\lambda_1 = 1$ and $\lambda_{2,3} = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$. Hence, we obtain the three linearly independent real solutions

$$e^t, \; e^{-\frac{1}{2}t} \cos \frac{\sqrt{3}}{2} t, \; e^{-\frac{1}{2}t} \sin \frac{\sqrt{3}}{2} t,$$

and the real general solution is

$$C_1e^t + e^{-\frac{1}{2}t}\left(C_2\cos\frac{\sqrt{3}}{2}t + C_3\sin\frac{\sqrt{3}}{2}t\right).$$

What to do when $P(\lambda)$ has fewer than n distinct roots? Recall the fundamental theorem of algebra (which is normally proved in a course of Complex Analysis): any polynomial $P(\lambda)$ of degree n with complex coefficients has exactly n complex roots counted with multiplicity. What is it the multiplicity of a root? If λ_0 is a root of $P(\lambda)$ then its multiplicity is the maximal natural number m such that $P(\lambda)$ is divisible by $(\lambda-\lambda_0)^m$, that is, the following identity holds

$$P(\lambda) = (\lambda-\lambda_0)^m Q(\lambda),$$

where $Q(\lambda)$ is another polynomial of λ. Note that $P(\lambda)$ is always divisible by $\lambda-\lambda_0$ so that $m \geq 1$. The fundamental theorem of algebra can be stated as follows: if $\lambda_1,...,\lambda_r$ are all distinct roots of $P(\lambda)$ and the multiplicity of λ_j is m_j then

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

and, hence,

$$P(\lambda) = (\lambda-\lambda_1)^{m_1}...(\lambda-\lambda_r)^{m_r}.$$

In order to obtain n independent solutions to the ODE (3.6), each root λ_j should give rise to m_j independent solutions.

Theorem 3.3 *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* (3.6)*:*

$$\left\{t^k e^{\lambda_j t}\right\},\ j=1,...,r,\ k=0,...,m_j-1. \tag{3.10}$$

Consequently, the general solution of (3.6) *is*

$$x(t) = \sum_{j=1}^{r}\sum_{k=0}^{m_j-1} C_{kj}t^k e^{\lambda_j t}, \tag{3.11}$$

where C_{kj} are arbitrary complex constants.

If $\lambda = \alpha + i\beta$ is a non-real root of P 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 (3.10) *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 (3.11)

$$x(t) = \sum_{j=1}^{r} P_j(t)e^{\lambda_j t}. \tag{3.12}$$

Hence, any solution to (3.6) has the form (3.12) 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 3.3, 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:

$$e^t,\ te^t,\ e^{-t},\ te^{-t},\ t^2e^{-t}.$$

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:

$$1,\ e^{it},\ te^{it},\ e^{-it},\ te^{-it}. \tag{3.13}$$

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 (3.13) 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 3.3. Given a polynomial $P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + ... + a_0$ 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

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = 0 \tag{3.14}$$

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 (3.14).

Example. Let us prove the following identity:

$$P\left(\frac{d}{dt}\right) e^{\lambda t} = P(\lambda) e^{\lambda t}. \tag{3.15}$$

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.

Lemma 3.4 *If $f(t), g(t)$ are n times differentiable functions on an open interval then, for any polynomial P of the order at most n, the following identity holds:*

$$P\left(\frac{d}{dt}\right)(fg) = \sum_{j=0}^{n} \frac{1}{j!} f^{(j)} P^{(j)}\left(\frac{d}{dt}\right) g. \tag{3.16}$$

Example. Let $P(\lambda) = \lambda^2 + \lambda + 1$. Then $P'(\lambda) = 2\lambda + 1$, $P'' = 2$, and (3.16) 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 (3.16) in the case when $P(\lambda) = \lambda^k$, $k \leq n$, because then for a general polynomial (3.16) 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,

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

and (3.16) becomes

$$(fg)^{(k)} = \sum_{j=0}^{k} \frac{k(k-1)\ldots(k-j+1)}{j!} f^{(j)} g^{(k-j)} = \sum_{j=0}^{k} \binom{k}{j} f^{(j)} g^{(k-j)}. \tag{3.17}$$

The latter identity is known from Analysis and is called the *Leibniz formula*[7]. ■

Lemma 3.5 *A complex number λ is a root of a polynomial P with the multiplicity m if and only if*

$$P^{(k)}(\lambda) = 0 \text{ for all } k = 0, \ldots, m-1 \text{ and } P^{(m)}(\lambda) \neq 0. \tag{3.18}$$

Proof. If P has a root λ with multiplicity m then we have the identity for all $z \in \mathbb{C}$

$$P(z) = (z-\lambda)^m Q(z)$$

where Q is a polynomial such that $Q(\lambda) \neq 0$. 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,

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

[7] If $k = 1$ then (3.17) amounts to the familiar product rule

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

For arbitrary $k \in \mathbb{N}$, (3.17) is proved by induction in k.

Conversely, if (3.18) holds then by the Taylor formula for a polynomial at λ, we have

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

where

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

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

Lemma 3.6 *If $\lambda_1, ..., \lambda_r$ are distinct complex numbers and if, for some polynomials $P_j(t)$,*

$$\sum_{j=1}^{r} P_j(t) e^{\lambda_t t} = 0 \quad \textit{for all } t \in \mathbb{R}, \tag{3.19}$$

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 (3.19) by $e^{\lambda_r t}$ and setting $\mu_j = \lambda_j - \lambda_r$, we obtain the identity

$$\sum_{j=1}^{r-1} P_j(t) e^{\mu_j t} + P_r(t) = 0. \tag{3.20}$$

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 of the degree at most k, which is only possible if $P_j \equiv 0$. Substituting into (3.20), we obtain that also $P_r \equiv 0$. ■

Proof of Theorem 3.3. Let $P(\lambda)$ be the characteristic polynomial of (3.14). We first prove that if λ is a root of multiplicity m then the function $t^k e^{\lambda t}$ solves (3.14) for any $k = 0, ..., m-1$. By Lemma 3.4, 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

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

that is, the function $x(t) = t^k e^{\lambda t}$ solves (3.14).

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

$$\left\{t^k e^{\lambda_j t}\right\},\ j = 1, ..., r,\ k = 0, ..., m_j - 1, \tag{3.21}$$

is a solution of (3.14). Let us show that these functions are linearly independent. Clearly, each linear combination of functions (3.21) has the form

$$\sum_{j=1}^{r} \sum_{k=0}^{m_j-1} C_{jk} t^k e^{\lambda_j t} = \sum_{j=1}^{r} P_j(t) e^{\lambda_j t} \tag{3.22}$$

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 3.6 $P_j \equiv 0$, which implies that all C_{jk} are 0. Hence, the functions (3.21) are linearly independent, and by Theorem 3.1 the general solution of (3.14) has the form (3.22).

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 3.5, λ satisfies the relations (3.18). 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 (3.21) 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

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

multiplied by t^k (compare the proof of Theorem 3.2). ∎

3.3 Space of solutions of inhomogeneous systems

Consider now an inhomogeneous linear ODE

$$x' = A(t)x + B(t), \tag{3.23}$$

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 $I \subset \mathbb{R}$.

Theorem 3.7 *If $x_0(t)$ is a particular solution of* (3.23) *and $x_1(t), ..., x_n(t)$ is a sequence of n linearly independent solutions of the homogeneous ODE $x' = Ax$ then the general solution of* (3.23) *is given by*

$$x(t) = x_0(t) + C_1 x_1(t) + ... + C_n x_n(t). \tag{3.24}$$

Proof. If $x(t)$ is also a solution of (3.23) then the function $y(t) = x(t) - x_0(t)$ solves $y' = Ay$, whence by Theorem 3.1

$$y = C_1 x_1(t) + ... + C_n x_n(t), \tag{3.25}$$

and $x(t)$ satisfies (3.24). Conversely, for all $C_1, ... C_n$, the function (3.25) solves $y' = Ay$, whence it follows that the function $x(t) = x_0(t) + y(t)$ solves (3.23). ∎

Consider now a scalar ODE

$$x^{(n)} + a_1(t) x^{(n-1)} + ... + a_n(t) x = f(t) \tag{3.26}$$

where all functions $a_1, ..., a_n, f$ are continuous on an interval I.

Corollary *If $x_0(t)$ is a particular solution of* (3.26) *and $x_1(t), ..., x_n(t)$ is a sequence of n linearly independent solutions of the homogeneous ODE*

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

then the general solution of (3.26) *is given by*

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

The proof is trivial and is omitted.

3.4 Linear inhomogeneous ODEs with constant coefficients

Here we consider the ODE

$$x^{(n)} + a_1 x^{(n-1)} + ... + a_n x = f(t), \tag{3.27}$$

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 we know, the general solution of the inhomogeneous equation (3.27) is obtained as a sum of the general solution of the homogeneous equation and a particular solution of (3.27). Hence, we focus on finding a particular solution of (3.27).

As before, denote by $P(\lambda)$ the characteristic polynomial of (3.27), that is

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

Then the equation (3.27) 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 $f = c_1 f_1 + ... + c_k f_k$ and $x_1(t), ..., x_k(t)$ are solutions to the equation $P\left(\frac{d}{dt}\right)x_j = f_j$, then $x = c_1 x_1 + ... + c_k x_k$ solves the equation $P\left(\frac{d}{dt}\right)x = f$.*

Proof. This is trivial because

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

■

Hence, we can assume that the function f in (3.27) is of the form $f(t) = R(t)e^{\mu t}$ where $R(t)$ is a polynomial.

To illustrate the method, which will be used in this Section, consider first the following example.

Example. Consider the ODE

$$P\left(\frac{d}{dt}\right)x = e^{\mu t} \tag{3.28}$$

where μ is *not* a root of the characteristic polynomial $P(\lambda)$ (*non-resonant case*). We claim that (3.28) 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 (3.15)

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

whence

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

provided

$$a = \frac{1}{P(\mu)}. \tag{3.29}$$

Consider some concrete examples of ODE. 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 (3.29), 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:

$$x'' + 2x' + x = \sin t \tag{3.30}$$

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 (3.30).

Consider yet another ODE

$$x'' + 2x' + x = e^{-t}\cos t. \tag{3.31}$$

Here $e^{-t}\cos t$ is a real part of $e^{\mu t}$ where $\mu = -1 + i$. Hence, first solve

$$x'' + 2x' + x = e^{\mu t}.$$

Setting $x(t) = ae^{\mu t}$, we obtain

$$a = \frac{1}{P(\mu)} = \frac{1}{(-1+i)^2 + 2(-1+i) + 1} = -1.$$

Hence, the complex solution is $x(t) = -e^{(-1+i)t} = -e^{-t}\cos t - ie^{-t}\sin t$, and the solution to (3.31) is $x(t) = -e^{-t}\cos t$.

Finally, let us combine the above examples into one:

$$x'' + 2x' + x = 2e^t - \sin t + e^{-t}\cos t. \tag{3.32}$$

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 (3.32)

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

Consider one more equation

$$x'' + 2x' + x = e^{-t}.$$

This time $\mu = -1$ is a root of $P(\lambda) = \lambda^2 + 2\lambda + 1$ and the above method does not work. Indeed, if we look for a solution in the form $x = ae^{-t}$ then after substitution we get 0 in the left hand side because e^{-t} solves the homogeneous equation.

The case when μ is a root of $P(\lambda)$ is referred to as a *resonance*. This case as well as the case of the general quasi-polynomial in the right hand side is treated in the following theorem.

Theorem 3.8 *Let $R(t)$ be a non-zero polynomial of degree $k \geq 0$ and μ be a complex number. Let m be the multiplicity of μ if μ is a root of P and $m = 0$ if μ is not a root of P. Then the equation*

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

has a solution of the form

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

where $Q(t)$ is a polynomial of degree k (which is to be found).

Example. Come back to the equation

$$x'' + 2x' + x = e^{-t}.$$

Here $\mu = -1$ is a root of multiplicity $m = 2$ and $R(t) = 1$ is a polynomial of degree 0. Hence, the solution should be sought in the form

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

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

$$a\left(\left(t^2 e^{-t}\right)'' + 2\left(t^2 e^{-t}\right)' + t^2 e^{-t}\right) = e^{-t} \tag{3.33}$$

Expanding the expression in the brackets, we obtain the identity

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

so that (3.33) becomes $2a = 1$ and $a = \frac{1}{2}$. Hence, a particular solution is

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

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

$$x(t) = (at + b)\, t^2 e^{-t} = \left(at^3 + bt^2\right) e^{-t}$$

into the equation, we obtain

$$\begin{aligned} x'' + 2x' + x &= \left(\left(at^3 + bt^2\right) e^{-t}\right)'' + 2\left(\left(at^3 + bt^2\right) e^{-t}\right)' + \left(at^3 + bt^2\right) e^{-t} \\ &= (2b + 6at)\, e^{-t}. \end{aligned}$$

Hence, comparing with the equation, we obtain

$$2b + 6at = t$$

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

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

Proof of Theorem 3.8. Let us prove that the equation

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

has a solution in the form

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

where m is the multiplicity of μ and $\deg Q = k = \deg R$. Using Lemma 3.4, we have

$$\begin{aligned} P\left(\frac{d}{dt}\right) x &= P\left(\frac{d}{dt}\right)\left(t^m Q(t)\, e^{\mu t}\right) = \sum_{j \geq 0} \frac{1}{j!} \left(t^m Q(t)\right)^{(j)} P^{(j)}\left(\frac{d}{dt}\right) e^{\mu t} \\ &= \sum_{j \geq 0} \frac{1}{j!} \left(t^m Q(t)\right)^{(j)} P^{(j)}(\mu)\, e^{\mu t}. \end{aligned} \tag{3.34}$$

By Lemma 3.4, 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 $P^{(j)}(\mu) = 0$ for all $j \leq m - 1$, we can restrict the summation to $j \geq m$. Set

$$y(t) = \left(t^m Q(t)\right)^{(m)} \tag{3.35}$$

and observe that $y(t)$ is a polynomial of degree k, provided so is $Q(t)$. Conversely, for any polynomial $y(t)$ of degree k, there is a polynomial $Q(t)$ of degree k such that (3.35) holds. Indeed, integrating (3.35) m times without adding constants and then dividing by t^m, we obtain $Q(t)$ as a polynomial of degree k.

It follows from (3.34) that y must satisfy the ODE

$$\frac{P^{(m)}(\mu)}{m!} y + \frac{P^{(m+1)}(\mu)}{(m+1)!} y' + ... + \frac{P^{(m+i)}(\mu)}{(m+i)!} y^{(m+i)} + ... = R(t),$$

which we rewrite in the form

$$b_0 y + b_1 y' + ... + b_i y^{(i)} + ... = R(t) \tag{3.36}$$

where $b_i = \frac{P^{(m+i)}(\mu)}{(m+i)!}$ (in fact, the index i in the left hand side of (3.36) can be restricted to $i \le k$ since $y^{(i)} \equiv 0$ for $i > k$). Note that

$$b_0 = \frac{P^{(m)}(\mu)}{m!} \neq 0. \tag{3.37}$$

Hence, the problem amounts to the following: given a polynomial

$$R(t) = r_0 t^k + r_1 t^{k-1} + ... + r_k$$

of degree k, prove that there exists a polynomial $y(t)$ of degree k that satisfies (3.36). Let us prove the existence of y by induction in k.

The inductive basis. If $k = 0$, then $R(t) \equiv r_0$ and $y(t) \equiv a$, so that (3.36) becomes $ab_0 = r_0$ whence $a = r_0/b_0$ (where we use that $b_0 \neq 0$).

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

$$y = at^k + z(t), \tag{3.38}$$

where z is a polynomial of degree $< k$. Substituting (3.38) into (3.36), we obtain the equation for z

$$b_0 z + b_1 z' + ... + b_i z^{(i)} + ... = R(t) - \left(ab_0 t^k + ab_1 \left(t^k\right)' + ... + ab_k \left(t^k\right)^{(k)}\right) =: \widetilde{R}(t).$$

Choosing a from the equation $ab_0 = r_0$ we obtain that the term t^k in the right hand side of (3.38) cancels out, whence it follows that $\widetilde{R}(t)$ is a polynomial of degree $< k$. By the inductive hypothesis, the equation

$$b_0 z + b_1 z' + ... + b_i z^{(i)} + ... = \widetilde{R}(t)$$

has a solution $z(t)$ which is a polynomial of degree $< k$. Hence, the function $y = at^k + z$ solves (3.36) and is a polynomial of degree k. ■

Remark. If $k = 0$, that is, $R(t) \equiv r_0$ is a constant then (3.36) yields

$$y = \frac{r_0}{b_0} = \frac{m! r_0}{P^{(m)}(\mu)}.$$

The equation (3.35) becomes

$$\left(t^m Q(t)\right)^{(m)} = \frac{m! r_0}{P^{(m)}(\mu)}$$

whence after m integrations we find

$$Q(t) = \frac{r_0}{P^{(m)}(\mu)}.$$

Therefore, the ODE $P\left(\frac{d}{dt}\right)x = r_0 e^{\mu t}$ has a particular solution

$$x(t) = \frac{r_0}{P^{(m)}(\mu)} t^m e^{\mu t}. \tag{3.39}$$

Example. Consider again the ODE $x'' + 2x' + x = e^{-t}$. Then $\mu = -1$ has multiplicity $m = 2$, and $R(t) \equiv 1$. Hence, by the above Remark, we find a particular solution

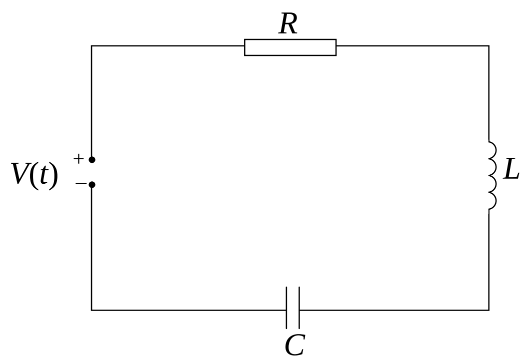
$$x(t) = \frac{1}{P''(-1)} t^2 e^{-t} = \frac{1}{2} t^2 e^{-t}.$$

3.5 Second order ODE with periodic right hand side

Consider a second order ODE

$$x'' + px' + qx = f(t), \tag{3.40}$$

which occurs in various physical phenomena. For example, (3.40) 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 (3.40), is an electrical circuit:



As before, let R the resistance, L be the inductance, and C be the capacitance of the circuit. Let $V(t)$ be the voltage of the power source in the circuit and $x(t)$ be the current in the circuit at time t. Then we have seen that the equation for $x(t)$ is

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

If $L > 0$ then dividing by L we obtain an ODE of the form (3.40).

As an example of application of the above methods of solving such ODEs, we investigate here the case when function $f(t)$ is periodic. More precisely, consider the ODE

$$x'' + px' + qx = A \sin \omega t, \tag{3.41}$$

where A, ω are given positive reals. The function $A \sin \omega t$ is a model for a more general periodic force, which makes good physical sense in all the above examples. For example, in the case of electrical circuit the external force has the form $A \sin \omega t$ if the power source is an electrical socket with the alternating current (AC). The number ω is called the *frequency* of the external force (note that the period $= \frac{2\pi}{\omega}$) or the external frequency, and the number A is called the *amplitude* (the maximum value) of the external force.

Assume in the sequel that $p \geq 0$ and $q > 0$, which is physically most interesting case. To find a particular solution of (3.41), let us consider the ODE with complex right hand side:

$$x'' + px' + qx = Ae^{i\omega t}. \tag{3.42}$$

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

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

and the particular solution of (3.42) is

$$(a + ib) e^{i\omega t} = (a \cos \omega t - b \sin \omega t) + i (a \sin \omega t + b \cos \omega t).$$

Taking its imaginary part, we obtain a particular solution to (3.41)

$$x(t) = a \sin \omega t + b \cos \omega t = B \sin (\omega t + \varphi) \tag{3.43}$$

where

$$B = \sqrt{a^2 + b^2} = |c| = \frac{A}{\sqrt{(q - \omega^2)^2 + \omega^2 p^2}} \tag{3.44}$$

and $\varphi \in [0, 2\pi)$ 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 the *phase*.

To obtain the general solution to (3.41), we need to add to (3.43) the general solution to the homogeneous equation

$$x'' + px' + qx = 0.$$

Let λ_1 and λ_2 are the roots of $P(\lambda)$, that is,

$$\lambda_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}.$$

Consider the following possibilities for the roots.

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

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

In the both cases, it decays exponentially in t as $t \to +\infty$. Hence, the general solution of (3.41) 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 also periodic with the same frequency ω as the external force.

λ_1 *and* λ_2 *are complex.*

Let $\lambda_{1,2} = \alpha \pm i\beta$ where

$$\alpha = -p/2 \le 0 \quad \text{and} \quad \beta = \sqrt{q - \frac{p^2}{4}} > 0.$$

The general solution to the homogeneous equation is

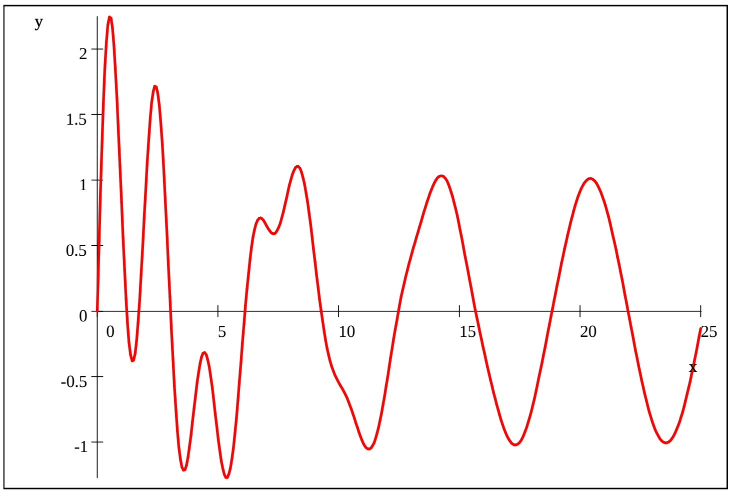
$$e^{\alpha t}(C_1 \cos\beta t + C_2 \sin\beta t) = Ce^{\alpha t}\sin(\beta t + \psi).$$

The number β is called the *natural frequency* of the physical system in question (pendulum, electrical circuit, spring) for the obvious reason - in absence of the external force, the system oscillate with the natural frequency β.

Hence, the general solution to (3.41) is

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

If $\alpha < 0$ then the leading term is again $B\sin(\omega t + \varphi)$. Here is a particular example of such a function: $\sin t + 2e^{-t/4}\sin \pi t$



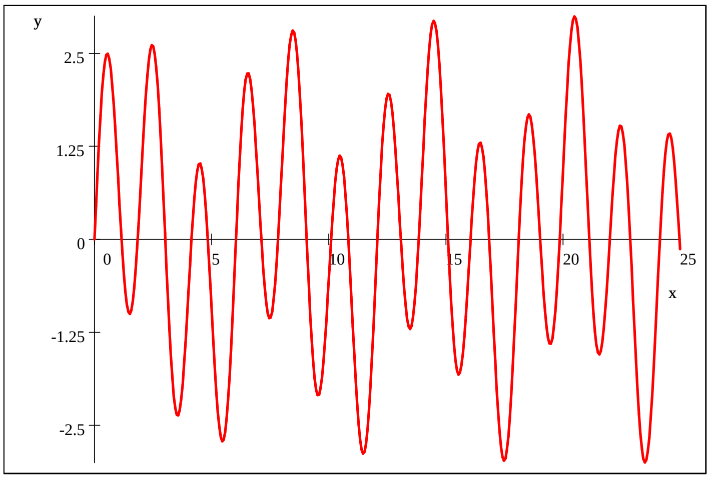
λ_1 *and* λ_2 *are purely imaginary, that is,* $\alpha = 0$. In this case, $p = 0$, $q = \beta^2$, and the equation has the form

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

The assumption that $i\omega$ is not a root implies $\omega \neq \beta$. The general solution is

$$x(t) = B\sin(\omega t + \varphi) + C\sin(\beta t + \psi),$$

which is the sum of two sin waves with different frequencies - the natural frequency and the external frequency. Here is a particular example of such a function: $\sin t + 2\sin \pi t$:



Strictly speaking, in practice such electrical circuits do not occur since the resistance is always positive.

Let us come back to the formula (3.44) 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 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 then the system exhibits the highest response to this force. This phenomenon is called a *resonance*.

Note for comparison that the natural frequency is equal to $\beta = \sqrt{q - p^2/4}$, which is in general different from ω_0. In terms of ω_0 and β, we can write

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

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 when $\omega = \omega_0$ is $B_{\max} = \frac{A}{p\beta}$.

In conclusion, consider the case, when $i\omega$ is a root of $P(\lambda)$, that is

$$(i\omega)^2 + pi\omega + q = 0,$$

which implies $p = 0$ and $q = \omega^2$. In this case $\alpha = 0$ and $\omega = \omega_0 = \beta = \sqrt{q}$, and the equation has the form

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

Considering the ODE

$$x'' + \omega^2 x = Ae^{i\omega t},$$

and searching a particular solution in the form $x(t) = cte^{i\omega t}$, we obtain by (3.39)

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

Hence, the complex particular solution is

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

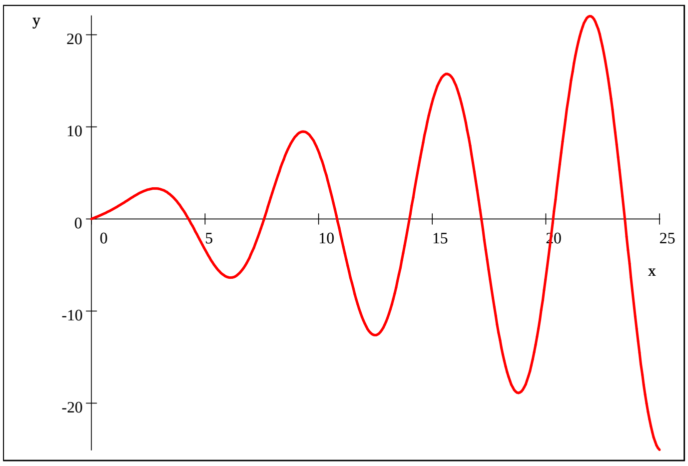
and its imaginary part is

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

Hence, the general solution is

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

Here is an example of such a function: $-t\cos t + 2\sin t$



Hence, we have a *complete resonance*: the external frequency ω is simultaneously equal to the natural frequency and the resonant frequency. In the case of a complete resonance, the amplitude increases in time unbounded. Since unbounded oscillations are physically impossible, either the system breaks down over time or the mathematical model becomes unsuitable for describing the physical system.

3.6 The method of variation of parameters

3.6.1 A system of the 1st order

We present here the method of variation of parameters in order to solve a general linear system

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

where as before $A(t) : I \to \mathbb{R}^{n\times n}$ and $B(t) : I \to \mathbb{R}^n$ are continuous. Let $x_1(t), ..., x_n(t)$ be n linearly independent solutions of the homogeneous system $x' = A(t)\,x$, defined on I. We start with the following observation.

Lemma 3.9 *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. Indeed, assume that for some constant $C_1, ..., C_n$

$$C_1x_1(t_0) + ... + C_nx_n(t_0) = 0.$$

Consider the function $x(t) = C_1x_1(t) + ... + C_nx_n(t)$. 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 $x' = A(t)x$, 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 3.9 that the column of $X(t)$ are linearly independent for any $t \in I$, which in particular means that the inverse matrix $X^{-1}(t)$ is also defined for all $t \in I$. This allows us to solve the inhomogeneous system as follows.

Theorem 3.10 *The general solution to the system*

$$x' = A(t)x + B(t), \tag{3.45}$$

is given by

$$x(t) = X(t) \int X^{-1}(t) B(t) \, dt, \tag{3.46}$$

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

Note that $X^{-1}B$ is a time dependent n-dimensional vector, which can be integrated in t componentwise.

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 (3.46) 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, $x(t)$ solves (3.45). Let us show that (3.46) gives all the solutions. Note that the integral in (3.46) is indefinite so that it can be presented in the form

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

where $V(t)$ is a vector function and $C = (C_1, ..., C_n)$ is an arbitrary constant vector. Hence, (3.46) gives

$$\begin{aligned} x(t) &= X(t) V(t) + X(t) C \\ &= x_0(t) + C_1 x_1(t) + ... + C_n x_n(t), \end{aligned}$$

where $x_0(t) = X(t) V(t)$ is a solution of (3.45). By Theorem 3.7 we conclude that $x(t)$ is indeed the general solution. ∎

Second proof. Let us show a different way of derivation of (3.46) that is convenient in practical applications and also explains the term "variation of parameters". Let us look for a solution to (3.45) in the form

$$x(t) = C_1(t)\,x_1(t) + \ldots + C_n(t)\,x_n(t) \tag{3.47}$$

where $C_1, C_2, .., C_n$ are now unknown real-valued functions to be determined. Since $x_1(t), \ldots, x_n(t)$ are for any t linearly independent vectors, any $\mathbb{R}^n$-valued function $x(t)$ can be represented in the form (3.47). The identity (3.47) can be considered as a linear system of algebraic equations with respect to the unknowns $C_1, \ldots, C_n$. Solving it by Cramer's rule, we obtain $C_1, \ldots, C_n$ in terms of rational functions of $x_1, \ldots, x_n, x$. Since the latter functions are all differentiable in t, we obtain that also $C_1, \ldots, C_n$ are differentiable in t.

Differentiating the identity (3.47) in time and using $x_k' = Ax_k$, we obtain

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

Hence, the equation $x' = Ax + B$ becomes

$$C_1'x_1 + C_2'x_2 + \ldots + C_n'x_n = B. \tag{3.48}$$

If $C(t)$ denotes the column-vector with components $C_1(t), \ldots, C_n(t)$ then (3.48) can be written in the form

$$XC' = B$$

whence

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

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

and

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

∎

The term "variation of parameters" comes from the identity (3.47). Indeed, if $C_1, \ldots., C_n$ are constant parameters then this identity determines the general solution of the homogeneous ODE $x' = Ax$. By allowing $C_1, \ldots, C_n$ to be variable, we obtain the general solution to $x' = Ax + B$.

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

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

Hence, the corresponding fundamental matrix is

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

and

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

Consider now the ODE

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

where $B(t) = \begin{pmatrix} b_1(t) \\ b_2(t) \end{pmatrix}$. By (3.46), we obtain the general solution

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

Consider a particular example $B(t) = \begin{pmatrix} 1 \\ -t \end{pmatrix}$. Then the integral is

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

whence

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

3.6.2 A scalar ODE of n-th order

Consider now a scalar ODE of order n

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

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 solution, we obtain the fundamental matrix

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

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

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

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 solution to*

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

and X be the corresponding fundamental matrix. Then, for any continuous function $f(t)$, the general solution to the ODE

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

is given by

$$x(t) = \sum_{j=1}^{n} x_j(t) \int f(t)\, y_{jn}(t)\, dt \tag{3.49}$$

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

Example. Consider the ODE

$$x'' + x = \sin t$$

The independent solutions are $x_1(t) = \cos t$ and $x_2(t) = \sin t$, so that

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

The inverse is

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

Hence, the solution is

$$\begin{aligned}
x(t) &= x_1(t) \int f(t)\, y_{12}(t)\, dt + x_2(t) \int f(t)\, y_{22}(t)\, dt \\
&= \cos t \int \sin t\, (-\sin t)\, dt + \sin t \int \sin t \cos t dt \\
&= -\cos t \int \sin^2 t dt + \frac{1}{2} \sin t \int \sin 2t dt \\
&= -\cos t \left(\frac{1}{2} t - \frac{1}{4} \sin 2t + C_1 \right) + \frac{1}{4} \sin t\, (-\cos 2t + C_2) \\
&= -\frac{1}{2} t \cos t + \frac{1}{4} (\sin 2t \cos t - \sin t \cos 2t) + C_3 \cos t + C_4 \sin t \\
&= -\frac{1}{2} t \cos t + C_3 \cos t + C_5 \sin t.
\end{aligned}$$

Of course, the same result can be obtained by Theorem 3.8.

Consider one more example, when the right hand side is not a quasi-polynomial:

$$x'' + x = \tan t. \tag{3.50}$$

Then as above we obtain[8]

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

[8] 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

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

Let us show how one can use the method of variation of parameters directly, without using the formula (3.49). Consider the ODE

$$x'' + x = f(t). \tag{3.51}$$

The general solution to the homogeneous ODE $x'' + x = 0$ is

$$x(t) = C_1 \cos + C_2 \sin t, \tag{3.52}$$

where C_1 and C_2 are constant parameters. let us look for the solution of (3.50) in the form

$$x(t) = C_1(t)\cos t + C_2(t)\sin t, \tag{3.53}$$

which is obtained from (3.52) by replacing the constants by functions (hence, the name of the method "variation of parameters"). To obtain the equations for the unknown functions $C_1(t), C_2(t)$, differentiate (3.53):

$$\begin{aligned} x'(t) &= -C_1(t)\sin t + C_2(t)\cos t \\ &\quad + C_1'(t)\cos t + C_2'(t)\sin t. \end{aligned} \tag{3.54}$$

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,

$$C_1' \cos t + C_2' \sin t = 0. \tag{3.55}$$

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 (3.55).

It follows from (3.54) and (3.55) 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

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

and

$$x(t) = -\cos t \int f(t)\sin t dt + \sin t \int f(t)\cos t dt.$$

3.7 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 3.11 (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 system 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 3.9, 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 3.12 (The Liouville formula) *Let $\{x_i\}_{i=1}^n$ be a sequence of n solutions of the ODE $x' = A(t)\,x$, where $A : I \to \mathbb{R}^{n\times n}$ is continuous. Then the Wronskian $W(t)$ of this sequence satisfies the identity*

$$W(t) = W(t_0)\exp\left(\int_{t_0}^{t} \operatorname{trace} A(\tau)\,d\tau\right), \tag{3.56}$$

for all $t, t_0 \in I$.

Recall that the trace (*Spur*) $\operatorname{trace} A$ of the matrix A is the sum of all the diagonal entries of the matrix.

Proof. 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:

$$W'(t) = \det\begin{pmatrix} r_1' \\ r_2 \\ ... \\ r_n \end{pmatrix} + \det\begin{pmatrix} r_1 \\ r_2' \\ ... \\ r_n \end{pmatrix} + ... + \det\begin{pmatrix} r_1 \\ r_2 \\ ... \\ r_n' \end{pmatrix}. \tag{3.57}$$

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, (3.57).

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

$$x_{ij}' = \sum_{k=1}^{n} A_{ik}x_{kj}. \tag{3.58}$$

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, (3.58) 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,

$$r_1' = A_{11} r_1 + A_{12} r_2 + \ldots + A_{1n} r_n$$

which implies that

$$\det \begin{pmatrix} r_1' \\ r_2 \\ \ldots \\ r_n \end{pmatrix} = A_{11} \det \begin{pmatrix} r_1 \\ r_2 \\ \ldots \\ r_n \end{pmatrix} + A_{12} \det \begin{pmatrix} r_2 \\ r_2 \\ \ldots \\ r_n \end{pmatrix} + \ldots + A_{1n} \det \begin{pmatrix} r_n \\ r_2 \\ \ldots \\ r_n \end{pmatrix}.$$

All the determinants except for the 1st one vanish since they have equal rows. Hence,

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

Evaluating similarly the other terms in (3.57), we obtain

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

By Lemma 3.11, $W(t)$ 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 (3.56) follows. ∎

Corollary. *Consider a scalar ODE*

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

where $a_k(t)$ are continuous functions on an interval $I \subset \mathbb{R}$. If $x_1(t), \ldots, x_n(t)$ are n solutions to this equation then their Wronskian $W(t)$ satisfies the identity

$$W(t) = W(t_0) \exp\left(-\int_{t_0}^{t} a_1(\tau)\, d\tau\right). \tag{3.59}$$

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

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

Since the Wronskian of the normal system coincides with $W(t)$, (3.59) follows from (3.56) because $\operatorname{trace} A = -a_1$. $\blacksquare$

In the case of the ODE of the 2nd order

$$x'' + a_1(t)\,x' + a_2(t)\,x = 0$$

the Liouville formula can help in finding the general solution if a particular solution is known. Indeed, if $x_0(t)$ is a particular non-zero solution and $x(t)$ is any other solution then we have by (3.59)

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

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

$$\left(\frac{x}{x_0}\right)' = \frac{C\exp\left(-\int a_1(t)\,dt\right)}{x_0^2}, \tag{3.60}$$

and by integrating it we obtain $\frac{x}{x_0}$ and, hence, x (cf. Exercise 35).

Example. Consider the ODE

$$x'' - 2\left(1+\tan^2 t\right)x = 0.$$

One solution can be guessed $x_0(t) = \tan t$ using the fact that

$$\frac{d}{dt}\tan t = \frac{1}{\cos^2 t} = \tan^2 t + 1$$

and

$$\frac{d^2}{dt^2}\tan t = 2\tan t\left(\tan^2 t + 1\right).$$

Hence, for $x(t)$ we obtain from (3.60)

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

whence[9]

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

Renaming the constants, we obtain the answer

$$x(t) = C_1\left(t\tan t + 1\right) + C_2\tan t.$$

[9] To evaluate the integral $\int\frac{dt}{\tan^2 t} = \int\cot^2 t\,dt$ use the identity

$$(\cot t)' = -\cot^2 t - 1$$

that yields

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

3.8 Linear homogeneous systems with constant coefficients

Here we will be concerned with finding the general solution to linear systems of the form $x' = Ax$ where $A \in \mathbb{C}^{n\times n}$ 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, it suffices to find n linearly independent solutions and then take their linear combination. We start with a simple observation. Let us try to find a solution in the form $x = e^{\lambda t}v$ where v is a non-zero vector in $\mathbb{C}^n$ that does not depend on t. Then the equation $x' = Ax$ becomes

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

that is, $Av = \lambda v$. 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*. Hence, 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.

The fact that λ is an eigenvalue means that the matrix $A - \lambda\,\mathrm{id}$ is not invertible, that is,

$$\det(A - \lambda\,\mathrm{id}) = 0. \tag{3.61}$$

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

$$(A - \lambda\,\mathrm{id})\, v = 0. \tag{3.62}$$

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

The function

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

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

Lemma 3.13 *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 general complex solution of the ODE $x' = Ax$ is given by*

$$x(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t} v_k, \tag{3.63}$$

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 (3.63) *can be replaced by the couple* $\mathrm{Re}\left(e^{\lambda t}v\right)$, $\mathrm{Im}\left(e^{\lambda t}v\right)$.

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 3.1.

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

$$\mathrm{Re}\, e^{\lambda t} v = \frac{e^{\lambda t}v + e^{\overline{\lambda} t}\overline{v}}{2} \quad \text{and} \quad \mathrm{Im}\, e^{\lambda t} v = \frac{e^{\lambda t}v - e^{\overline{\lambda} t}\overline{v}}{2i}$$

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 (3.63) the terms $e^{\lambda t}, e^{\overline{\lambda}t}$ by the couple $\operatorname{Re}\left(e^{\lambda t}v\right)$, $\operatorname{Im}\left(e^{\lambda t}v\right)$ does not change the set of functions, which finishes the proof. ■

It is known from Linear Algebra that if A has n distinct eigenvalues then their eigenvectors are automatically linearly independent, and Lemma 3.13 applies. Or if A is a symmetric matrix then there is a basis of eigenvectors, and Lemma 3.13 applies.

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 $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. 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 (3.62) 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 $v = \binom{a}{b}$

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

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

$$\mathbf{x}_1(t) = e^{it} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin t + i \cos t \\ \cos t + i \sin t \end{pmatrix}.$$

Since this solution is complex, we obtain the general solution using the second claim of Lemma 3.13:

$$\mathbf{x}(t) = C_1 \operatorname{Re} \mathbf{x}_1 + C_2 \operatorname{Im} \mathbf{x}_1 = C_1 \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} + C_2 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} = \begin{pmatrix} -C_1 \sin t + C_2 \cos t \\ C_1 \cos t + C_2 \sin t \end{pmatrix}.$$

Example. Consider a normal system

$$\begin{cases} x' = y \\ y' = 0. \end{cases}$$

This system is trivially solved to obtain $y = C_1$ and $x = C_1 t + C_2$. However, if we try to solve it using the above method, we fail. Indeed, the matrix of the system is $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, the characteristic polynomial is

$$P(\lambda) = \det \begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix} = \lambda^2,$$

and the characteristic equation $P(\lambda) = 0$ yields only one eigenvalue $\lambda = 0$. The eigenvector $v = \binom{a}{b}$ satisfies the equation

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

whence $b = 0$. That is, the only eigenvector (up to a constant multiple) is $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, and the only solution we obtain in this way is $\mathbf{x}(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The problem lies in the properties of this matrix – it does not have a basis of eigenvectors, which is needed for this method.

In order to handle such cases, we use a different approach.

3.8.1 Functions of operators and matrices

Recall that an scalar ODE $x' = Ax$ has a solution $x(t) = Ce^{At}t$. Now if A is a $n \times n$ matrix, we may be able to use this formula if we define what is e^{At}. It suffices to define what is e^A for any matrix A. It is convenient to do this for linear operators acting in $\mathbb{C}^n$.

Recall that a *linear operator* in $\mathbb{C}^n$ is a mapping $A : \mathbb{C}^n \to \mathbb{C}^n$ such that, for all $x, y \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$,

$$\begin{aligned} A(x+y) &= Ax + Ay \\ A(\lambda x) &= \lambda Ax. \end{aligned}$$

Any $n \times n$ matrix defines a linear operator in $\mathbb{C}^n$ using multiplication of column-vectors by this matrix. Moreover, any linear operator can be represented in this form so that there is an one-to-one correspondence[10] between linear operators and matrices.

Denote the family of all linear operators in $\mathbb{C}^n$ by $\mathcal{L}(\mathbb{C}^n)$. For any two operators A, B, define their sum $A + B$ by

$$(A+B)x = Ax + Bx$$

and the product by a scalar $\lambda \in \mathbb{C}$ by

$$(\lambda A)(x) = \lambda Ax,$$

for all $x \in \mathbb{C}^n$. With these operation, $\mathcal{L}(\mathbb{C}^n)$ is a linear space over $\mathbb{C}$. Since any operator can be identified with a $n \times n$ matrix, the dimension of the linear space $\mathcal{L}(\mathbb{C}^n)$ is n^2.

Apart from the linear structure, the product AB of operators is defined in $\mathcal{L}(\mathbb{C}^n)$ as composition that is,

$$(AB)x = A(Bx).$$

Fix a norm $\|\cdot\|$ in $\mathbb{C}^n$, for example, the ∞-norm

$$\|x\|_\infty := \max_{1 \le k \le n} |x_k|$$

where $x_1, ..., x_n$ are the components of the vector x. Define the associated *operator norm* in $\mathcal{L}(\mathbb{C}^n)$ by

$$\|A\| = \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{3.64}$$

Claim. *The operator norm is a norm in* $\mathcal{L}(\mathbb{C}^n)$.

Proof. Let us first show that $\|A\|$ is finite. Represent A is a matrix (A_{kj}) in the standard basis. Since all norms in any finitely dimensional linear space are equivalent, we can assume in the sequel that $\|x\| = \|x\|_\infty$. Then

$$\begin{aligned} \|Ax\|_\infty &= \max_k |(Ax)_k| = \max_k \left| \sum_j A_{kj} x_j \right| \\ &\le \max_k \left| \sum_j A_{kj} \right| \max_j |x_j| = C\|x\|_\infty, \end{aligned}$$

where $C < \infty$. Therefore, $\|A\| \le C < \infty$.

[10] This correspondence depends on the choice of a basis in $\mathbb{C}^n$.

2. Clearly, $\|A\| \geq 0$. Let us show that $\|A\| > 0$ if $A \neq 0$. Indeed, if $A \neq 0$ then there is $x \in \mathbb{C}^n$ such that $Ax \neq 0$ and $\|Ax\| > 0$, whence

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

3. Let us prove the triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$. Indeed, by definition (3.64)

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

4. Let us prove the scaling property: $\|\lambda A\| = |\lambda| \, \|A\|$ for any $\lambda \in \mathbb{C}$. Indeed, by (3.64)

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

■

In addition to the general properties of a norm, the operator norm satisfies the inequality

$$\|AB\| \leq \|A\| \, \|B\| \,. \tag{3.65}$$

Indeed, it follows from (3.64) that $\|Ax\| \leq \|A\| \, \|x\|$ whence

$$\|(AB)x\| = \|A(Bx)\| \leq \|A\| \, \|Bx\| \leq \|A\| \, \|B\| \, \|x\|$$

which yields (3.65).

Hence, $\mathcal{L}(\mathbb{C}^n)$ is a normed linear space. Since this space is finite dimensional, it is complete as a normed space. As in any complete normed linear space, one can define in $\mathcal{L}(\mathbb{C}^n)$ the notion of the limit of a sequence of operators. Namely, we say that a sequence $\{A_k\}$ of operators converges to an operator A if

$$\|A_k - A\| \to 0 \text{ as } k \to \infty.$$

Representing an operator A as a matrix $(A_{ij})_{i,j=1}^n$, one can consider the ∞-norm on operators defined by

$$\|A\|_\infty = \max_{1 \leq i,j \leq n} |A_{ij}| \,.$$

Clearly, the convergence in the ∞-norm is equivalent to the convergence of each component A_{ij} separately. Since all norms in $\mathcal{L}(\mathbb{C}^n)$ are equivalent, we see that convergence of a sequence of operators in any norm is equivalent to the convergence of the individual components of the operators.

Given a series $\sum_{k=1}^\infty A_k$ of operators, the sum of the series is defined as the limit of the sequence of partial sums $\sum_{k=1}^N A_k$ as $N \to \infty$. That is, $S = \sum_{k=1}^\infty A_k$ if

$$\|S - \sum_{k=1}^N A_k\| \to 0 \quad \text{as } N \to \infty.$$

Claim. *Assume that*

$$\sum_{k=1}^{\infty} \|A_k\| < \infty. \tag{3.66}$$

Then the series $\sum_{k=1}^{\infty} A_k$ *converges.*

Proof. Indeed, since all norms in $\mathcal{L}(\mathbb{C}^n)$ are equivalent, we can assume that the norm in (3.66) is the ∞-norm. Denoting by $(A_k)_{ij}$ the ij-components of the matrix A, we obtain that then the condition (3.66) is equivalent to

$$\sum_{k=1}^{\infty} \left|(A_k)_{ij}\right| < \infty \tag{3.67}$$

for any indices $1 \leq i, j \leq n$. Then (3.67) implies that the numerical series

$$\sum_{k=1}^{\infty} (A_k)_{ij}$$

converges, which implies that the operator series $\sum_{k=1}^{\infty} A_k$ also converges. ■

If the condition (3.66) is satisfied then the series $\sum_{k=1}^{\infty} A_k$ *is called absolutely convergent.*

Hence, the above Claim means that absolute convergence of an operator series implies the usual convergence.

Definition. If $A \in \mathcal{L}(\mathbb{C}^n)$ then define $e^A \in \mathcal{L}(\mathbb{C}^n)$ by means of the identity

$$e^A = \mathrm{id} + A + \frac{A^2}{2!} + \ldots + \frac{A^k}{k!} + \ldots = \sum_{k=0}^{\infty} \frac{A^k}{k!}, \tag{3.68}$$

where id is the identity operator.

Of course, in order to justify this definition, we need to verify the convergence of the series (3.68).

Lemma 3.14 *The exponential series* (3.68) *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 (3.65) 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 3.15 *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 3.14) 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 theorem. ■

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 and, hence, can be used to form the fundamental matrix. In particular, choosing $v_1, ..., v_n$ to be the canonical basis in $\mathbb{C}^n$, we obtain that $e^{tA}v_k$ is the k-th column of the matrix e^{tA}. Hence, the matrix e^{tA} is itself a fundamental matrix of the system $x' = Ax$.

Example. Let A be the diagonal matrix

$$A = \operatorname{diag}(\lambda_1, ..., \lambda_n).$$

Then

$$A^k = \operatorname{diag}\left(\lambda_1^k, ..., \lambda_n^k\right)$$

and

$$e^{tA} = \operatorname{diag}\left(e^{\lambda_1 t}, ..., e^{\lambda_n t}\right).$$

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,

$$(A + B)^2 = A^2 + 2AB + B^2. \tag{3.69}$$

Indeed, in general we have

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2,$$

which yields (3.69) if $AB = BA$.

Lemma 3.16 *If A and B commute then*

$$e^{A+B} = e^A e^B.$$

Proof. Let us prove a sequence of claims.

Claim 1. *If A, B, C commute pairwise then so do AC and B.*

Indeed,

$$(AC)B = A(CB) = A(BC) = (AB)C = (BA)C = B(AC).$$

Claim 2. *If A and B commute then so do e^A and B.*

Indeed, it follows from Claim 1 that A^k and B commute for any natural k, whence

$$e^A B = \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) B = B \left(\sum_{k=0}^{\infty} \frac{A^k}{k!} \right) = B e^A.$$

Claim 3. *If $A(t)$ and $B(t)$ are differentiable functions from $\mathbb{R}$ to $\mathcal{L}(\mathbb{C}^n)$ then*

$$(A(t) B(t))' = A'(t) B(t) + A(t) B'(t). \tag{3.70}$$

Warning: watch the correct order of the multiples.

Indeed, we have for any component

$$(AB)'_{ij} = \left(\sum_k A_{ik} B_{kj} \right)' = \sum_k A'_{ik} B_{kj} + \sum_k A_{ik} B'_{kj} = (A'B)_{ij} + (AB')_{ij} = (A'B + AB')_{ij},$$

whence (3.70) follows.

Now we can finish the proof of the lemma. Consider the function $F:\mathbb{R}\to\mathcal{L}\left(\mathbb{C}^n\right)$ defined by
$$F\left(t\right)=e^{tA}e^{tB}.$$
Differentiating it using Theorem 3.15, Claims 2 and 3, we obtain
$$F'\left(t\right)=\left(e^{tA}\right)'e^{tB}+e^{tA}\left(e^{tB}\right)'=Ae^{tA}e^{tB}+e^{tA}Be^{tB}=Ae^{tA}e^{tB}+Be^{tA}e^{tB}=\left(A+B\right)F\left(t\right).$$
On the other hand, by Theorem 3.15, the function $G\left(t\right)=e^{t(A+B)}$ satisfies the same equation
$$G'=\left(A+B\right)G.$$
Since $G\left(0\right)=F\left(0\right)=\mathrm{id}$, we obtain that the vector functions $F\left(t\right)$ and $G\left(t\right)$ solve the same IVP, whence by the uniqueness theorem they are identically equal. In particular, $F\left(1\right)=G\left(1\right)$, 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
$$\left(A+B\right)^n=\sum_{k=0}^{n}\binom{n}{k}A^kB^{n-k}$$
using the fact that A and B commute (this can be done by induction in the same way as for numbers). Then we have
$$e^{A+B}=\sum_{n=0}^{\infty}\frac{\left(A+B\right)^n}{n!}=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{A^kB^{n-k}}{k!\left(n-k\right)!}$$
and, using the Cauchy product formula,
$$e^Ae^B=\sum_{m=0}^{\infty}\frac{A^m}{m!}\sum_{l=0}^{\infty}\frac{B^l}{l!}=\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{A^kB^{n-k}}{k!\left(n-k\right)!}.$$
Of course, one need to justify the Cauchy product formula for absolutely convergent series of operators. ■

3.8.2 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
$$A=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0\\ 0 & \lambda & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ \vdots & & \ddots & \lambda & 1\\ 0 & \cdots & \cdots & 0 & \lambda\end{pmatrix},\tag{3.71}$$
where λ is any complex number.

Here all the entries on the main diagonal are λ and all the entries just above the main diagonal are 1 (and all other values are 0). Let us use Lemma 3.16 in order to evaluate e^{tA} where A is a Jordan cell. Clearly, we have $A = \lambda \operatorname{id} + N$ where

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}. \tag{3.72}$$

A matrix (3.72) is called a *nilpotent Jordan cell.* Since the matrices $\lambda \operatorname{id}$ and N commute (because id commutes with anything), Lemma 3.16 yields

$$e^{tA} = e^{t\lambda \operatorname{id}} e^{tN} = e^{t\lambda} e^{tN}. \tag{3.73}$$

Hence, we need to evaluate e^{tN}, and for that we first evaluate the powers N^2, N^3, etc. Observe that the components of matrix N are as follows

$$N_{ij} = \begin{cases} 1, & \text{if } j = i + 1 \\ 0, & \text{otherwise} \end{cases},$$

where i is the row index and j is the column index. It follows that

$$\left(N^2\right)_{ij} = \sum_{k=1}^{n} N_{ik} N_{kj} = \begin{cases} 1, & \text{if } j = i + 2 \\ 0, & \text{otherwise} \end{cases}$$

that is,

$$N^2 = \begin{pmatrix} 0 & 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}.$$

Here the entries with value 1 are located on the diagonal that is two positions above the main diagonal. Similarly, we obtain

$$N^k = \begin{pmatrix} 0 & \ddots & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & & \ddots & \ddots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

where the entries with value 1 are located on the diagonal that is k positions above the main diagonal, provided $k < n$, and $N^k = 0$ if $k \geq n$.

Any matrix A with the property that $A^k = 0$ for some natural k is called *nilpotent.* Hence, N is a nilpotent matrix, which explains the term "a nilpotent Jordan cell". It follows that

$$e^{tN} = \mathrm{id} + \frac{t}{1!}N + \frac{t^2}{2!}N^2 + \ldots + \frac{t^{n-1}}{(n-1)!}N^{n-1} = \begin{pmatrix} 1 & \frac{t}{1!} & \frac{t^2}{2!} & \ddots & \frac{t^{n-1}}{(n-1)!} \\ 0 & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!} \\ \vdots & & \ddots & \ddots & \frac{t}{1!} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}. \tag{3.74}$$

Combining with (3.73), we obtain the following statement.

Lemma 3.17 *If A is a Jordan cell* (3.71) *then, for any $t \in \mathbb{R}$,*

$$e^{tA} = \begin{pmatrix} e^{\lambda t} & \frac{t}{1!}e^{t\lambda} & \frac{t^2}{2!}e^{t\lambda} & \ddots & \frac{t^{n-1}}{(n-1)!}e^{t\lambda} \\ 0 & e^{t\lambda} & \frac{t}{1!}e^{t\lambda} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \frac{t^2}{2!}e^{t\lambda} \\ \vdots & & \ddots & \ddots & \frac{t}{1!}e^{t\lambda} \\ 0 & \cdots & \cdots & 0 & e^{t\lambda} \end{pmatrix}. \tag{3.75}$$

By Lemma 3.15, the general solution of the system $x' = Ax$ is $x(t) = e^{tA}v$ where v is an arbitrary vector from $\mathbb{C}^n$. Setting $v = (C_1, ..., C_n)$, we obtain that the general solution is

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

where $x_1, ..., x_n$ are the columns of the matrix e^{tA} (which form a sequence of n linearly independent solutions). Using (3.75), we obtain

$$\begin{aligned} x_1(t) &= e^{\lambda t}(1, 0, ..., 0) \\ x_2(t) &= e^{\lambda t}\left(\frac{t}{1!}, 1, 0, ..., 0\right) \\ x_3(t) &= e^{\lambda t}\left(\frac{t^2}{2!}, \frac{t}{1!}, 1, 0, ..., 0\right) \\ &\cdots \\ x_n(t) &= e^{\lambda t}\left(\frac{t^{n-1}}{(n-1)!}, ..., \frac{t}{1!}, 1\right). \end{aligned}$$

3.8.3 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 3.18 *The following identity is true:*

$$e^{A\otimes B} = e^A \otimes e^B. \tag{3.76}$$

In extended notation, (3.76) 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

$$(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1A_2) \otimes (B_1B_2). \tag{3.77}$$

Indeed, in the extended form this identity means

$$\left(\begin{array}{|c|c|}\hline A_1 & 0 \\ \hline 0 & B_1 \\ \hline\end{array}\right)\left(\begin{array}{|c|c|}\hline A_2 & 0 \\ \hline 0 & B_2 \\ \hline\end{array}\right) = \left(\begin{array}{|c|c|}\hline A_1A_2 & 0 \\ \hline 0 & B_1B_2 \\ \hline\end{array}\right)$$

which follows easily from the rule of multiplication of matrices. Hence, the tensor product commutes with the matrix multiplication. It is also obvious that the tensor product commutes with addition of matrices and taking limits. Therefore, we obtain

$$e^{A\otimes B} = \sum_{k=0}^{\infty} \frac{(A\otimes B)^k}{k!} = \sum_{k=0}^{\infty} \frac{A^k \otimes B^k}{k!} = \left(\sum_{k=0}^{\infty} \frac{A^k}{k!}\right) \otimes \left(\sum_{k=0}^{\infty} \frac{B^k}{k!}\right) = e^A \otimes e^B.$$

■

Definition. A tensor product of a finite number of Jordan cells is called a *Jordan normal form.*

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 3.17 and 3.18 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 3.17, we obtain

$$e^{tJ_1} = \begin{pmatrix} e^t & te^t \\ 0 & e^t \end{pmatrix} \text{ and } e^{tJ_2} = \begin{pmatrix} e^{2t} & te^{2t} \\ 0 & e^{2t} \end{pmatrix}$$

whence by Lemma 3.18,

$$e^{tA} = \begin{pmatrix} e^t & te^t & 0 & 0 \\ 0 & e^t & 0 & 0 \\ 0 & 0 & e^{2t} & te^{2t} \\ 0 & 0 & 0 & e^{2t} \end{pmatrix}.$$

The columns of this matrix form 4 linearly independent solutions

$$\begin{aligned} x_1 &= \left(e^t, 0, 0, 0\right) \\ x_2 &= \left(te^t, e^t, 0, 0\right) \\ x_3 &= \left(0, 0, e^{2t}, 0\right) \\ x_4 &= \left(0, 0, te^{2t}, e^{2t}\right) \end{aligned}$$

and the general solution is

$$\begin{aligned} x(t) &= C_1x_1 + C_2x_2 + C_3x_3 + C_4x_4 \\ &= \left(C_1e^t + C_2te^t, C_2e^t, C_3e^{2t} + C_4te^{2t}, C_4e^{2t}\right). \end{aligned}$$

3.8.4 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_i^b is the i-th component of x^b then

$$x = x_1^b b_1 + x_2^b b_2 + ... + x_n^b b_n = \sum_{i=1}^n x_i^b b_i.$$

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, $(b_i)^b = (0, ...1, ...0)$ where 1 is at position i, which implies that $(Ab_i)^b = A^b (b_i)^b$ is the i-th column of A^b. In other words, we have the identity

$$A^b = \left((Ab_1)^b \mid (Ab_2)^b \mid \cdots \mid (Ab_n)^b \right),$$

that can be stated as the following rule:

the i-th column of A^b is the column vector Ab_i written in the basis $b_1, ..., b_n$.

Example. Consider the operator A in $\mathbb{C}^2$ 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.*

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

$$A^b - \lambda\,\mathrm{id} = \begin{array}{cc} \begin{array}{cccccccc} & & j & \cdots & \cdots & j+p-1 & & \\ & & \downarrow & & & \downarrow & & \end{array} & \\ \left(\begin{array}{cccccccc} \ddots &&&&&&& \\ & \ddots &&&&&& \\ && 0 & 1 & \cdots & 0 && \\ && 0 & \ddots & \ddots & \vdots && \\ && \vdots & \ddots & \ddots & 1 && \\ && 0 & \cdots & 0 & 0 && \\ &&&&&& \ddots & \\ &&&&&&& \ddots \end{array}\right) & \begin{array}{c} \\ \\ \leftarrow j \\ \cdots \\ \cdots \\ \leftarrow j+p-1 \\ \\ \\ \end{array} \end{array}$$

and the k-th column of $A^b - \lambda\,\mathrm{id}$ is the vector $(A - \lambda\,\mathrm{id})\, b_k$ written in the basis b, we conclude that

$$\begin{aligned} (A - \lambda\,\mathrm{id})\, b_j &= 0 \\ (A - \lambda\,\mathrm{id})\, b_{j+1} &= b_j \\ (A - \lambda\,\mathrm{id})\, b_{j+2} &= b_{j+1} \\ &\cdots \\ (A - \lambda\,\mathrm{id})\, b_{j+p-1} &= b_{j+p-2}. \end{aligned}$$

In particular, b_j is an eigenvector of A with the eigenvalue λ. The vectors $b_{j+1}, ..., b_{j+p-1}$ are called the *generalized eigenvectors* of A (more precisely, b_{j+1} is the $1st$ generalized eigenvector, b_{j+2} is the second generalized eigenvector, etc.). Hence, any Jordan chain contains exactly one eigenvector and the rest vectors are the generalized eigenvectors.

Theorem 3.19 *Consider the system $x' = Ax$ with a constant linear operator A and let A^b be the Jordan normal form of A. Then each Jordan cell J of A^b of dimension p with λ on the diagonal gives rise to p linearly independent solutions as follows:*

$$
\begin{aligned}
x_1(t) &= e^{\lambda t} v_1 \\
x_2(t) &= e^{\lambda t}\left(\frac{t}{1!}v_1 + v_2\right) \\
x_3(t) &= e^{\lambda t}\left(\frac{t^2}{2!}v_1 + \frac{t}{1!}v_2 + v_3\right) \\
&\dots \\
x_p(t) &= e^{\lambda t}\left(\frac{t^{p-1}}{(p-1)!}v_1 + ... + \frac{t}{1!}v_{p-1} + v_p\right),
\end{aligned}
$$

where $\{v_1, ..., v_p\}$ is the Jordan chain of J. The set of all n solutions obtained across all Jordan cells is linearly independent.

Proof. In the basis b, we have by Lemmas 3.17 and 3.18

$$
e^{tA^b} = \begin{pmatrix}
\ddots & & & & & & \\
& \ddots & & & & & \\
& & e^{\lambda t} & \frac{t}{1!}e^{t\lambda} & \cdots & \frac{t^{p-1}}{(p-1)!}e^{t\lambda} & & \\
& & 0 & e^{t\lambda} & \ddots & \vdots & & \\
& & \vdots & \ddots & \ddots & \frac{t}{1!}e^{t\lambda} & & \\
& & 0 & \cdots & 0 & e^{t\lambda} & & \\
& & & & & & \ddots & \\
& & & & & & & \ddots
\end{pmatrix},
$$

where the block in the middle is e^{tJ}. By Lemma 3.15, 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,

$$
\begin{aligned}
x_1(t) &= (\dots \underbrace{e^{\lambda t}, 0, \dots, 0}_{p} \dots) \\
x_2(t) &= (\dots \underbrace{\tfrac{t}{1!}e^{\lambda t}, e^{\lambda t}, 0, \dots, 0}_{p} \dots) \\
&\dots \\
x_p(t) &= (\dots \underbrace{\tfrac{t^{p-1}}{(p-1)!}e^{\lambda t}, \dots, \tfrac{t}{1!}e^{\lambda t}, e^{t\lambda}}_{p} \dots),
\end{aligned}
$$

where all the vectors are written in the basis b, the horizontal braces mark the columns of the cell J, and all the terms outside the horizontal braces are zeros. Representing these vectors in the coordinateless form via the Jordan chain $v_1, ..., v_p$, we obtain the solutions as in the statement of Theorem 3.19. ■

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[11] $P(\lambda) = \det(A - \lambda\,\mathrm{id})$. Denote by g the *geometric multiplicity* of λ, that is the dimension of the eigenspace of λ:

$$g = \dim\ker(A - \lambda\,\mathrm{id}).$$

In other words, g is the maximal number of linearly independent eigenvectors of λ. The numbers m and g can be characterized in terms of the Jordan normal form A^b of A as follows: m is the total number of occurrences of λ on the diagonal[12] of A^b, whereas g is equal to the number of the Jordan cells with λ on the diagonal[13]. It follows that $g \leq m$ and the equality occurs if and only if all the Jordan cells with the eigenvalue λ have dimension 1.

Despite this relation to the Jordan normal form, m and g can be determined without a priori finding the Jordan normal form, as it is clear from the definitions of m and g.

Theorem 3.19′ *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*

$$x(t) = e^{\lambda t}\left(u_1 + u_2 t + ... + u_s t^{s-1}\right) \tag{3.78}$$

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 (3.78) into the system $x' = Ax$ considering u_{ij} as unknowns (where u_{ij} is the i-th component of the vector u_j) and solve the resulting linear algebraic system with respect to u_{ij}. The result will contain m arbitrary constants, and the solution in the form (3.78) 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

$$\sum_{j=1}^{g} p_j = m.$$

[11] To compute $P(\lambda)$, one needs to write the operator A in some basis b as a matrix A_b and then evaluate $\det(A_b - \lambda\,\mathrm{id})$. The characteristic polynomial does not depend on the choice of basis b. Indeed, if b' is another basis then the relation between the matrices A_b and $A_{b'}$ is given by $A_b = CA_{b'}C^{-1}$ where C is the matrix of transformation of basis. It follows that $A_b - \lambda\,\mathrm{id} = C(A_{b'} - \lambda\,\mathrm{id})C^{-1}$ whence $\det(A_b - \lambda\,\mathrm{id}) = \det C \det(A_{b'} - \lambda\,\mathrm{id}) \det C^{-1} = \det(A_{b'} - \lambda\,\mathrm{id})$.

[12] 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, $k = m$.

[13] Note that each Jordan cell correponds to exactly one eigenvector.

Hence, the total number of linearly independent solutions that are given by Theorem 3.19 for the eigenvalue λ is equal to m. Let us show that each of the solutions of Theorem 3.19 has the form (3.78). Indeed, each solution of Theorem 3.19 is already in the form

$$e^{\lambda t} \text{ times a polynomial of } t \text{ of degree } \leq p_j - 1.$$

To ensure that these solutions can be represented in the form (3.78), we only need to verify that $p_j - 1 \leq s - 1$. Indeed, we have

$$\sum_{j=1}^{g} (p_j - 1) = \left(\sum_{j=1}^{g} p_j \right) - g = m - g = s - 1,$$

whence the inequality $p_j - 1 \leq s - 1$ follows. ∎

In particular, if $m = g$, that is, $s = 1$, then m independent solutions can be found in the form $x(t) = e^{\lambda t} v$, where v is one of m independent eigenvectors of λ. This case has been already discussed above. Consider some examples, where $g < m$.

Example. Solve the system

$$x' = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix} x.$$

The characteristic polynomial is

$$P(\lambda) = \det(A - \lambda \operatorname{id}) = \det \begin{pmatrix} 2-\lambda & 1 \\ -1 & 4-\lambda \end{pmatrix} = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2,$$

and the only eigenvalue is $\lambda_1 = 3$ with the algebraic multiplicity $m_1 = 2$. The equation for an eigenvector v is

$$(A - \lambda \operatorname{id}) v = 0$$

that is, for $v = (a, b)$,

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0,$$

which is equivalent to $-a + b = 0$. Setting $a = 1$ and $b = 1$, we obtain the unique (up to a constant multiple) eigenvector

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Hence, the geometric multiplicity is $g_1 = 1$. Hence, there is only one Jordan cell with the eigenvalue λ_1, which allows to immediately determine the Jordan normal form of the given matrix:

$$\begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}.$$

By Theorem 3.19, 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

$$\begin{aligned} P(\lambda) &= \det(A - \lambda \operatorname{id}) = \det \begin{pmatrix} 2-\lambda & 1 & 1 \\ -2 & -\lambda & -1 \\ 2 & 1 & 2-\lambda \end{pmatrix} \\ &= -\lambda^3 + 4\lambda^2 - 5\lambda + 2 = (2-\lambda)(\lambda-1)^2. \end{aligned}$$

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

$$(A - \lambda_1 \operatorname{id}) v = 0,$$

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 3.19, 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

$$x_3(t) = e^t (t v_1 + v_2),$$

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

$$x(t) = C_1x_1 + C_2x_2 + C_3x_3 = \begin{pmatrix} C_1e^{2t} - C_3e^t \\ -2C_1e^{2t} + (C_2+C_3t)\,e^t \\ 2C_1e^{2t} + (C_3-C_2-C_3t)\,e^t \end{pmatrix}.$$

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 and 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 phase trajectories is called a *phase diagram* or a *phase portrait.*

Recall that the graph of a solution (or the integral curve) is the set of points $(t, x(t))$ in $\mathbb{R}\times\Omega$. Hence, a phase trajectory can be regarded as the projection of an integral curve onto Ω.

Assume in the sequel that f is continuously differentiable in Ω. For any $y \in \Omega$, denote by $x(t,y)$ the maximal solution to the IVP

$$\begin{cases} x' = f(x) \\ x(0) = y. \end{cases}$$

Recall that, by Theorem 2.14, the domain of function $x(t,y)$ is an open subset of $\mathbb{R}^{n+1}$ and $x(t,y)$ is continuously differentiable in this domain.

The fact that f does not depend on t, implies 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 $x(t-t_0, y)$ solves the following IVP

$$\begin{cases} x' = f(x) \\ x(t_0) = y. \end{cases}$$

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*[14] of the ODE $x' = f(x)$.

It follows from the above observation that if x_0 is a stationary point if and only if $x(t, x_0) \equiv x_0$.

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 $y \in \Omega$ such that $\|y - x_0\| < \delta$, the solution $x(t, y)$ is defined for all $t > 0$ and

$$\sup_{t \in (0,+\infty)} \|x(t, y) - x_0\| < \varepsilon. \tag{4.2}$$

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 (4.2) the interval $(0, +\infty)$ by any bounded interval $[a, b]$ containing 0 then by the continuity of $x(t, y)$,

$$\sup_{t \in [a,b]} \|x(t, y) - x_0\| = \sup_{t \in [a,b]} \|x(t, y) - x(t, x_0)\| \to 0 \text{ as } y \to x_0.$$

Hence, the main issue for the stability is the behavior of solutions as $t \to +\infty$.

Definition. A stationary point x_0 is called *asymptotically stable* for the system $x' = f(x)$ (or the system is called asymptotically stable at x_0), if it is Lyapunov stable and, in addition,

$$\|x(t, y) - x_0\| \to 0 \text{ as } t \to +\infty,$$

provided $\|y - x_0\|$ 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.

[14] In the literature one can find the following synonyms for the term "stationary point": rest point, singular point, equilibrium point, fixed point.

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, $\operatorname{Re}\lambda > 0$ then 0 is unstable.*

Proof. By Theorem 3.19′, the general complex solution of $x' = Ax$ has the form

$$x(t) = \sum_{k=1}^{n} C_k e^{\lambda_k t} P_k(t), \tag{4.3}$$

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 $P_k(t) = u_1 + u_2 t + ... + u_s t^{s-1}$ for some $s \in \mathbb{N}$ and for some vectors $u_1, ..., u_s$. Note that this solution is obtained by taking a linear combination of n independent solutions $e^{\lambda_k t} P_k(t)$. Since

$$x(0) = \sum_{k=1}^{n} C_k P_k(0),$$

we see that the coefficients C_k are the components of $x(0)$ in the basis $\{P_k(0)\}_{k=1}^{n}$.

It follows from (4.3) that

$$\begin{aligned} \|x(t)\| &\leq \sum_{k=1}^{n} \left|C_k e^{\lambda_k t}\right| \|P_k(t)\| \\ &\leq \max_k |C_k| \, e^{(\operatorname{Re}\lambda_k)t} \sum_{k=1}^{n} \|P_k(t)\|. \end{aligned}$$

Set

$$\alpha = \max_k \operatorname{Re}\lambda_k < 0.$$

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. Hence, it follows that

$$\|x(t)\| \leq C e^{\alpha t}\left(1 + t^N\right) \|x(0)\|_\infty \tag{4.4}$$

Clearly, by adjusting the constant C, we can replace $\|x(0)\|_\infty$ by $\|x(0)\|$.

Since the function $\left(1 + t^N\right) e^{\alpha t}$ is bounded on $(0, +\infty)$, we obtain that there is a constant K such that, for all $t > 0$,

$$\|x(t)\| \leq K \|x(0)\|,$$

whence it follows that the stationary point 0 is Lyapunov stable. Moreover, since

$$\left(1 + t^N\right) e^{\alpha t} \to 0 \quad \text{as} \quad t \to +\infty,$$

we conclude from (4.4) that $\|x(t)\| \to 0$ as $t \to \infty$, that is, the stationary point 0 is asymptotically stable.

Let now $\operatorname{Re}\lambda > 0$ for some eigenvalue λ. To prove that 0 is unstable is suffices to show that there exists an unbounded real solution $x(t)$, that is, a solution for which $\|x(t)\|$ is not bounded on $(0, +\infty)$ as a function of t. Indeed, if such a solution exists then the function $\varepsilon x(t)$ is also an unbounded solution for any $\varepsilon > 0$, while its initial value $\varepsilon x(0)$ can be made arbitrarily small by choosing ε appropriately.

To construct an unbounded solution, consider an eigenvector v of the eigenvalue λ. It gives rise to the solution

$$x(t) = e^{\lambda t} v$$

for which

$$\|x(t)\| = \left|e^{\lambda t}\right| \|v\| = e^{t \operatorname{Re}\lambda} \|v\|.$$

Hence, $\|x(t)\|$ is unbounded. If $x(t)$ is a real solution then this finishes the proof. In general, if $x(t)$ is a complex solution then then either $\operatorname{Re} x(t)$ or $\operatorname{Im} x(t)$ is unbounded (in fact, both are), whence the instability of 0 follows. ■

This theorem does not answer the question what happens when $\operatorname{Re}\lambda = 0$. We will investigate this for the case $n = 2$ where we also give a more detailed description of the phase diagrams.

Consider now a linear system $x' = Ax$ in $\mathbb{R}^2$ where A is a constant operator in $\mathbb{R}^2$. Let $b = \{b_1, b_2\}$ be the Jordan basis of A so that A^b has the Jordan normal form. Consider first the case when the Jordan normal form of A has two Jordan cells, that is,

$$A^b = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Then b_1 and b_2 are the eigenvectors of the eigenvalues λ_1 and λ_2, respectively, and the general solution is

$$x(t) = C_1 e^{\lambda_1 t} b_1 + C_2 e^{\lambda_2 t} b_2.$$

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

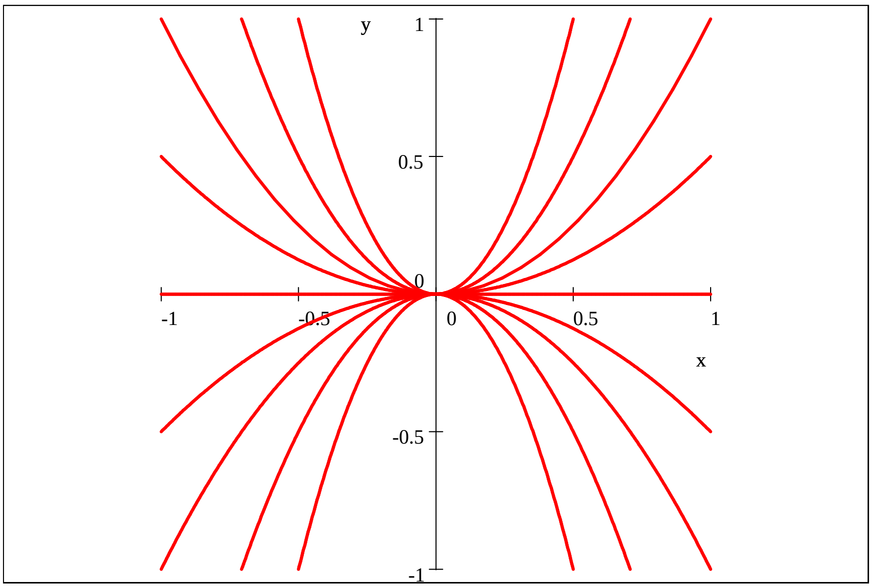
$$x_1 = C_1 e^{\lambda_1 t} \quad \text{and} \quad x_2 = C_2 e^{\lambda_2 t}.$$

Assuming that $\lambda_1, \lambda_2 \neq 0$, we obtain the relation between x_1 and x_2 as follows:

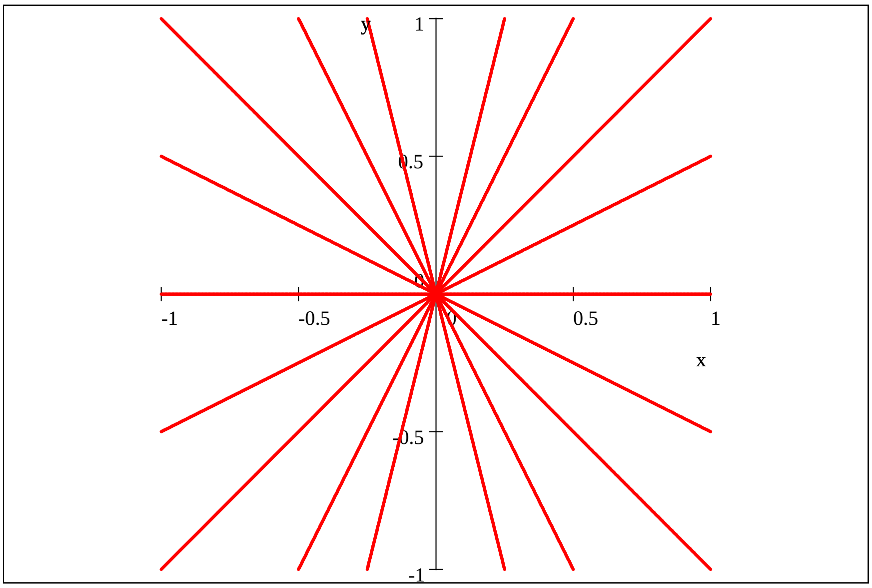
$$x_2 = C |x_1|^{\gamma},$$

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 $\gamma > 1$:

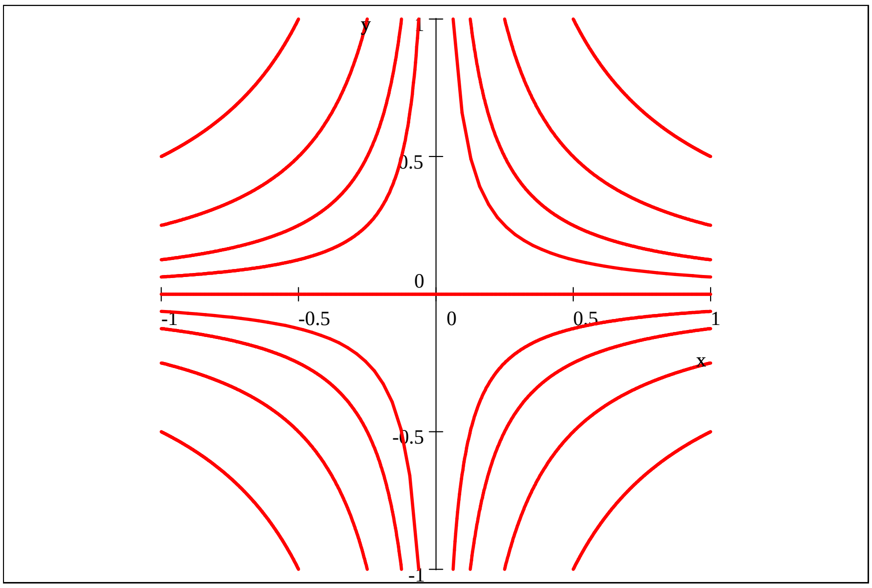


and here is a node with $\gamma = 1$:



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 $\beta \neq 0$.

Then we rewrite the general solution in the real form

$$x(t) = C_1 \operatorname{Re} e^{(\alpha - i\beta)t} b_1 + C_2 \operatorname{Im} e^{(\alpha - i\beta)t} b_1.$$

Note that b_1 is an eigenvector of λ_1 and, hence, must have a non-trivial imaginary part in any real basis. We claim that in some real basis b_1 has the form $(1, i)$. Indeed, if $b_1 = (p, q)$ in the canonical basis e_1, e_2 then by rotating the basis we can assume $p, q \neq 0$. Since b_1 is an eigenvector, it is defined up to a constant multiple, so that we can take $p = 1$. Then, setting $q = q_1 + iq_2$ we obtain

$$b_1 = e_1 + (q_1 + iq_2) e_2 = (e_1 + q_1 e_2) + iq_2 e_2 = e_1' + ie_2'$$

where $e_1' = e_1 + q_1 e_2$ and $e_2' = q_2 e_2$ is a new basis (the latter follows from the fact that q is imaginary and, hence, $q_2 \neq 0$). Hence, in the basis $e' = \{e_1', e_2'\}$ we have $b_1 = (1, i)$.

It follows that in the basis e'

$$e^{(\alpha + \beta i)t} b_1 = e^{\alpha t} (\cos \beta t + i \sin \beta t) \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} e^{\alpha t} \cos \beta t - ie^{\alpha t} \sin \beta t \\ e^{\alpha t} \sin \beta t + ie^{\alpha t} \cos \beta t \end{pmatrix}$$

and

$$x(t) = C_1 \begin{pmatrix} e^{\alpha t} \cos \beta t \\ e^{\alpha t} \sin \beta t \end{pmatrix} + C_2 \begin{pmatrix} -e^{\alpha t} \sin \beta t \\ e^{\alpha t} \cos \beta t \end{pmatrix} = C \begin{pmatrix} e^{\alpha t} \cos (\beta t + \psi) \\ e^{\alpha t} \sin (\beta t + \psi) \end{pmatrix},$$

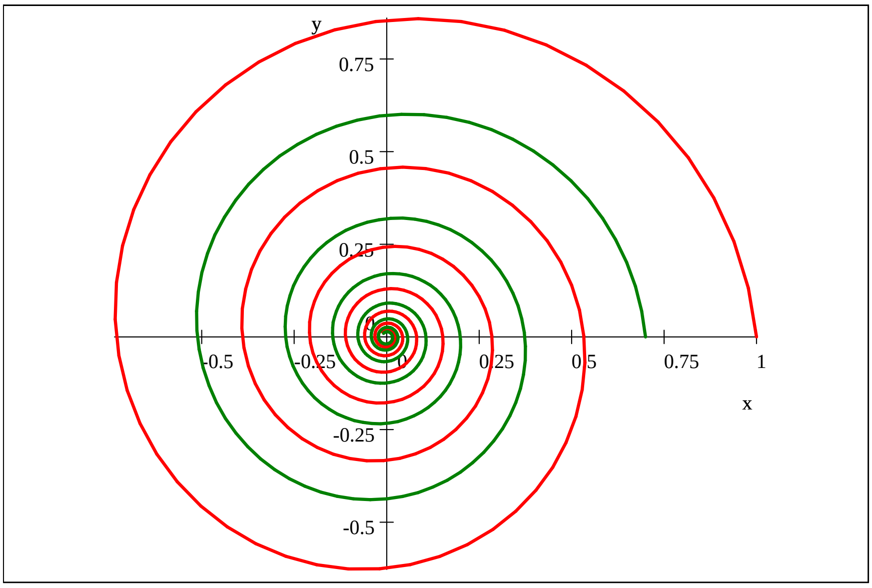
where $C = \sqrt{C_1^2 + C_2^2}$ and

$$\cos \psi = \frac{C_1}{C}, \ \sin \psi = \frac{C_2}{C}.$$

If (r, θ) are the polar coordinates on the plane in the basis e', then the polar coordinates for the solution $x(t)$ are

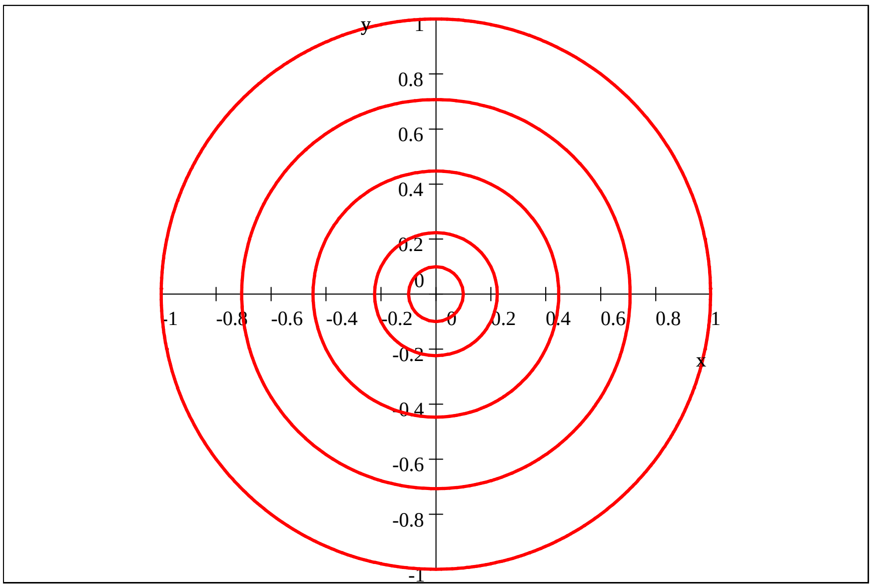
$$r(t) = Ce^{\alpha t} \text{ and } \theta(t) = \beta t + \psi.$$

If $\alpha \neq 0$ then these equations define a *logarithmic spiral*, and the phase diagram is called a *focus* or a *spiral*:



The focus is stable is $\alpha < 0$ and unstable if $\alpha > 0$.

If $\alpha = 0$ (that is, the both eigenvalues λ_1 and λ_2 are purely imaginary), then $r(t) = C$, that is, we get a family of concentric circles around 0, and this phase diagram is called a *center:*



In this case, the stationary point is stable but not asymptotically stable.

Consider now the case when the Jordan normal form of A has only one Jordan cell, that is,

$$A^b = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

In this case, λ must be real because if λ is an imaginary root of a characteristic polynomial then $\overline{\lambda}$ must also be a root, which is not possible since $\overline{\lambda}$ does not occur on the diagonal of A^b. Then the general solution is

$$x(t) = C_1 e^{\lambda t} b_1 + C_2 e^{\lambda t} (b_1 t + b_2) = (C_1 + C_2 t) e^{\lambda t} b_1 + C_2 e^{\lambda t} b_2$$

whence $x(0) = C_1 b_1 + C_2 b_2$. That is, in the basis b, we can write $x(0) = (C_1, C_2)$ and

$$x(t) = \left(e^{\lambda t}(C_1 + C_2 t), e^{\lambda t} C_2\right) \tag{4.5}$$

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), while 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

$$\|x(t)\|_1 = e^{\lambda t}(t+1),$$

which is unbounded.

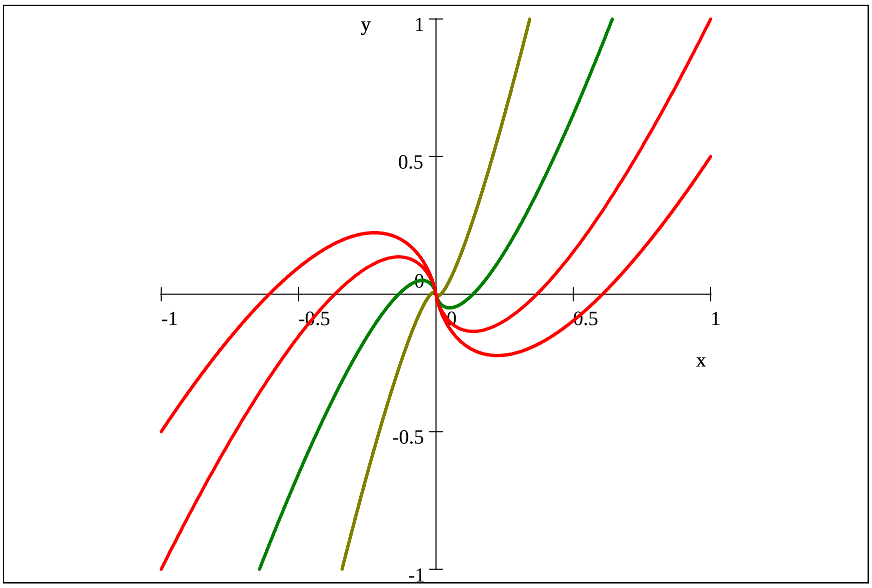
If $\lambda \neq 0$ then it follows from (4.5) that the components x_1, x_2 of x are related as follows:

$$\frac{x_1}{x_2} = \frac{C_1}{C_2} + t \quad \text{and} \quad t = \frac{1}{\lambda} \ln \frac{x_2}{C_2}$$

whence

$$x_1 = C x_2 + \frac{x_2 \ln |x_2|}{\lambda}$$

for some constant C. Here is the phase diagram in this case:



This phase diagram is also called a node. It is stable if $\lambda < 0$ and unstable if $\lambda > 0$. If $\lambda = 0$ then we obtain a degenerate phase diagram - parallel straight lines.

Hence, the main types of the phases diagrams are the *node* (λ_1, λ_2 are real, non-zero and of the same sign), the *saddle* (λ_1, λ_2 are real, non-zero and of opposite signs), *focus/spiral* (λ_1, λ_2 are imaginary and $\operatorname{Re} \lambda \neq 0$) and *center* ($\lambda_1, \lambda_2$ are purely imaginary). Otherwise, the phase diagram consists of parallel straight lines or just dots, and is referred to as degenerate.

To summarize the stability investigation, let us emphasize that in the case $\operatorname{Re} \lambda = 0$ both stability and instability can happen, depending on the structure of the Jordan normal form.

4.3 Lyapunov's theorem

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 $A = f'(x_0)$ (that is, A is the Jacobian matrix of f at x_0). If $\operatorname{Re}\lambda < 0$ for all eigenvalues λ of A then the stationary point x_0 is asymptotically stable for the system $x' = f(x)$.*

Remark. This theorem has the second part that says the following: if $\operatorname{Re}\lambda > 0$ for some eigenvalue λ of A then x_0 is unstable for $x' = f(x)$. However, the proof of that 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

$$A = f'(0,0) = \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ \partial_x f_2 & \partial_y f_2 \end{pmatrix} = \begin{pmatrix} -2 & -1 \\ 3 & -4 \end{pmatrix}.$$

Another way to obtain this matrix is to expand each component of $f(x,y)$ by the Taylor formula:

$$\begin{aligned} f_1(x,y) &= 2\sqrt{1+y} - 2e^{x+y} = 2\left(1 + \frac{y}{2} + o(x)\right) - 2\left(1 + (x+y) + o(|x|+|y|)\right) \\ &= -2x - y + o(|x|+|y|) \end{aligned}$$

and

$$\begin{aligned} f_2(x,y) &= \sin 3x + \ln(1-4y) = 3x + o(x) - 4y + o(y) \\ &= 3x - 4y + o(|x|+|y|). \end{aligned}$$

Hence,

$$f(x,y) = \begin{pmatrix} -2 & -1 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + o(|x|+|y|),$$

whence we obtain the same matrix A.

The characteristic polynomial of A is

$$\det \begin{pmatrix} -2-\lambda & -1 \\ 3 & -4-\lambda \end{pmatrix} = \lambda^2 + 6\lambda + 11,$$

and the eigenvalues are

$$\lambda_{1,2} = -3 \pm i\sqrt{2}.$$

Hence, $\operatorname{Re}\lambda < 0$ for all λ, whence we conclude that 0 is asymptotically stable.

The main tool for the proof of theorem 4.2 is the following lemma, that is of its own interest. Recall that for any vector $v \in \mathbb{R}^n$ and a differentiable function F in a domain in $\mathbb{R}^n$, the directional derivative $\partial_v F$ can be determined by

$$\partial_v F(x) = F'(x)\, v = \sum_{k=1}^{n} \frac{\partial F}{\partial x_k}(x)\, v_k.$$

Lemma 4.3 (Lyapunov's lemma) *Consider the system $x' = f(x)$ where $f \in C^1(\Omega)$ and let x_0 be a stationary point of it. Let $V(x)$ be a C^1 scalar function in an open set U such that $x_0 \in U \subset \Omega$ and the following conditions hold:*

1. *$V(x) > 0$ for any $x \in U \setminus \{x_0\}$ and $V(x_0) = 0$.*
2. *For all $x \in U$,*
$$\partial_{f(x)} V(x) \leq 0. \tag{4.6}$$

Then the stationary point x_0 is stable.
Furthermore, if all $x \in U$
$$\partial_{f(x)} V(x) \leq -W(x), \tag{4.7}$$
where $W(x)$ is a continuous function on U such that $W(x) > 0$ for $x \in U \setminus \{x_0\}$, then the stationary point x_0 is asymptotically stable.

Function V with the properties 1-2 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 $x' = f(x)$.

Before the proof, let us show examples of the Lyapunov functions.

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 $f(x) = Ax$, 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

$$\partial_f V = \sum_{k=1}^{n} \frac{\partial V}{\partial x_k} f_k = 2 \sum_{j,k=1}^{n} A_{kj} x_j x_k.$$

The matrix (A_{kj}) is called a *non-positive definite* if

$$\sum_{j,k=1}^{n} A_{kj}x_jx_k \leq 0 \text{ for all } x \in \mathbb{R}^n.$$

Hence, in the case when A is non-positive definite, we have $\partial_f V \leq 0$ so that V is a Lyapunov function. It follows that in this case 0 is Lyapunov stable. Matrix A is called *negative definite* if

$$\sum_{j,k=1}^{n} A_{kj}x_jx_k < 0 \text{ for all } x \in \mathbb{R}^n \setminus \{0\}.$$

Then setting $W(x) = -\sum_{j,k=1}^{n} A_{kj}x_jx_k$, we obtain $\partial_f V = -W$ so that by the second part of Lemma 4.3, 0 is asymptotically stable.

For example, if $A = \operatorname{diag}(\lambda_1, ..., \lambda_n)$ then A is negative definite if all $\lambda_k < 0$, and A is non-positive definite if all $\lambda_k \leq 0$.

Example. Consider the second order scalar ODE $x'' + kx' = F(x)$ which describes the movement of a body under the external potential force $F(x)$ and friction with the coefficient k. This can be written as a system

$$\begin{cases} x' = y \\ y' = -ky + F(x). \end{cases}$$

Note that the phase space is $\mathbb{R}^2$ (assuming that F is defined on $\mathbb{R}$) and a point (x, y) in the phase space is a couple position-velocity.

Assume $F(0) = 0$ so that $(0, 0)$ is a stationary point. We would like to answer the question if $(0, 0)$ is stable or not. The Lyapunov function can be constructed in this case as the full energy

$$V(x, y) = \frac{y^2}{2} + U(x),$$

where

$$U(x) = -\int F(x)\,dx$$

is the potential energy and $\frac{y^2}{2}$ is the kinetic energy. More precisely, assume that $k \geq 0$ and

$$F(x) < 0 \;\text{ for }\; x > 0, \qquad F(x) > 0 \text{ for } x < 0,$$

and set

$$U(x) = -\int_0^x F(s)\,ds,$$

so that $U(0) = 0$ and $U(x) > 0$ for $x \neq 0$. Then the function $V(x, y)$ is positive away from $(0, 0)$ and vanishes at $(0, 0)$.

Setting

$$f(x, y) = (y, -ky + F(x)),$$

let us compute the orbital derivative $\partial_f V$:

$$\begin{aligned} \partial_f V &= y\frac{\partial V}{\partial x} + (-ky + F(x))\frac{\partial V}{\partial y} \\ &= yU'(x) + (-ky + F(x))\,y \\ &= -yF(x) - ky^2 + F(x)\,y = -ky^2 \leq 0. \end{aligned}$$

Hence, V is indeed the Lyapunov function, and by Lemma 4.3 the stationary point $(0,0)$ is Lyapunov stable.

Physically this has a simple meaning. The fact that $F(x) < 0$ for $x > 0$ and $F(x) > 0$ for $x < 0$ means that the force always acts in the direction of the origin thus trying to return the displaced body to the stationary point, which causes the stability.

Proof of Lemma 4.3. By shrinking U, we can assume that U is bounded and that V is defined on $\overline{U}$. Set

$$B_r = B(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| < r\}$$

and observe that, by the openness of U, $B_\varepsilon \subset U$ provided $\varepsilon > 0$ is small enough. 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

$$V(x) \geq m(\varepsilon) \quad \text{for all} \quad x \in \overline{U} \setminus B_\varepsilon. \tag{4.8}$$

Since $V(x_0) = 0$, for any given $\varepsilon > 0$ there is $\delta > 0$ so small that

$$V(x) < m(\varepsilon) \quad \text{for all} \quad x \in B_\delta.$$

Fix $y \in B_\delta$ and let $x(t)$ be the maximal solution in $\mathbb{R} \times U$ of the IVP

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

We will show that $x(t) \in B_\varepsilon$ for all $t > 0$, which means that the system is Lyapunov stable at x_0.

For any solution $x(t)$ in U, we have by the chain rule

$$\frac{d}{dt} V(x(t)) = V'(x)\, x'(t) = V'(x)\, f(x) = \partial_{f(x)} V(x) \leq 0. \tag{4.9}$$

Therefore, the function V is decreasing along any solution $x(t)$ as long as $x(t)$ remains inside U.

If the initial point y is in B_δ then $V(y) < m(\varepsilon)$ and, hence, $V(x(t)) < m(\varepsilon)$ for $t > 0$ as long as $x(t)$ is defined in U. It follows from (4.8) that $x(t) \in B_\varepsilon$. We are left to verify that $x(t)$ is defined[15] for all $t > 0$. Indeed, assume that $x(t)$ is defined only for $t < T$ where T is finite. By Theorem 2.8, if $t \to T-$, 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 set $[0, T] \times \overline{B_\varepsilon}$. This contradiction shows that $T = +\infty$, which finishes the proof of the first part.

For the second part, we obtain by (4.7) and (4.9)

$$\frac{d}{dt} V(x(t)) \leq -W(x(t)).$$

[15] Since $x(t)$ has been defined as the maximal solution in the domain $\mathbb{R} \times U$, the solution $x(t)$ is always contained in U as long as it is defined.

It suffices to show that

$$V(x(t)) \to 0 \text{ as } t \to \infty$$

since this will imply that $x(t) \to 0$ (recall that 0 is the only point where V vanishes). Since $V(x(t))$ is decreasing in t, the limit

$$L = \lim_{t \to +\infty} V(x(t))$$

exists. Assume from the contrary that $L > 0$. Then, for all $t > 0$, $V(x(t)) \geq L$. By the continuity of V, there is $r > 0$ such that

$$V(y) < L \text{ for all } y \in B_r.$$

Hence, $x(t) \notin B_r$ for all $t > 0$. Set

$$m = \inf_{y \in \overline{U} \setminus B_r} W(y) > 0.$$

It follows that $W(x(t)) \geq m$ for all $t > 0$ whence

$$\frac{d}{dt} V(x(t)) \leq -W(x(t)) \leq -m$$

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. ■

Proof of Theorem 4.2. Without loss of generality, set $x_0 = 0$. Using that $f \in C^2$, we obtain by the Taylor formula, for any component f_k of f,

$$f_k(x) = f_k(0) + \sum_{i=1}^{n} \partial_i f_k(0)\, x_i + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} f_k(0)\, x_i x_j + o\left(\|x\|^2\right) \text{ as } x \to 0.$$

Noticing that $\partial_i f_k(0) = A_{ki}$ write

$$f(x) = Ax + h(x)$$

where $h(x)$ is defined by

$$h_k(x) = \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} f_k(0)\, x_i x_j + o\left(\|x\|^2\right).$$

Setting $B = \max_{i,j,k} |\partial_{ij} f_k(0)|$, we obtain

$$\|h(x)\|_\infty = \max_{1 \leq k \leq n} |h_k(x)| \leq B \sum_{i,j=1}^{n} |x_i x_j| + o\left(\|x\|^2\right) = B \|x\|_1^2 + o\left(\|x\|^2\right).$$

Hence, for any choice of the norms, there is a constant C such that

$$\|h(x)\| \leq C \|x\|^2 \tag{4.10}$$

provided $\|x\|$ is small enough.

Assuming that $\operatorname{Re}\lambda < 0$ for all eigenvalues of A, consider the following function

$$V(x) = \int_0^\infty \left\|e^{sA}x\right\|_2^2 ds \tag{4.11}$$

and prove that $V(x)$ is the Lyapunov function.

Let us first verify that $V(x)$ is finite, that is, the integral in (4.11) converges. Indeed, in the proof of Theorem 4.1 we have established the inequality

$$\left\|e^{tA}x\right\| \leq Ce^{\alpha t}\left(t^N + 1\right)\|x\|, \tag{4.12}$$

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 $\alpha < 0$, (4.12) implies that $\left\|e^{sA}x\right\|$ decays exponentially as $s \to +\infty$, whence the convergence of the integral in (4.11) follows.

Next, let us show that $V(x)$ is of the class C^1 (in fact, C^∞). For that, represent x in the canonical basis $v_1, ..., v_n$ as $x = \sum x_i v_i$ and notice that

$$\|x\|_2^2 = \sum_{i=1}^n |x_i|^2 = x \cdot x.$$

Therefore,

$$\begin{aligned} \left\|e^{sA}x\right\|_2^2 &= e^{sA}x \cdot e^{sA}x = \left(\sum_i x_i\left(e^{sA}v_i\right)\right) \cdot \left(\sum_j x_j\left(e^{sA}v_j\right)\right) \\ &= \sum_{i,j} x_i x_j \left(e^{sA}v_i \cdot e^{sA}v_j\right). \end{aligned}$$

Integrating in s, we obtain

$$V(x) = \sum_{i,j} b_{ij} x_i x_j$$

where $b_{ij} = \int_0^\infty \left(e^{sA}v_i \cdot e^{sA}v_j\right) ds$ are constants, which clearly implies that $V(x)$ is infinitely many times differentiable in x.

Remark. Usually we work with any norm in $\mathbb{R}^n$. In the definition (4.11) of $V(x)$, we have specifically chosen the 2-norm to ensure the smoothness of $V(x)$.

Function $V(x)$ is obviously non-negative and $V(x) = 0$ if and only if $x = 0$. In order to complete the proof of the fact that $V(x)$ is the Lyapunov function, we need to estimate $\partial_{f(x)}V(x)$. Let us first evaluate $\partial_{Ax}V(x)$ for any $x \in U$. Since the function $y(t) = e^{tA}x$ solves the ODE $y' = Ay$, we have by (4.9)

$$\partial_{Ay(t)}V(y(t)) = \frac{d}{dt}V(y(t)).$$

Setting $t = 0$ and noticing that $y(0) = x$, we obtain

$$\partial_{Ax}V(x) = \left.\frac{d}{dt}V\left(e^{tA}x\right)\right|_{t=0}. \tag{4.13}$$

On the other hand,

$$V\left(e^{tA}x\right) = \int_0^\infty \left\| e^{sA}\left(e^{tA}x\right)\right\|_2^2 ds = \int_0^\infty \left\| e^{(s+t)A}x\right\|_2^2 ds = \int_t^\infty \left\| e^{\tau A}x\right\|_2^2 d\tau$$

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 (4.13), we obtain

$$\partial_{Ax}V(x) = \left.\frac{d}{dt}V\left(e^{tA}x\right)\right|_{t=0} = -\|x\|_2^2.$$

Now we can estimate $\partial_{f(x)}V(x)$ as follows:

$$\begin{aligned} \partial_{f(x)}V(x) &= \partial_{Ax}V(x) + \partial_{h(x)}V(x) \\ &= -\|x\|_2^2 + V'(x)\cdot h(x) \\ &\leq -\|x\|_2^2 + \|V'(x)\|_2\,\|h(x)\|_2, \end{aligned}$$

where we have used the Cauchy-Schwarz inequality $u\cdot v \leq \|u\|_2\,\|v\|_2$ for all $u, v \in \mathbb{R}^n$. Next, let us use the estimate (4.10) in the form

$$\|h(x)\|_2 \leq C\,\|x\|_2^2,$$

which is true provided $\|x\|_2$ is small enough. Observe also that the function $V(x)$ has minimum at 0, which implies that $V'(0) = 0$. Hence, if $\|x\|_2$ is small enough then

$$\|V'(x)\|_2 \leq \frac{1}{2}C^{-1}.$$

Combining together the above three 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 Lemma 4.3, that the ODE $x' = f(x)$ is asymptotically stable at 0. ■

Now consider some examples of investigation of stationary points of an autonomous system $x' = f(x)$.

The first step is to find the stationary points, that is, to solve the equation $f(x) = 0$. In general, it may have many roots. Then each root requires a separate investigation.

Let x_0 denote as before one of the stationary points of the system. The second step is to compute the matrix $A = f'(x_0)$. Of course, the matrix A can be found as the Jacobian matrix componentwise by $A_{kj} = \partial_{x_j}f_k(x_0)$. However, in practice is it frequently more convenient to do as follows. Setting $X = x - x_0$, we obtain that the system $x' = f(x)$ transforms to

$$X' = f(x) = f(x_0 + X) = f(x_0) + f'(x_0)X + o(\|X\|)$$

as $X \to 0$, that is, to

$$X' = AX + o(\|X\|).$$

Hence, the linear term AX appears in the right hand side if we throw away the terms of the order $o(\|X\|)$. The equation $X' = AX$ is called the *linearized system* for $x' = f(x)$ at x_0.

The third step is the investigation of the stability of the linearized system, which amounts to evaluating the eigenvalues of A and, possibly, the Jordan normal form.

The fours step is the conclusion of the stability of the non-linear system $x' = f(x)$ using Lyapunov's theorem or Lyapunov lemma. If $\operatorname{Re}\lambda < 0$ for all eigenvalues λ of A then both linearized and non-linear system are asymptotically stable at x_0, and if $\operatorname{Re}\lambda > 0$ for some eigenvalue λ then both are unstable. The other cases require additional investigation.

Example. Consider the system

$$\begin{cases} x' = y + xy, \\ y' = -x - xy. \end{cases} \tag{4.14}$$

For the stationary points we have the equation

$$\begin{cases} y + xy = 0 \\ x + xy = 0 \end{cases}$$

whence we obtain two roots: $(x, y) = (0, 0)$ and $(x, y) = (-1, -1)$.

Consider first the stationary point $(-1, -1)$. Setting $X = x + 1$ and $Y = y + 1$, we obtain the system

$$\begin{cases} X' = (Y - 1)X = -X + XY = -X + o(\|(X, Y)\|) \\ Y' = -(X - 1)Y = Y - XY = Y + o(\|(X, Y)\|) \end{cases} \tag{4.15}$$

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. The linearized and non-linear system are unstable at $(-1, -1)$ because one of the eigenvalues is positive.

Consider now the stationary point $(0, 0)$. Near this point, the system 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 $(0, 0)$ is a center. Hence, the linearized system is stable at $(0, 0)$ but not asymptotically stable.

For the non-linear system (4.14), no conclusion can be drawn just from the eigenvalues. In this case, one can use the following Lyapunov function:

$$V(x, y) = x - \ln(x+1) + y - \ln(y+1),$$

which is defined for $x > -1$ and $y > -1$. Indeed, the function $x - \ln(x+1)$ take the minimum 0 at $x = 0$ and is positive for $x \neq 0$. It follows that $V(x, y)$ takes the minimal value 0 at $(0, 0)$ and is positive away from $(0, 0)$. The orbital derivative of V is

$$\begin{aligned}
\partial_f V &= (y + xy)\,\partial_x V - (x + xy)\,\partial_y V \\
&= (y + xy)\left(1 - \frac{1}{x+1}\right) - (x + xy)\left(1 - \frac{1}{y+1}\right) \\
&= xy - xy = 0.
\end{aligned}$$

Hence, V is the Lyapunov function, which implies that $(0, 0)$ is stable for the non-linear system.

Since $\partial_f V = 0$, it follows from (4.9) that V remains constant along the trajectories of the system. Using that one can easily show that $(0, 0)$ is not asymptotically stable and the type of the stationary point $(0, 0)$ for the non-linear system is also a center. The phase trajectories of this system around $(0, 0)$ are shown on the diagram.