GMIT

DIFFERENTIAL EQUATIONS AND APPLICATIONS

Fernando Pestana da Costa
(Univ. Aberta & CAMGSD, IST; Lisbon, Portugal)

September 2024

fcosta@uab.pt

GMIT

2/9/2024 – 27/9/2024

Differential Equations and Aplications

Topics to be studied:

1. ODEs: basic theory; concepts and results of qualitative theory. Examples from applications: population balances, pendulum equations, problems from Biology

2. PDEs: diffusion equations: deduction, solution via separation of variables and Fourier series. A detour to Analysis: basic results from Fourier series (definitions, convergence). Wave equation: deduction, solutions via separation of variables and Fourier series; wave equation in unbounded regions (d'Alembert solution, propagation of waves). Nonlinear waves: Burgers' equation: shocks, notion of weak solution.

References

Churchill & Brown: Fourier Series and Boundary Value Problems, 3rd Ed McGraw-Hill Kogakusha, Tokyo, 1978

Costa: Equações Diferenciais Ordinárias, IST Press, Lisboa, 2001 [in Portuguese]

Figueiredo: Análise de Fourier e Equações Diferenciais Parciais, IMPA, Rio de Janeiro, 1987 [in Portuguese]

Holmes: Introduction to the Foundations of Applied Mathematics, Springer, New York, 2009

Salsa, Vegni, Zaretti & Zunino: A Primer on PDEs: Models, Methods, Simulations, Springer, Milan, 2013

Schroers, Ordinary Differential Equations, Cambridge UP, 2011

Plan of the sessions

SEMINAR ON APPLICATIONS - population models - radioactive decay - pendulum equations 2/9	(ODEs) Revisions Linear equations Separable equations Change of variables Integrating factors 3/9	(ODEs) Existence and uniqueness of solutions (Picard-Lindelöf) Continuous dependence Maximal intervals 4/9	(ODEs) Linear systems: e^{At} and its computations 5/9
(ODEs) Linear systems: eigenvalues and eigenvectors; phase portraits; notions of qualitative theory 9/9	(ODEs) Conservative systems (examples in $\mathbb{R}^2$). Linearization about equilibria 10/9	(ODEs) Lyapunov functions and stability 11/9	SEMINAR ON APPLICATIONS Application to the study of electrical impulses in the brain: Hodgkin-Huxley equation 12/9
SEMINAR ON APPLICATIONS Diffusion equations: deduction; linear and nonlinear 16/9	(PDEs) Linear diffusion eq.: separation of variables and the need for Fourier series 17/9	(PDEs) Fourier series: definition, examples, convergence results 18/9	(PDEs) Application of Fourier series to linear diffusion eq. 19/9
SEMINAR ON APPLICATIONS Wave equations: deduction; conservation equation in a traffic flow problem 23/9	(PDEs) Application of Fourier series for the wave equation in a bounded interval 24/9	(PDEs) Wave equation in unbounded intervals: d'Alembert solution and propagation of waves 25/9	(PDEs) Nonlinear waves (Burgers' equation) shocks and need for a weaker notion of solution 26/9

GMIT

Differential Equations and Applications

Session 1: Applications of Differential Equations: population dynamics, radioactive decay, gravitational pendulum.

GMIT

2/9/2024 - 27/9/24

Differential Equations and Applications

this notes support the sessions of the course and are to be complemented with daily exercise sheets illustrating the topics explored in the sessions.

the extension of the course is 4 weeks, with 4 sessions each week of 1h30 each. As the students attending the course are students of Science and Engineering, not of Mathematics, the emphasis is more on motivation, presenting the methods and illustrating them with examples, than on proving theorems (although there are some proofs!).

Sessions 1 - 8 (first 2 weeks) are about ODEs, and Sessions 9 - 16 (last 2 weeks) are about PDEs, namely the linear diffusion and wave equations and its study using Fourier series. In every week there is always one session (usually the first of the week) titled "Applications of Differential Equations", where the differential equations to be studied are introduced as models to some ("simple") problems in Science or Engineering, or some other applications

— Why are differential equations important?
— Some examples.

- Humans are better at understanding small (local) changes than big (global) ones. [back to Newton, in 17th Century!]

- This leads to the consideration of <u>rates of change</u> at a given instant (<u>instantaneous</u> rates of change), which can be measured by the operation of "taking derivatives"

$$\underbrace{\underbrace{\lim_{h\to 0} \frac{f(t+h)-f(t)}{h}}_{\text{rate of change}} =: f'(t)}_{\text{instantaneous rate of change}}$$

- this leads directly to <u>(ordinary) differential equations</u>, which are equations that relate the <u>instantaneous rate of change of a quantity</u> with the <u>value of that quantity</u>, or to a function of that quantity.

Examples

1. From a population dynamics modelling situation

$N(t)$: density of a given population in a given region (population sufficiently large) at time t. Suppose the region is isolated

$$N(t+\Delta t) = N(t) + (\text{births in } \Delta t) - (\text{deaths in } \Delta t)$$

$$N(t+\Delta t) = N(t) + \underbrace{\beta(t, N(t))}_{\text{per capita birth rate}} N(t)\,\Delta t - \underbrace{\mu(t, N(t))}_{\text{per capita death rate}} N(t)\,\Delta t$$

$$\frac{N(t+\Delta t) - N(t)}{\Delta t} = \beta(t, N(t))\,N(t) - \mu(t, N(t))\,N(t)$$

$$\lim_{\Delta t \to 0} \frac{N(t+\Delta t) - N(t)}{\Delta t} =: N'(t) = (\beta - \mu)(t, N(t))\,N(t)$$

So we get the (ordinary) differential equation

$$N'(t) = r(t, N(t))\,N(t) \qquad (*)$$

where $r(t, N(t)) := \beta(t, N(t)) - \mu(t, N(t))$

is the net rate change per capita.

As a first approximation, usually valid if the population $N(t)$ is small, we can consider that $r(t, N(t))$ is constant, in which case $(*)$ becomes

$$N'(t) = r \cdot N(t) \qquad (**)$$

When either t or $N(t)$ is large, in many cases $r(t, N(t))$ can no loger be reasonably assumed to be constant either because of seasonal effects or by large numbers in the population affecting β or μ. A simple case is when $r(t, N(t)) = r - \alpha N(t)$, giving rise to the ODE

$$N'(t) = (r - \alpha N(t)) N(t) \qquad (***)$$

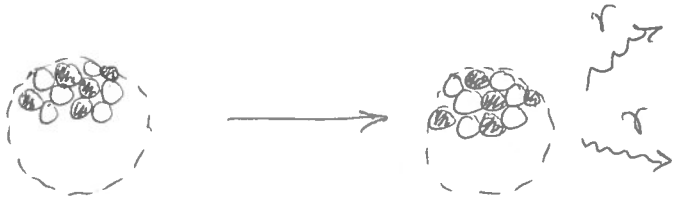
Examples $(**)$ and $(***)$, although they are very simple models, are quite relevant in applications: they are the <u>linear growth equation</u> (giving rise to the exponential increase law, as we will see)

and the logistic growth equation, respectively.

An important situation in which $r(t, N(t))$ do not change with t or $N(t)$, whatever their values might be, is the case of radioactive decay that we will consider next.

In radioactive chemical elements the nucleus of the atom is unstable and can decay into other nucleus after some time. Typical processes are:

γ-decay



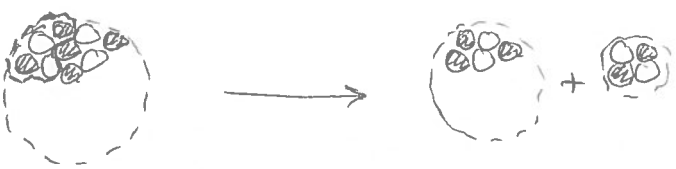
$[\ X \longrightarrow X + \gamma\text{-ray photons}\ ]$

β-decay

$+ e^- + \bar{\nu}_e$

$X \begin{cases} p \text{ protons} \\ n \text{ neutrons} \end{cases}$ $\qquad$ $Y \begin{cases} p+1 \text{ protons} \\ n-1 \text{ neutrons} \end{cases}$

α-decay



$X \begin{cases} p \text{ protons} \\ n \text{ neutrons} \end{cases}$ $\qquad$ $Y \begin{cases} p-2 \text{ protons} \\ n-2 \text{ neutrons} \end{cases}$

In all cases, denoting by X the number of nucleus of atoms of the element X (an abuse of language...) the rate of change of X is related with the value of X by equation (*), which takes the form

$$\frac{dX}{dt} = -kX$$

where the constant $k > 0$ depends on the element but not on its quantity, neither on time.

If a sequence of these reactions takes place (which is usually the case) we get a system of ODEs for the concentration of the several radioactive species involved. An example with 3 species, (X, Y, Z), that decay radioactively, and a final stable element, W, is now presented:

$$X \xrightarrow{k_x} Y \xrightarrow{k_y} Z \xrightarrow{k_z} W$$

$$\begin{cases} x' = -k_x X \\ Y' = k_x X - k_y Y \\ Z' = k_y Y - k_z Z \\ W' = k_z Z \end{cases}$$

or, in vector notation

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}' = \begin{bmatrix} -k_x & 0 & 0 & 0 \\ k_x & -k_y & 0 & 0 \\ 0 & k_y & -k_z & 0 \\ 0 & 0 & k_z & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \qquad (\dagger)$$

which can be written in a more compact notation as

$$\underline{x}' = A\,\underline{x}$$

where $\underline{x} = \underline{x}(t) : I \subset \mathbb{R} \to \mathbb{R}^4$ is the vector $\underline{x} = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$

and A is the matrix is $(\dagger)$

2. Let us consider now the equation describing the movement of a gravitational pendulum without friction or damping, and the vibrating spring.

the whole of Classical Mechanics is essentially an exploration of Newton's laws of motion. One of them leads directly to differential equations:

$$f = m\,a$$

force — mass — acceleration

Why is this a differential equation? Note that

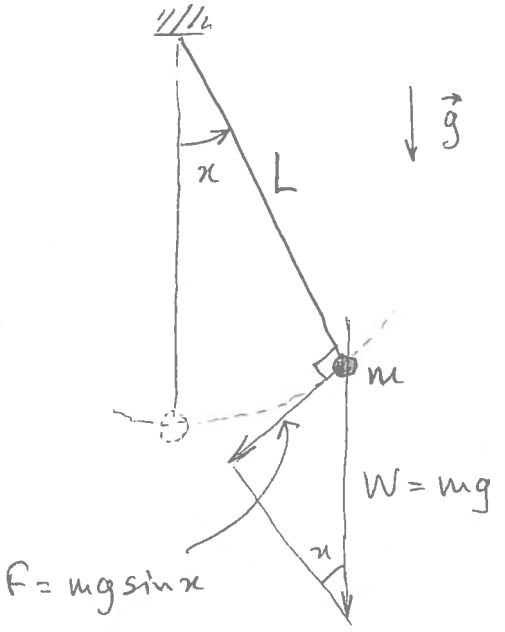
$$a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}$$

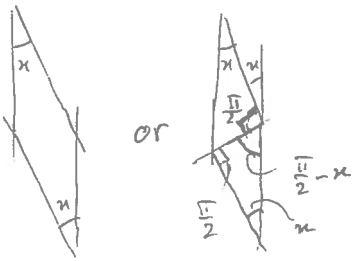
where v is the velocity and x the displacement. Typically $f = f(x, v) = F\left(x, \frac{dx}{dt}\right)$, and so $F=ma$ is, in fact

$$\frac{d^2x}{dt^2} = \frac{1}{m} F\left(x, \frac{dx}{dt}\right)$$

where the force f describes the physical problem we are studying.

Let us consider the gravitational pendulum





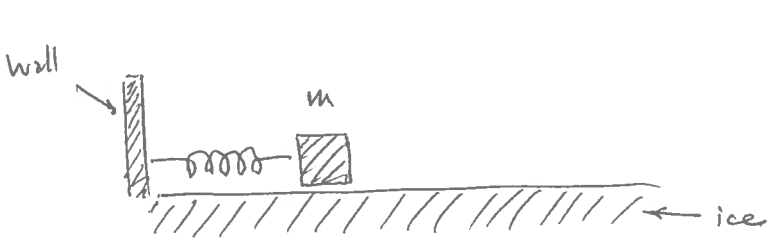
thus, the second law of Newton gives

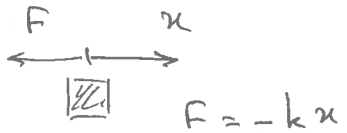
$$m(Lx)'' = -mg\sin x$$

$$\boxed{x'' = -\frac{g}{L}\sin x} \qquad (\sharp)$$

which is an ordinary differential equation!

A similar case is the vibration of a mass-spring system without friction:





$$\boxed{mx'' = F(x) = -kx} \qquad (\sharp\sharp)$$

Hooke's law, valid for small displacements $x(t)$.

Note that if we consider only small displacements in $(\sharp)$ then, since $\sin x \approx x$, equation $(\sharp)$ becomes (approximatelly!) like $(\sharp\sharp)$. Are the solutions of $(\sharp)$ for small $x(t)$ well approximated by the solutions of $(\sharp\sharp)$? this an important problem and we'll see later in the

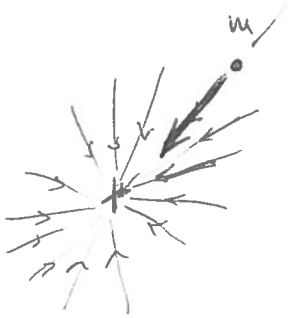
course tools that allow us to answer it, under appropriate assumptions.

3. To finish this set of examples let us consider the case of Celestial Mechanics, an historically important problem that is still relevant today.

[See: Murray & Dermott: Solar System Dynamics, and Geiges: the Geometry of Celestial Mechanics]

One particle movement in a central force field F, where $F(\vec{x}) = f(\|\vec{x}\|)\frac{\vec{x}}{\|\vec{x}\|}$ satisfies Newton's equation $m\vec{x}'' = F(\vec{x})$. In Newton's gravitation theory

$$f(r) = -\frac{\mu m}{r^3}$$



where $\mu = GM$, with $G = 6,67408 \times 10^{-11}\ m^3 kg^{-1} s^{-2}$ and M is mass of the body "creating" the field, m is the mass of the "moving" particle.

So, in the 3-dim space $\mathbb{R}^3$ the equation of motion of the particle is

$$x_i'' = -\mu \frac{x_i}{(x_1^2 + x_2^2 + x_3^2)^{3/2}}, \quad i = 1,2,3$$

For a system of N gravitating bodies we get a system of $3N$ ODEs for the positions of all the particles

GMIT

Differential Equations and Applications

Session 2 : Basic notions of Ordinary Differential Equations

Revisions : linear and separable equations ; the use of changes of variables

— Basic notions
— Revisions: linear and separable ODEs
— Using changes of variables

Given an ordinary differencial equation

$$F(t, x(t), x'(t), \dots, x^{(m)}(t)) = 0 \qquad (*)$$

- the order of the equation is the higher order of derivation occurring in the DE, in the above case: m;
- the equation $(*)$ is called autonomous if F does not depend explicitly on t, i.e., if $(*)$ is of the Type

$$f(x(t), x'(t), \dots, x^{(m)}(t)) = 0;$$

- the equation $(*)$ is linear if the function

$$(x_0, \dots, x_m) \longmapsto F(t, x_0, x_1, \dots, x_m) - F(t, 0, \dots, 0)$$

is a linear function;

Defining auxiliary unknown functions in a convenient way (more about this later) an m^{th} order differential equation can be transformed into a system of m differential equations of 1^{st} order.

So, we will concentrate our efforts in 1^{st} order systems.

Definition

Let $f : D \subset \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$, where D is an open set, be a continuous function

Consider the ODE defined by

$$x' = f(t, x) \qquad (*)$$

for some unknow function $x = x(t)$.

A function $x : (a, b) \subset \mathbb{R} \longrightarrow \mathbb{R}^n$ is a solution of $(*)$ if (i) $x \in C^1$

(ii) $(t, x(t)) \in D$, $\forall t \in (a, b)$ [or $x(t) \in D$ if f aut

(iii) $x'(t) = f(t, x(t))$, $\forall t \in (a, b)$

(We can extend this definition to non-open intervals by considering lateral derivatives when needed.)

We now consider some instances where we can solve equations explicitly. These cases should be known from undergraduate courses but it is important to refresh our memory about them.

Examples

0) the simplest differential equation is $x' = b(t)$, for some continuous function $b(\cdot)$. this is "just" a primitivation problem: the solutions are $x(t) = \int b(t) + C$.

1) the simplest authentic differential equation is the linear equation $x' = ax$, which we saw as a model of population dynamics. let's study it carefully.

- clearly the function $x(t) = 0, \forall t$, is a solution of $x' = ax$; check: (i) x is $C^1(\mathbb{R})$, (ii) $x(t) \equiv 0 \in D_{ax} = \mathbb{R}$
  (iii) $x' = 0 = a0 = ax$ for all t.

- now look for solutions that are not identically zero (if they exist...):

  Suppose we have $x(t)$ which is not zero everywhere. So it is $\neq 0$ at some t value and hence (because we are looking for solutions, which need to be C^1, and hence continuous functions) we have $x(t) \neq 0$ is some interval of t. In that interval we can divide $x' = ax$ by $x(\cdot)$ obtaining $\frac{x'}{x} = a$:

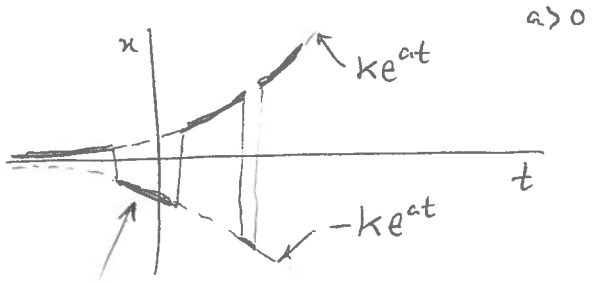
$$\frac{x'}{x} = a \iff \int \frac{x'(t)}{x(t)}\,dt = \int a\,dt$$

$$\iff \int \frac{1}{x(t)} \frac{dx(t)}{dt}\,dt = \int a\,dt$$

$$\Leftrightarrow \int \frac{1}{x}\,dx = at + C$$

$$\Leftrightarrow \log|x(t)| = at + C$$

$$\Leftrightarrow |x(t)| = e^C e^{at} =: Ke^{at}$$



this is not possible because $x(t)$ must be C^1 (and hence C^0) to be a solution of $x' = ax$.

So, if $|x(t)| = Ke^{at}$, we must have

$$x(t) = (\text{constant})\, e^{at}$$

where $(\text{constant}) \in \mathbb{R}\setminus\{0\}$.

Note, however, that if we put $(\text{constant}) = 0$ we get the solution $x(t) \equiv 0, \forall t$, that we considered before. We conclude that all functions of the form

$$x(t) = \alpha e^{at}, \qquad (*)$$

with $\alpha \in \mathbb{R}$, are solutions of $x' = ax$.

To fix one of these functions, i.e., to fix the constant α we need to know the value of $x(t)$ at some value of $t = t_0$;

the resulting problem is

$$(*) \quad \begin{cases} x' = a x \\ x(t_0) = x_0 \end{cases}$$

where (t_0, x_0) is a given datum.
This is called an initial value problem and (t_0, x_0) is the initial condition, or initial datum.

Using the initial condition in (x) we get

$$x_0 = x(t_0) = \alpha e^{a t_0} \iff \alpha = x_0 e^{-a t_0}$$

which means that the solution of the ODE is

$$x(t) = x_0 e^{a(t-t_0)}.$$

An equivalent way to obtain this result is that, if we have a problem $(*)$ we can repeat the above process but integrate in t between t_0 and some arbitrary $\hat{t}$, instead of computing the primitive:

$$\frac{x'}{x} = a \iff \int_{t_0}^{\hat{t}} \frac{1}{x(t)} \frac{dx(t)}{dt} dt = \int_{t_0}^{\hat{t}} a \, dt$$

$$\iff \int_{x(t_0)}^{x(\hat{t})} \frac{1}{x} dx = \int_{t_0}^{\hat{t}} a \, dt$$

$$\Leftrightarrow \quad \log\left|\frac{x(\hat{t})}{x_0}\right| = a(\hat{t}-t_0)$$

$$\Leftrightarrow \quad \left|\frac{x(\hat{t})}{x_0}\right| = e^{a(\hat{t}-t_0)}$$

and by the same argument as before we have

$$\frac{x(\hat{t})}{x_0} = e^{a(\hat{t}-t_0)}$$

which is the previous expression (changing $\hat{t}$ to t)

$$x(t) = x_0 e^{a(t-t_0)}.$$

Clearly, if the equation is $x' = a(t)x$ where $a(\cdot)$ is a nonconstant continuous function the same process is valid and gives

$$\frac{x'}{x} = a(t) \quad \Leftrightarrow \quad \underbrace{\int_{t_0}^{\hat{t}} \frac{1}{x(t)}\frac{dx(t)}{dt}\,dt}_{\text{this is the same as before}} = \int_{t_0}^{\hat{t}} a(t)\,dt$$

$$\vdots$$

$$\Leftrightarrow \quad x(t) = x_0 e^{\int_{t_0}^{t} a(s)\,ds} \qquad (**)$$

Are there any other functions that might be solution of $x' = a(t)x$? Actually: is $(**)$ really a solution of $(*)$?

this last question is easy to answer:

(i) because $a(\cdot)$ is continuous, $\int_{t_0}^{t} a(s)ds$ is C^1 and hence $x(t)$ is also C^1

(ii) the values of $x(t)$ are in $\mathbb{R}$, so $(t, x(t))$ are in the set $D_a \times \mathbb{R}$, which is the domain of $a(t)x$

(iii) clearly $x'(t) = \frac{d}{dt} x_0 e^{\int_{t_0}^{t} a(s)ds} = x_0 a(t) e^{\int_{t_0}^{t} a(s)ds} = a(t)x(t)$ is valid $\forall t \in D_a$.

the first question is what is known as a uniqueness result. It is somewhat more abstract. If our "guess" is that all solutions of (*) are given by (**) then a possible way to prove that no other solutions exist is to assume that they do and try to get a contradiction.

Let us see how this works in this simple case:

To prove that $x(t) = \alpha e^{\int_{t_0}^{t} a(s)ds}$ is the only solution of

$$\begin{cases} x' = a(t)x \\ x(t_0) = \alpha \end{cases}$$

suppose there is another solution $y(t)$.

Compute $\left(\frac{y}{x}\right)'$:

$$\left(\frac{y}{x}\right)' = \frac{y'x - yx'}{x^2} = \frac{a(t)y\alpha e^{\int_{t_0}^{t} a(s)ds} - y\,a(t)\alpha e^{\int_{t_0}^{t} a(s)ds}}{x^2} = 0, \ \forall t$$

So, if $\left(\frac{y}{x}\right)' = 0$, $\forall t$, we can conclude that

$$\frac{y}{x}(t) = \text{constant}, \quad \forall t$$

But putting $t = t_0$ we have $\frac{y}{x}(t_0) = \frac{y(t_0)}{x(t_0)} = \frac{\alpha}{\alpha} = 1$

which, together with the above gives constant $= 1$, i.e.,

$$y(t) = x(t), \quad \forall t.$$

This is a contradiction with the assumption that $y(t)$ was <u>another</u> solution, and proves that the solution given by (**) is the only solution of (*).

2) Another class of equations for which we can get explicit (or "almost explicit") solutions are the so called <u>separable equations</u>. these are of the form

$$\frac{dy}{dt} = g(t)\,h(y). \qquad (\times)$$

Suppose $h(\cdot)$ is not identically zero (if it is the equation is $\frac{dy}{dt} = 0$, which we have considered before).

Let's assume $g(\cdot)$, $h(\cdot)$ are continuous. the process used before (for the case $h(y) \equiv y$) can be used here

again: separating the variables y and t and integrating both sides of the resulting equation in the variable t:

$$\frac{1}{h(y)}\frac{dy}{dt} = g(t) \iff \int \frac{1}{h(y)}\frac{dy}{dt}\,dt = \int g(t)\,dt + C$$

$$\| $$

$$\int \frac{1}{h(y)}\,dy$$

$$\| $$

$$H(y)$$

thus, the solutions of the ODE are given by solutions of

$$H(y) = \int g(t)\,dt + C$$

Since, by assumption, $H'(y) = \frac{1}{h(y)} \neq 0$, the inverse function theorem can be applied to guarantee that for every y there is a neighborhood for which H has a differentiable local inverse $H^{-1}(\cdot)$ and thus the solution of (X) can be written as

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

Let's see a concrete example of this argument.

Let us consider the initial value problem

$$(\ast)\quad \begin{cases} \dfrac{dy}{dt} = ty^3 \\ y(0) = \dfrac{1}{2} \end{cases}$$

$$\frac{dy}{dt} = ty^3 \iff \frac{1}{y^3}\frac{dy}{dt} = t$$

$$\iff \int \frac{1}{y^3}\frac{dy}{dt}\,dt = \int t\,dt$$

$$\iff \int \frac{1}{y^3}\,dy + C = \frac{1}{2}t^2$$

$$\iff \frac{1}{-2}y^{-2} + C = \frac{1}{2}t^2$$

$$\iff y^{-2} = -t^2 + 2C$$

as the inverse of $y \mapsto y^{-2}$ is $w \mapsto w^{-1/2}$ (note that $(y^{-2})^{-1/2} = y$) we conclude that

$$y(t) = (2C - t^2)^{-1/2} = \frac{1}{\sqrt{2C - t^2}}$$

Observing the initial condition is $y(0) = \frac{1}{2}$ we have

$$\frac{1}{2} = y(0) = \frac{1}{\sqrt{2C - 0}} = \frac{1}{\sqrt{2C}} \iff \sqrt{2C} = 2$$

$$\iff 2C = 4$$

$$\Rightarrow C = 2$$

thus a solution of (*) is $x(t) = \dfrac{1}{\sqrt{4-t^2}}$, whose interval of existance is $t \in (-\sqrt{2}, \sqrt{2})$.

[Note that, by definition, solutions are defined on open intervals that — naturally! — must contain the initial "time" t_0.]

3) to conclude we consider an impotnt tool in Mathematics (not only in differential equations) which is the change of variables.

Unfortunately there is no rule (or trick) of universal applicability to discover the "best" change of variables to a given situation / equation (if such a change exists...)

So, we will present an example of such a technique.

- Consider the ODE

$$y' = f\left(\frac{y}{t}\right)$$

with f sufficiently smooth (say C^1).

Since f depends on t and y only through the ratio $\frac{y}{t}$ it is reasonable to call this $z(t)$

and try to see if the resulting equation in the (t,z) variables is easier than in the (t,y).

$$z(t) = \frac{y(t)}{t} \iff y(t) = t\,z(t)$$

$$\Rightarrow \quad y' = z(t) + t\,z'(t)$$

thus $\quad y' = f\left(\frac{y}{t}\right) \Rightarrow \quad z + t\,z' = f(z)$

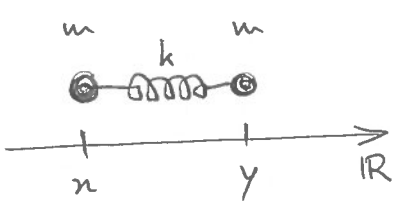
and this last equation is a separable equation:

$$z + t\,z' = f(z) \iff t\,z' = f(z) - z$$

$$\iff \frac{1}{f(z) - z}\, z' = \frac{1}{t}$$

this can be integrated as seen before.

- Another example, more interesting from the point of view of applications is the following:



natural length of spring: ℓ
spring stiffness: k

$$(\dagger) \quad \begin{cases} m x'' = k\,((y-x) - \ell) & \text{[if } y-x > \ell\text{, } x \text{ goes to right]} \\ m y'' = -k\,((y-x) - \ell) & \text{[if } y-x > \ell\text{, } y \text{ goes to left]} \end{cases}$$

Observing that the rhs of $(\dagger)$ involve just the combination $y-x$ of the variables y and x it seems natural to

Consider a new variable $v := y - x$.
the equation for this new variable is

$$v'' = y'' - x'' = -\frac{k}{m}(v-l) - \frac{k}{m}(v-l) = -2\frac{k}{m}(v-l)$$

$$v'' + 2\frac{k}{m}v - 2\frac{kl}{m} = 0$$

which is like the vibrating mass equation (but with an extra constant term $-2\frac{kl}{m}$).

This does not allow us to compute x and y separately. For this we need to plug the solution $v(t)$ into the equations for x and for y, written as

$$(\ddagger) \begin{cases} m x'' = k(v-l) \\ m y'' = -k(v-l) \end{cases}$$

and solve them for $x(t)$ and $y(t)$. Note that, once we know $v(t)$ these are just primitivation exercises (twice for each equation).

It is interesting to observe that if we let $u := y + x$ then $(\ddagger)$ gives (summing the two equations)

$$u'' = 0$$

which implies that $u(t) = \alpha t + \beta$, where $\alpha, \beta \in \mathbb{R}$ are constants. Thus, making the change of variables

$$(x, y) \mapsto (u; v)$$

where $u := y + x$, $v := y - x$, we get from (†) to the system

$$(\ddagger) \quad \begin{cases} u'' = 0 \\ v'' + 2\frac{k}{m} v - 2\frac{kl}{m} = 0 \end{cases}$$

which is far easier to study because the two equations are now decoupled! After getting the solutions to (‡) we get the solutions $x(t)$ and $y(t)$ of (†) by inverting the change of variables: $y(t) = \frac{u(t) + v(t)}{2}$, $x(t) = \frac{u(t) - v(t)}{2}$.

[Note: observe that $\frac{1}{2} u(t)$ is the coordinate of the center of mass of the system.]

Another technique that is very useful (in some cases) is the multiplication by an integrating factor. This is exemplified next in a very important case: the linear equation with non-homogeneous right-hand side.

Let us consider the equation and initial value

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

then, writing

$$x' - a(t)x = b(t)$$

let us multiply by a function $\mu = \mu(t)$ and try to find μ such that the left-hand side is a total derivative with respect to t

$$\underbrace{\mu x' - \mu a(t) x} = \mu(t) b(t)$$

$$\stackrel{?}{=} \frac{d}{dt}(\mu x) = \mu x' + \mu' x$$

this can be done if $\mu' = -\mu a(t)$, i.e., choosing

$$\mu(t) = e^{-\int_{t_0}^{t} a(s)\,ds}.$$

Then the equation becomes

$$\frac{d}{dt}\left(e^{-\int_{t_0}^{t} a(s)\,ds} x(t) \right) = b(t)\, e^{-\int_{t_0}^{t} a(s)\,ds}$$

integrating both terms in t

$$\int_{t_0}^{t} \frac{d}{d\tau}\left(e^{-\int_{t_0}^{\tau} a(s)\,ds} x(\tau) \right) d\tau = \int_{t_0}^{t} b(\tau)\, e^{-\int_{t_0}^{\tau} a(s)\,ds}\, d\tau$$

$$e^{-\int_{t_0}^{t} a(s)\,ds} x(t) - 1 \cdot x_0 = \int_{t_0}^{t} b(\tau)\, e^{-\int_{t_0}^{\tau} a(s)\,ds}\, d\tau$$

and thus

$$x(t) = x_0 e^{\int_{t_0}^{t} a(s)ds} + \int_{t_0}^{t} b(\tau)\, e^{\int_{\tau}^{t} a(s)ds}\, d\tau.$$

Another way to write this expression is

$$x(t) = \left(x_0 + \int_{t_0}^{t} b(\tau)\, e^{-\int_{t_0}^{\tau} a(s)ds}\, d\tau \right) e^{\int_{t_0}^{t} a(s)ds}.$$

Both are called the variation of constants formula because the constant x_0 that shows up in the solution $x(t) = x_0 e^{\int_{t_0}^{t} a(s)ds}$ of the homogeneous equation

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

is now a function of t, in parentesis above, and thus the constant has changed...

GMIT

Differential Equations and Applications

Session 3: Existence and uniqueness of solutions: Picard-Lindelöf theorem. Extension of solutions to maximal intervals. Continuous dependence on initial data and parameters

We saw that for some classes of equations we can, in some cases, have explicit formulas for the solutions. However, this is a rather rare case: given a differential equation we do not usually have a way to write its solutions in terms of known functions of Mathematical Analysis (polynomials, exponential, trigonometric functions, and their inverses, together with a finite number of algebraic operations $(+, -, \times, /)$ and compositions.

So, how can we be sure that, in such cases, a given differential equation has a solution?

The answer is provided by an <u>existence result</u>: a theorem that, under certain conditions on f, guarantees that the differential equation $x' = f(t, x)$ has a solution.

There are several existence theorems (under diverse assumptions on f). We will present the more common one, which also guarantees the uniqueness of solutions to the initial value problems. It is the celebrated <u>Picard-Lindelöf</u> (or Cauchy-Lipschitz) <u>theorem</u>.

the idea behind the proof of the theorem is to obtain the seeked for solution as the limit as $n \to \infty$ of a sequence of better and better approximations $x_n(t)$.

As the derivatives do not behave nicely with taking limits (for example: the limit of a sequence of differentiable functions can be nondifferentiable everywhere!) it is better to work with the integral version of the initial value problem:

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

↓ integrate $\left(\int_{t_0}^{t} \cdot\, dt\right)$ in both sides of the equation

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

Proposition 1

Let $f : D \subset \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a continuous function on an open set D; let $(t_0, x_0) \in D$

A function $x(t)$ is a solution of $(*)$ in $(a,b) \ni t_0$ iff the integral equation $(**)$ is satisfied $\forall t \in (a,b)$

Proof

"If" part ($\Leftarrow$):

Let $x(t)$ be a (continuous) solution of (**). Then, because f is continuous we have that $t \mapsto f(t, x(t))$ is also continuous, and thus $\int_{t_0}^{t} f(s, x(s))\,ds$ is a continuously differentiable function of t. Thus, by (**), $x(t)$ [in the left-hand side of (**)] is C^1 and differentiating both sides of (**) we get $x'(t) = f(t, x(t))$, $\forall t$. Clearly, from (**) computed at t_0 we get $x(t_0) = x_0$. Thus $x(t)$ is a solution of (*).

"Only if" part ($\Rightarrow$):

Let $x(\cdot)$ be a solution of (*). Then $x(\cdot) \in C^1 \subset C^0$ and $f(t, x(t))$ is a continuous function of t, hence it is integrable and integrating (*) in the interval with boundary t_0 and t we get (**), and so $x(\cdot)$ is a solution of (**). ∎

So, from now on, to prove the existence of a solution to (*) we rather prove the existence of a solution to (**) as the integrals behave better in passing to the limit and in obtaining [illegible] inequalities

How can we be sure that (under appropriate conditions on the fuction f) there is a solution $x(t)$ of (**)? meaning: there is a fuction $x(t)$ which is continuous and satisfy the equality

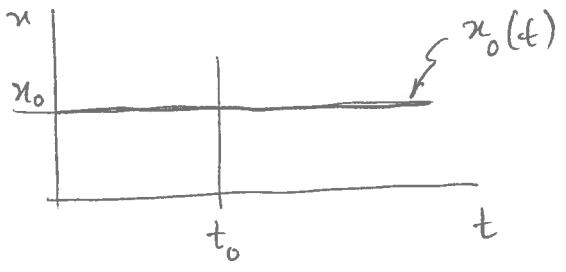
$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds .$$

Note that such a fuction must satisfy (*), i.e., it must satisfy two conditions:

(1) the graph must pass in the point (t_0, x_0)

(2) it must satisfy $x'(t) = f(t, x(t))$.

Condition (2) is hard, but (1) is very easy to fulfill: What's the essiest continuous fuction that satisfy (1)? clearly it is the constant fuction $x_0(t) = x_0, \forall t$



to check if $x_0(t)$ is a solution of (**), plug it into the right-hand side and check if it is

equal to the left-hand side:

is $x_0 + \int_{t_0}^{t} f(s, x_0(s))\,ds$ equal to $x_0(t)$?

This is the same as asking

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

Usually it is not: $x_0 + \int_{t_0}^{t} f(s, x_0)\,ds$ is a function different from $x_0(t)$. Call it $x_1(t)$:

$$x_1(t) := x_0 + \int_{t_0}^{t} f(s, x_0(s))\,ds.$$

This function $x_1(t)$ is clearly continuous (why?) and satisfy $x_1(t_0) = x_0$ (why?). Is it a solution of (**)? Same procedure: plug it in (**) and see if $x_0 + \int_{t_0}^{t} f(t, x_1(s))\,ds$ is equal to $x_1(t)$.

Usually it is not: it is a different function, call it $x_2(t)$.

In this way we get a sequence $(x_n(t))_{n \in \mathbb{N}}$ defined as

$$(P) \begin{cases} x_0(t) \equiv x_0 \\ x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, x_n(s))\,ds \end{cases}$$

If for some value of n we have $x_{n+1}(t) = x_n(t)$ then we got a solution to (**). If not, maybe we can have that in the limit as $n \to \infty$, if the limit exists.

Suppose the limit exists, i.e., there exists a continuous function $x(t) := \lim_{n\to\infty} x_n(t)$. then, if we can perform the operations indicated

$$x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

(if $\exists \lim$) $\Big\downarrow n\to\infty$ $\qquad\qquad$ $\Big\downarrow n\to\infty$ (if $\exists \lim$ and $\lim\int = \int\lim$)

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

then we see that we obtain the equality (**) with $x(\cdot)$, i.e., the limit function is a solution of (**), and hence of (*).

What can we do to make this idea work?

Maybe it is better to try and see if it works in a simple case for which we already now that a solution exists (and even know how to write it...)

Does this process works in the case

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

[Note that we know that the only solution is $x(t) = e^t$, so we better get this result in the end!]

In (X) we have $f(t,x) \equiv x$ and the sequence becomes

$$x_0(t) \equiv 1$$

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

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

$$x_3(t) = 1 + \int_0^t \left(1 + s + \frac{s^2}{2}\right) ds = 1 + t + \frac{t^2}{2} + \frac{t^3}{3\cdot 2}$$

$$x_4(t) = 1 + \int_0^t \left(1 + s + \frac{s^2}{2} + \frac{s^3}{3\cdot 2}\right) ds = 1 + t + \frac{t^2}{2} + \frac{t^3}{3\cdot 2} + \frac{t^4}{4\cdot 3\cdot 2}$$

We can actually prove (by induction) that

$$x_n(t) = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^n}{n!} = \sum_{j=0}^{n} \frac{t^j}{j!}$$

Clearly this sum is the partial sum of the series defining the exponential function

$$e^t := \sum_{j=0}^{\infty} \frac{t^j}{j!}$$

So we have that the limit of $x_n(t)$ as $n \to \infty$ exists $x_n(t) \xrightarrow[n\to\infty]{} e^t$ and the limit function is the known solution of (X).

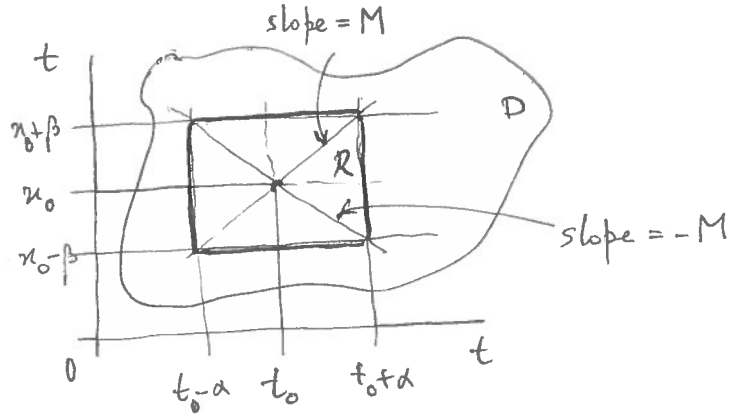
These are good news! The process works in at least one case! Let us now see what can we do for the case of a more general function f.

We need to prove that the (so called) Picard iterates (P)

(i) are a convergent sequence, or have a convergent subsequence

(ii) the limit is a continuous solution of (**), and hence, by Proposition 1, solves (*).

Prior to (i) we need to guarantee that the sequence is well defined, meaning: if $x_n(t)$ is such that $(t, x_n(t)) \in D$ then $(t, x_{n+1}(t))$ will also belong to D; for this, consider $t \in [t_0 - \alpha, t_0 + \alpha]$. for some $\alpha > 0$, and try to choose α such that $(t, x_n(t)) \in D$ when t is in this interval. Let B_β be the ball $B_\beta := \{x \in \mathbb{R}^n : \|x - x_0\| \leq \beta\}$ with β such that $R := I_\alpha \times B_\beta \subset D$. Furthermore, let's refine R as follows:



Let $M = \sup_{(t,x)\in R} \|f(t,x)\|$ and choose α, β (and hence R) such that $M\alpha = \beta$. Then we have

$$\|x_{n+1}(t) - x_0\| = \left\| \int_{t_0}^{t} f(s, x_n(s))\, ds \right\|$$

$$\leq \left| \int_{t_0}^{t} \|f(s, x_n(s))\|\, ds \right| \leq \left| \int_{t_0}^{t} M\, ds \right|$$

$$= M|t - t_0| \leq M\alpha = \beta.$$

Thus, with this choice of α and β the sequence is such that $(t, x_n(t)) \in R \subset D$, $\forall n$, and hence we are sure all the $x_n(t)$ are well defined.

To prove that $(x_n(\cdot))$ converges when $n \to \infty$ we prove that, for each $t \in I_\alpha$ fixed, the sequence of real vectors in $\mathbb{R}^n$, $x_n(t)$, is a Cauchy sequence, meaning:

$$\forall_{\varepsilon>0} \; \exists_{p\in\mathbb{N}_1} : \; \forall_{j>p,\, k\geq 1}, \; \|x_{j+k}(t) - x_j(t)\| < \varepsilon$$

Let us start with the case of $k=1$:

$$\|x_{j+1}(t) - x_j(t)\| = \left\| \cancel{x_0} + \int_{t_0}^{t} f(s, x_j(s))\,ds - \cancel{x_0} - \int_{t_0}^{t} f(s, x_{j-1}(s))\,ds \right\|$$

$$\leq \left| \int_{t_0}^{t} \| f(s, x_j(s)) - f(s, x_{j-1}(s)) \|\,ds \right| \qquad (*)$$

to estimate this last term we need to estimate $\| f(s, x_j(s)) - f(s, x_{j-1}(s)) \|$ in terms of $\| x_j(s) - x_{j-1}(s) \|$. In order to do this we need to assume a condition upon the function f, which is that $f(t,x)$ satisfies the following:

> there exists $L>0$ such that, for all $(t,x), (t,x^*)$ in R the following holds:
>
> $$\| f(t,x) - f(t,x^*) \| \leq L \| x - x^* \|.$$

A function such that, for all compact $R \subset D$ there exists a $L>0$ satisfying the above condition is called a <u>Locally Lipschitz function</u> relative to the variable x.

[<u>Obs.</u>: A continuously differentiable function in R is locally Lipschitz in R: this is a consequence of Lagrange's theorem

Now, if f is Locally Lipschitz relative to the variable x we can proceed the inequality (x) as follows

$$(x) \leq \left| L \int_{t_0}^{t} \| u_j(s) - u_{j-1}(s) \| \, ds \right|$$

$$\leq \alpha L \max_{s \in I_\alpha} \| u_j(s) - u_{j-1}(s) \|$$

Proceeding inductively we get

$$\| u_{j+1}(t) - u_j(t) \| \leq (\alpha L)^j \max_{s \in I_\alpha} \| u_1(s) - u_0 \| \leq (\alpha L)^j \beta. \tag{xx}$$

Now, if $k > 1$ we can write

$$\| u_{j+k}(t) - u_j(t) \| = \| u_{j+k}(t) - u_{j+k-1}(t) + u_{j+k-1}(t) - \cdots - u_j(t) \|$$

$$\leq \| u_{j+k}(t) - u_{j+k-1}(t) \| + \cdots + \| u_{j+1}(t) - u_j(t) \|$$

$$= \sum_{l=j}^{j+k-1} \| u_{l+1}(t) - u_l(t) \|$$

(using (xx)) ↓

$$\leq \sum_{l=j}^{j+k-1} (\alpha L)^l \beta = (\alpha L)^j \beta \sum_{l=0}^{k-1} (\alpha L)^l$$

$$\leq (\alpha L)^j \beta \sum_{l=0}^{\infty} (\alpha L)^l$$

$$= (\alpha L)^j \beta \, \frac{1}{1 - \alpha L} \tag{xxx}$$

thus, reducing α if needed so that $\alpha L < 1$ (in which case we need to reduce β so that in the new R we have $M\alpha = \beta$) we have that $(\alpha L)^j \longrightarrow 0$ as $j \to \infty$ and thus $\| x_{j+k}(t) - x_j(t) \| \longrightarrow 0$ as $j \to \infty$, for all k, and for fixed $t \in I_\alpha$. As the right-hand side of (xxx) does not depend on t, the above holds for all t in I_α.

This proves the Picard iterates sequence is a Cauchy sequence and thus it is convergent, i.e., there exists a fuction $x(t)$ defined on I_α such that $\lim_{n\to\infty} x_n(t) = x(t)$.

Proving that $x(t)$ is continuous in I_α is easy: for all $t, t^* \in I_\alpha$ write

$$\| x(t) - x(t^*) \| = \| (x(t) - x_j(t)) + (x_j(t) - x_j(t^*)) + (x_j(t^*) - x(t^*)) \|$$

$$\leq \underbrace{\| x(t) - x_j(t) \|}_{\downarrow\, j\to\infty,\ 0} + \underbrace{\| x_j(t) - x_j(t^*) \|}_{\text{can be made as small as wanted by making } |t-t^*| < \delta, \text{ because all } x_j(\cdot) \text{ are continuous fuctions}} + \underbrace{\| x_j(t^*) - x(t^*) \|}_{\downarrow\, j\to\infty,\ 0}$$

Hence $x(\cdot)$ is a continuous function.

Now, because of (xxx) we can write

$$\left\| \int_{t_0}^{t} f(s, x_{j-1}(s))\,ds - \int_{t_0}^{t} f(s, x(s))\,ds \right\| \le \left| \int_{t_0}^{t} \| f(s, x_{j-1}(s)) - f(s, x(s)) \|\,ds \right|$$

f locally Lipschitz ↘

$$\le L \left| \int_{t_0}^{t} \| x_{j-1}(s) - x(s) \|\,ds \right|$$

using (xxx) for j large enough ↘

$$\le \alpha L \varepsilon$$

thus we can write

$$x_j(t) = x_0 + \int_{t_0}^{t} f(s, x_{j-1}(s))\,ds$$

$j \to \infty$ ↓ , $j \to +\infty$ ↓

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

(**) [page 3.2]

But this means that $x(t)$ is a continuous solution of the integral equation (**) and hence, by Proposition 1, it is a solution of the initial value problem

(*) [page 3.2]

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

Actually we can easily prove that this solution is unique:

Suppose there are two solutions $x(\cdot)$ and $y(\cdot)$ to the problem (**) [and hence to (*)]. then

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

$$\leq L \left| \int_{t_0}^{t} \|x(s)-y(s)\|\,ds \right|$$

$$\leq \alpha L \sup_{s\in I_\alpha} \|x(s)-y(s)\|$$

Since $x(\cdot)$ and $y(\cdot)$ are continuous functions, Weierstrass theorem implies there exists $\tau \in I_\alpha$ such that $\|x(\tau)-y(\tau)\| = \max_{s\in I_\alpha} \|x(s)-y(s)\|$. Thus the above inequality for $t=\tau$ becomes

$$\|x(\tau)-y(\tau)\| \leq \alpha L \|x(\tau)-y(\tau)\|$$

But $\alpha L < 1$ (by the proof above) and thus, if $\|x(\tau)-y(\tau)\| \neq 0$, we have, by the above inequality,

$$1 \leq \alpha L$$

which contradicts $\alpha L < 1$.

Thus we must have $\|x(\tau)-y(\tau)\| = 0$ which implies that $\|x(t)-y(t)\| = 0 \quad \forall t \in I_\alpha$ (remember how τ was chosen!) which means the solution is unique.

We have prove the following important result:

Theorem (Picard-Lindelöf)

Let $D \subset \mathbb{R} \times \mathbb{R}^n$ be an open set, $(t_0, x_0) \in D$.
Let $f : D \to \mathbb{R}^n$ be a continuous function that is locally Lipschitz relative to the second variable of f

Then, the initial value problem

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

has a unique solution defined in an interval $I_\alpha = [t_0 - \alpha, t_0 + \alpha]$, for some sufficiently small $\alpha > 0$.

It is an important problem (both for the theory and for its applications) to know when can the domain of definition I_α given in the Picard-Lindelöf theorem be enlarged.

One result relevant to clarify this problem is the following:

Proposition 2

Let $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ be a <u>bounded</u> and <u>globally Lipschitz</u> function (i.e., there is a constant $L > 0$ valid for all $\mathbb{R}$)

then I_α can be extended to all $\mathbb{R}$ (the solution is global).

the following result (which will not be proved here) is a central result clarifying what can happen for a solutions not to be extended further

Theorem

With the assumptions of the Picard-Lindelöf theorem let $x(\cdot)$ be the solution of the Cauchy problem (*), defined in an interval J.
then, there exists an extension of $x(\cdot)$ to a maximal interval of existence $(a,b) \supseteq J$ and we have that, when $t \to b^-$ or $t \to a^+$, either

$$(t, x(t)) \longrightarrow \partial D \quad \text{or} \quad \|(t, x(t))\| \longrightarrow +\infty.$$

the following comparison theorem is useful in many instances:

Proposition 3

Let $D \subset \mathbb{R} \times \mathbb{R}^n$, $S \subset \mathbb{R} \times \mathbb{R}$ two open sets.
let $f : D \to \mathbb{R}^n$, $g : S \to \mathbb{R}$, satisfying Picard-Lindelöf conditions, be such that $\|f(t,x)\| \leq g(t, \|x\|)$.
Let $x(\cdot)$ be the solution of $x' = f(t,x)$ with $x(t_0) = x_0$, and $u(\cdot)$ be the solution of $u' = g(t,u)$ with $u(t_0) = u_0$
Let $\|x_0\| \leq u_0$.
then $\|x(t)\| \leq u(t)$ for all t for which both $x(\cdot)$ and $u(\cdot)$ are defined.

To complete the set of results of the basic theory of ODEs that will be needed later we will briefly refer to dependence of solutions on initial data and parameters:

Let $f(t,x;\gamma)$ depend on a parameter $\gamma \in \mathbb{R}^p$, then the solutions to initial value problems will depend also on γ and will be denoted by $x(t;t_0,x_0,\gamma)$ when we need to make clear its dependence on the initial condition (t_0,x_0) and on the parameter γ.

One can prove that under Picard-Lindelöf conditions (f continuous and Locally Lipschitz relative to x) then $x(\cdot)$ depend continuously of t_0 and locally Lipschitz of x_0. One can also prove that if f depends on γ in a C^k fashion, then the same will happens with $x(\cdot)$, and the same regularity results holds if f depends C^k of t and x, the regularity is inherited by $x(t;t_0,x_0,\gamma)$.

GMIT

Differential Equations and Applications

Session 4: Linear systems. Matrix exponential and fundamental solutions via eigenvalues and eigenvectors of the matrix system.

We now turn our attention to system of linear ODEs with constant coeficients. Remembering the case of radioactive decay the system of four ODEs can be written in vector form as

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}' = \begin{bmatrix} -k_x & 0 & 0 & 0 \\ k_x & -k_y & 0 & 0 \\ 0 & k_y & -k_z & 0 \\ 0 & 0 & k_z & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

or, in a more compact notation,

$$\underline{x}' = A\,\underline{x}$$

where $x = x(t) : \mathbb{R} \to \mathbb{R}^4$ and $A \in M_{4\times 4}(\mathbb{R})$.

The study of these systems can be done based on results from linear Algebra. First we introduce the notion of <u>matrix norm</u>:

Let $M_{n\times n}(\mathbb{R})$ be the vector space of all $n\times n$ matrices with real entries. A function $|||\cdot||| : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ is a matrix norm if, $\forall A, B \in M_{n\times n}(\mathbb{R})$, $\forall \lambda \in \mathbb{R}$,

(i) $|||A||| \geq 0$ and $|||A||| = 0$ iff $A = 0$

(ii) $|||cA||| = |c| \cdot |||A|||$

(iii) $|||A+B||| \leq |||A||| + |||B|||$

(iv) $|||AB||| \leq |||A||| \cdot |||B|||$

Some examples of matrix norms are, for $A = [a_{ij}]$,

a) $\|\!|A|\!\|_{\ell_1} = \sum_{i,j=1}^{n} |a_{ij}|$

b) $\|\!|A|\!\|_{\ell_2} = \left(\sum_{i,j=1}^{n} a_{ij}^2 \right)^{1/2}$

c) $\|\!|A|\!\|_{\ell_\infty} = n \max_{1 \le i,j \le n} |a_{ij}|$

d) $\|\!|A|\!\| = \max \{ \|Ax\| : \|x\| = 1 \}$, where $\|\cdot\|$ is a vector norm in $\mathbb{R}^n$.

Being $M_{n\times n}(\mathbb{R})$ a finite dimensional vector space (with dimension n^2) we know that all matrix norms are equivalent in the sense that a sequence of matrices being convergent does not depend on the matrix norm being used.

With the notion of matrix norm we can conclude that

$$\|Ax\| \le \|\!|A|\!\| \cdot \|x\|$$

(for instance using the "induced norm" in d) above).

Thus, taking $f(t,x) := Ax$, we have

$$\begin{aligned} \|f(t,x) - f(t,y)\| &= \|Ax - Ay\| \\ &= \|A(x-y)\| \\ &\le \|\!|A|\!\| \cdot \|x-y\| =: L\,\|x-y\| \end{aligned}$$

and thus we have that $x \mapsto Ax$ is locally Lipschitz (actually it is globally Lipschitz) and we can apply Picard-Lindelöf theorem to conclude that initial value problems $x' = Ax$, $x(t_0) = x_0$, have a unique local solution that, because $x \mapsto Ax$ is globally bounded and globally Lipschitz, it can be extended to all $t \in \mathbb{R}$.

In dimension $n=1$ the solution to

$$\begin{cases} x' = ax \\ x(0) = x_0 \end{cases}$$

is $x(t) = x_0 e^{at}$.

Could a similar expression be valid also when a is a $n \times n$ matrix? How can we define the exponential of a square matrix?

The idea is to do the same as with real (or complex) numbers:

Definition

Let $B \in M_{n\times n}(\mathbb{R})$.

We define $e^B := \sum_{j=0}^{\infty} \frac{1}{j!} B^j$, where $B^0 = I_n$.

the series is defined by $\sum_{j=0}^{\infty} \frac{1}{j!} B^j = \lim_{n\to\infty} \sum_{j=0}^{n} \frac{1}{j!} B^j$, and the sum (with $n+1$ terms) in the right-hand side is well defined because the product of matrices is associative. the radius of convergence of that power series (of matrices) is

$$r := \left(\overline{\lim_{k\to\infty}} \left(\frac{1}{k!}\right)^{1/k} \right)^{-1} = +\infty$$

and thus the exponential is convergent for all $B \in \mathcal{M}_{n\times n}(\mathbb{R})$.

the following result is relevant for the use of the matrix exponential in differential equations:

Proposition 4

Let $A, B \in \mathcal{M}_{n\times n}(\mathbb{R})$ be such that $AB = BA$.

Then: a) $e^{A+B} = e^A e^B$

b) $e^A e^B = e^B e^A$

c) $Be^A = e^A B$.

To illustrate where the commutativity $AB = BA$ enters observe that

$$e^{A+B} = I + A + B + \frac{1}{2!}(A+B)^2 + \cdots$$

$$= I + A + B + \frac{1}{2!}(A^2 + AB + BA + B^2) + \cdots$$

$$= I + A + B + \frac{1}{2!}A^2 + \frac{1}{2!}B^2 + AB + \cdots \quad \swarrow\ AB = BA$$

and $e^A e^B = \left(I + A + \frac{1}{2!}A^2 + \cdots\right)\left(I + B + \frac{1}{2!}B^2 + \cdots\right)$

$$= I + A + \frac{1}{2!}A^2 + \cdots + B + AB + \cdots + \frac{1}{2!}B^2 + \cdots$$

$$= I + A + \frac{1}{2!}A^2 + \cdots + B + \frac{1}{2!}B^2 + \cdots + AB + \cdots$$

the rigorous proof requires the use of induction to obtain

$$(A+B)^m = \sum_{k=0}^{m} \binom{m}{k} A^k B^{m-k}$$

and can be left as an exercise.

For application to ODE the interesting exponential is e^{At}, with $A \in M_{n\times n}(\mathbb{R})$ and $t \in \mathbb{R}$. Clearly $(At)(As) = (As)(At)$ because the real numbers t, s always commute.

Thus we have always

$$e^{A(t+s)} = e^{At+As} = e^{At}e^{As}, \quad e^{A0} = I_n$$

and also

$$\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A.$$

this proves the following result

Theorem

Let $A \in M_{n\times n}(\mathbb{R})$.
The matrix $\Phi(t) := e^{At}$ is a matrix solution of the ODE $x' = Ax$ with linearly independent

Corollary

- the general solution of the ODE system $x' = Ax$ can be written as $x(t) = e^{At}\alpha$ for arbitrary $\alpha \in \mathbb{R}^n$.
- the solution of the Cauchy problem

$$\begin{cases} x' = Ax \\ x(0) = x_0 \end{cases}$$

is $x(t) = e^{At}x_0$

- the solution of the Cauchy problem

$$\begin{cases} x' = Ax \\ x(t_0) = x_0 \end{cases}$$

is $x(t) = e^{A(t-t_0)}x_0$

let's see some examples

Example 1

Let $\begin{cases} x_1' = x_1 \\ x_2' = -2x_2 \end{cases}$

this is not really a system but two independent ODEs put together. We can solve them independently:

$$x_1' = x_1 \iff x_1(t) = \alpha e^t$$

$$x_2' = -2x_2 \iff x_1(t) = \beta e^{-2t}$$

and thus, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) = \begin{pmatrix} \alpha e^t \\ \beta e^{-2t} \end{pmatrix} = \begin{pmatrix} e^t & 0 \\ 0 & e^{-2t} \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

Let's try to reproduce this result using the matrix exponential

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$$

$$e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} t + \frac{1}{2!}\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}^2 t^2 + \frac{1}{3!}\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}^3 t^3 + \cdots$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} t + \frac{t^2}{2!}\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} + \frac{t^3}{3!}\begin{bmatrix} 1 & 0 \\ 0 & -8 \end{bmatrix} + \cdots$$

$$= \begin{bmatrix} 1 + t + \frac{t^2}{2!} + \cdots & 0 \\ 0 & 1 - 2t + \frac{(2t)^2}{2!} - \frac{(2t)^3}{3!} + \cdots \end{bmatrix}$$

$$= \begin{bmatrix} e^t & 0 \\ 0 & e^{-2t} \end{bmatrix}$$

and thus, for all $\alpha, \beta \in \mathbb{R}$,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) = \begin{bmatrix} e^t & 0 \\ 0 & e^{-2t} \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

is the general solution, as obtained before

Example 2

Consider the harmonic pendulum equation $x'' + x = 0$

We can solve it easily by a change of variables transforming it to a system of 1st order equations $y := x'$

$$y = x', \quad y' = x'' = -x$$

$$(*)\quad \begin{cases} x' = y \\ y' = -x \end{cases} \iff \begin{pmatrix} x \\ y \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

We can solve this system in many different ways:
Let's start by the exponential of the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

$$e^{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}t} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}t + \frac{1}{2!}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 t^2 + \frac{1}{3!}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^3 t^3 + \cdots$$

now observe that

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I_2$$

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^2 = -\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} I_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^4 = (-I_2)^2 = I_2$$

and from now on everything is repeated because we have obtained the identity matrix: $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^5 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} I_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, etc.

Thus

$$e^{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}t} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}t - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{t^2}{2!} - \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\frac{t^3}{3!} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{t^4}{4!} + \cdots$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} \cdots\right) + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\left(t - \frac{t^3}{3!} + \cdots\right)$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\cos t + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\sin t$$

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

and the general solution of (*) is

$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad \forall \alpha, \beta \in \mathbb{R}.$$

The general solution of $x'' + x = 0$ is obtained by picking just the result in the first line:

$$x(t) = \alpha \cos t + \beta \sin t, \quad \forall \alpha, \beta \in \mathbb{R}.$$

Example 3

Consider now the ODE system

$$(**) \qquad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}' = \underbrace{\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}}_{A}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

this system, like the one in Example 1, can be solved without using the matrix exponential: the equation for x_3 can be solved independently to get $x_3(t) = e^{\lambda t} r$, then the equation for x_2 becomes $x_2' = \lambda x_2 + r e^{\lambda t}$ which can be solved by an integrating factor, resulting in an expression for $x_2(t)$ that then can be substituted into the first equation $x_1' = \lambda x_1 + x_2(t)$.

Instead, let us solve (**) by computing the exponential of the matrix of the system (multiplied by t).

First observe that $A = \lambda I_3 + N_3$, where $N_3 = \begin{bmatrix} & 1 & \\ & & 1 \\ & 0 & \end{bmatrix}$.

Clearly $I_3 N_3 = N_3 I_3$, because the identity matrix commutes with all matrices. Then, by Proposition 4, we have

$$e^{At} = e^{\lambda I_3 t + N_3 t} = e^{\lambda t I_3} \cdot e^{N_3 t} = e^{\lambda t} I_3 e^{N_3 t} = e^{\lambda t} e^{N_3 t}$$

Now observe that

$$N_3 = \begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix}, \quad N_3^2 = \begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix}$$

$$N_3^3 = \begin{bmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix} = 0$$

and hence

$$e^{N_3 t} = I_3 + N_3 t + \frac{1}{2!} N_3^2 t^2$$

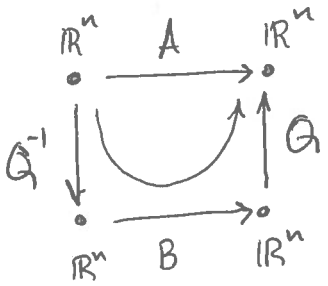
$$= \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{bmatrix} t + \begin{bmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix} \frac{t^2}{2!}$$

$$= \begin{bmatrix} 1 & t & t^2/2 \\ & 1 & t \\ & & 1 \end{bmatrix}.$$

Consequently, the general solution of (**) is, with $\alpha_i \in \mathbb{R}$,

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix}(t) = e^{\lambda t} \begin{pmatrix} 1 & t & t^2/2! \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

In linear Algebra it is very useful to change basis so that a linear transformation which is originally represented by a matrix A becomes represented by a different matrix B which is simpler than A



$$A = QBQ^{-1} \qquad (\dagger)$$

What is the relation between e^A and e^B ?

$$e^A = I_n + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

$$= I_n + QBQ^{-1} + \frac{1}{2!}QBQ^{-1}QBQ^{-1} + \frac{1}{3!}QBQ^{-1}QBQ^{-1}QBQ^{-1} + \cdots$$

($Q^{-1}Q = I_n$, $Q^{-1}Q = I_n$)

$$= Q\left(I_n + B + \frac{1}{2!}B^2 + \frac{1}{3!}B^3 + \cdots\right)Q^{-1}$$

$$= Qe^BQ^{-1}$$

Observe that if B is a diagonal matrix, then e^B is very easy to compute (we did it in Example 1):

$$B = \operatorname{diag}(\lambda_1, \cdots, \lambda_n) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \Rightarrow e^B = \operatorname{diag}(e^{\lambda_1}, \cdots, e^{\lambda_n}) = \begin{bmatrix} e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n} \end{bmatrix}$$

We know that if a matrix A has n linearly independent eigenvectors, then it is diagonalizable, i.e., there exists a diagonal matrix B and an invertible matrix Q such that (†) holds; furthermore Q is the matrix whose columns are the eigenvectors and the elements of the main diagonal of B are the eigenvectors (in the same order of the eigenvectors).

To recall this result let us consider the simple case of a 2×2 matrix, A.

Let (λ, w), (μ, v) be two eigenpairs (pairs of eigenvalue and associated eigenvector) with w and v linear independent then $A\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ and $A\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mu \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

We can write these two equalities in a condensed form as

$$A \left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2 \end{array}\right) = \left(\begin{array}{c|c} \lambda w_1 & \mu v_1 \\ \lambda w_2 & \mu v_2 \end{array}\right) = \left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2 \end{array}\right) \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$$

and thus

$$A = \left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2 \end{array}\right) \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2 \end{array}\right)^{-1}$$

thus, in this case, the solutions of the linear ODE

$$\underline{x}' = A\underline{x}$$

can be written as

$$\underline{x}(t) = e^{At}\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \underbrace{\left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2\end{array}\right)\begin{pmatrix} e^{\lambda t} & 0 \\ 0 & e^{\mu t}\end{pmatrix}}_{=:\underline{\Phi}(t)}\underbrace{\left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2\end{array}\right)^{-1}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}}_{=\begin{pmatrix}a\\ b\end{pmatrix}\in\mathbb{R}^2}$$

$$= \left(\begin{array}{c|c} w_1 & v_1 \\ w_2 & v_2\end{array}\right)\begin{pmatrix} a e^{\lambda t} \\ b e^{\mu t}\end{pmatrix}$$

$$= a e^{\lambda t}\begin{pmatrix} w_1 \\ w_2\end{pmatrix} + b e^{\mu t}\begin{pmatrix} v_1 \\ v_2\end{pmatrix}.$$

this is readily generalized for diagonalizable matrices of any (finite) dimension: solutions of $\underline{x}' = A\underline{x}$, with $A \in M_{n\times n}(\mathbb{R})$ diagonalizable, with (λ_i, v_i), $i=1,\dots,n$, eigenpairs with v_i and v_j linear independent if $i\neq j$, can be written as

$$\underline{x}(t) = \underline{\Phi}(t)\,\underline{\alpha}$$

where $\underline{\Phi}(t) := \left(v_1 \middle| \cdots \middle| v_n\right)\begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t}\end{pmatrix}$

is a fundamental solution matrix and $\alpha \in \mathbb{R}^n$. [illegible]

and, furthermore, this can be written as

$$\underline{x}(t) = \alpha_1 e^{\lambda_1 t} v_1 + \alpha_2 e^{\lambda_2 t} v_2 + \cdots + \alpha_n e^{\lambda_n t} v_n .$$

this way to write the solutions will be extremely helpful for the geometric interpretation of the result, which is our first inroad into the so called "qualitative theory", that will be studied next.

GMIT

Differential Equations and Applications

Session 5: Notions of qualitative theory. Applications to linear systems. Phase portraits of linear systems. the case of complex eigenvalues.

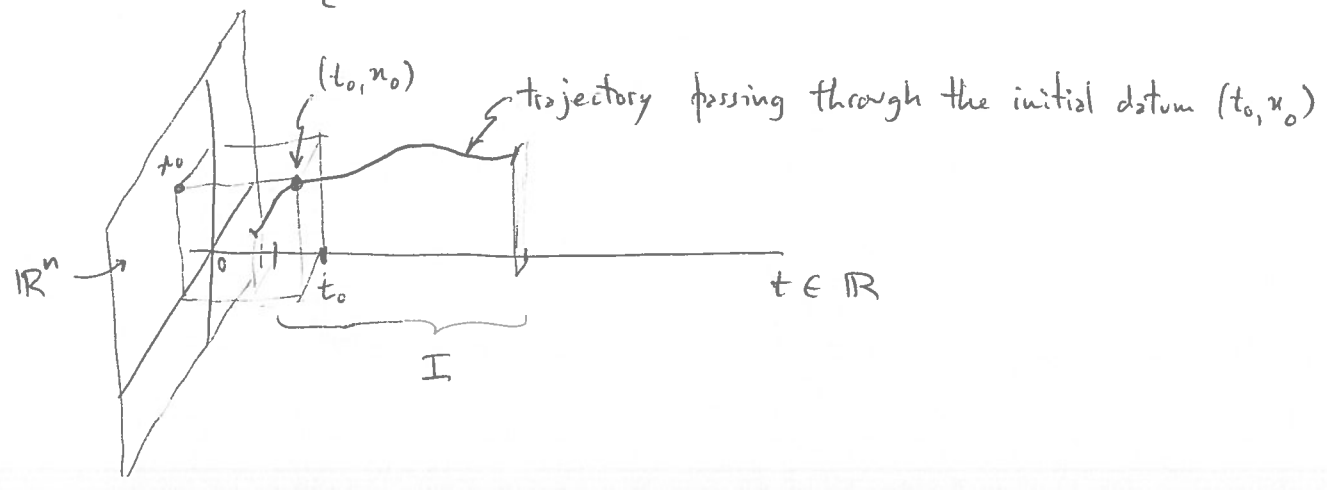
We saw that a system $\underline{x}' = A\underline{x}$ for which $A \in M_{n\times n}(\mathbb{R})$ is diagonalizable has solutions $\underline{x}(t)$ that can be written as

$$x(t) = \alpha_1 e^{\lambda_1 t} v_1 + \cdots + \alpha_n e^{\lambda_n t} v_n \qquad (*)$$

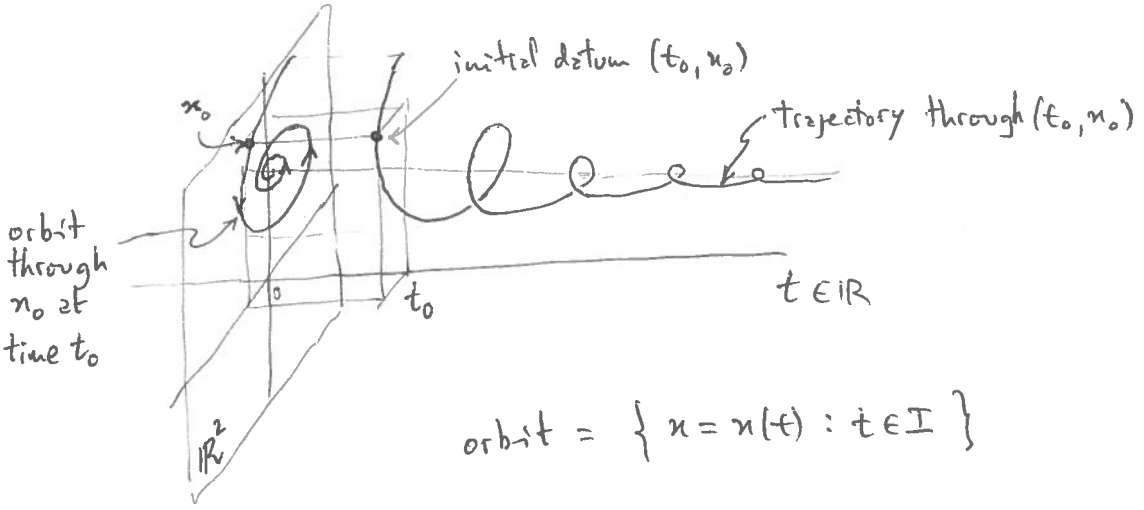
where (λ_i, v_i) are the eigenpairs of A, with v_i, v_j linearly independent for $i \neq j$. We now introduce concepts that will allow us to have a geometric interpretation of the solutions $(*)$ which is very useful in the applications and also constitute an extremely rich theoretical field.

First note that a solution of a differential equation $\underline{x}' = f(t, \underline{x})$ is a function $x(t)$ and so its graph is in the (t, x) space. In differential equations literature this graph is called the <u>trajectory</u>:

$$\text{trajectory} = \left\{ (t, x) : t \in I,\ x = x(t) \right\}$$



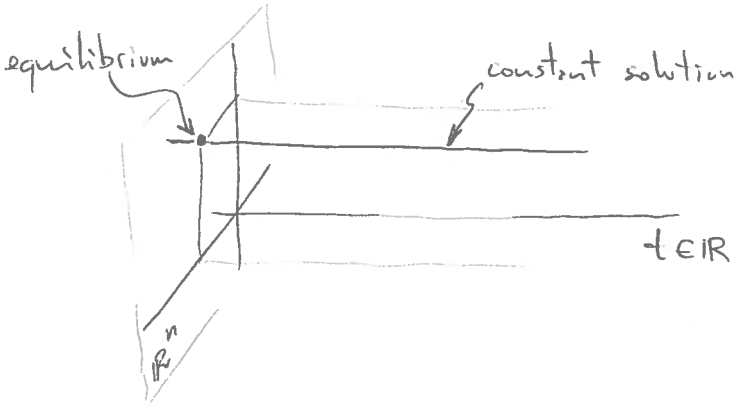
the space $\mathbb{R}^n$, where $x(t)$ "lives", is called the <u>phase space</u> and the projection of the trajectory on the phase space is called an <u>orbit</u> of the equation $\underline{x}' = f(t, \underline{x})$



$$\text{orbit} = \{ x = x(t) : t \in I \}$$

(the arrows in the orbit point to increasing t).

Clearly, a constant solution has a trajectory which is a line parallel to the t-axis and thus its projection in the phase plane is a single point, which is called an <u>equilibrium</u>



the collection of orbits in a sufficient number to suggest the behaviour of those orbits not represented is called the phase portrait of the system. For linear systems of the type $\underline{x}' = A\underline{x}$, with $A \in M_{n\times n}(\mathbb{R})$ diagonalizable and with all eigenvalues real valued, the phase portrait is easy to draw. We illustrate this with an example.

Example 1

Consider the system of ODEs

$$(x) \qquad \begin{cases} x' = 2x \\ y' = 4x - 2y. \end{cases}$$

Writing this system in vector form we have

$$(xx) \qquad \begin{pmatrix} x \\ y \end{pmatrix}' = \underbrace{\begin{pmatrix} 2 & 0 \\ 4 & -2 \end{pmatrix}}_{A} \begin{pmatrix} x \\ y \end{pmatrix}$$

the only equilibrium is $(x,y) = (0,0)$.
the eigenvalues and eigenvectors of A are:

$$0 = \det(A - \lambda I_2) = \det \begin{pmatrix} 2-\lambda & 0 \\ 4 & -2-\lambda \end{pmatrix} = (\lambda-2)(\lambda+2) - 0$$

$$\lambda_+ = 2, \quad \lambda_- = -2$$

$$\downarrow$$

$$\underline{0} = \begin{pmatrix} 0 & 0 \\ 4 & -4 \end{pmatrix} \begin{pmatrix} v_1^+ \\ v_2^+ \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ v_1^+ - v_2^+ \end{pmatrix} \Leftrightarrow v_2^+ = v_1^+ \quad : \quad v^+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$\lambda_- = -2$

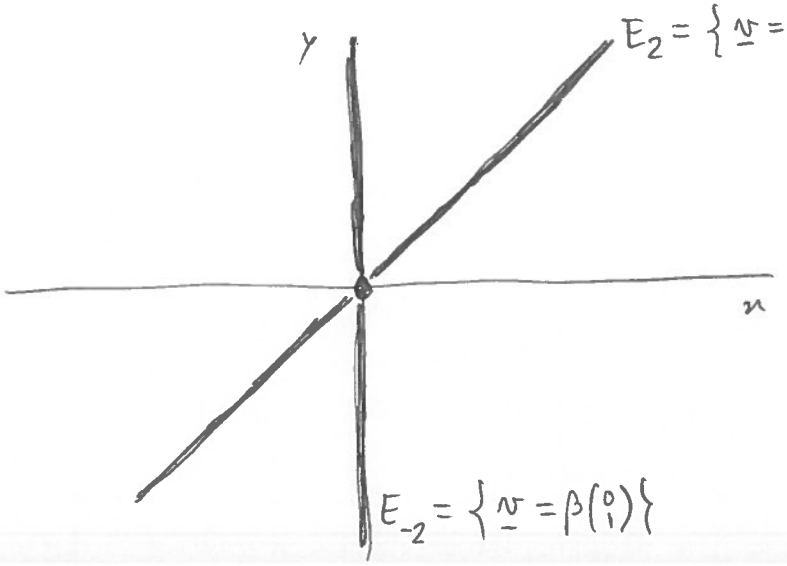
$\downarrow$

$$\underline{0} = \begin{pmatrix} 4 & 0 \\ 4 & 0 \end{pmatrix}\begin{pmatrix} v_1^- \\ v_2^- \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4v_1^- \\ 4v_1^- \end{pmatrix} \Leftrightarrow v_1^- = 0 \quad : \quad v^- = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

thus, we can write the general solution of (xx) [=(x)] as

$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = \alpha_1 e^{2t}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 e^{-2t}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \text{for } \alpha_1, \alpha_2 \in \mathbb{R}. \tag{xxx}$$

the orbits of (xx) are now easily drawn by plotting the equilibria first, the eigenspaces next (which are invariant sets for the equation, i.e., if the initial condition is in an eigenspace (= a space generated by an eigenvector) it remains there for all t, because of the structure of (xxx), as explained next:



Observe that a solution for which the initial condition at $t=0$ (say) is colinear with the eigenvector $\begin{pmatrix}1\\1\end{pmatrix}$ is a solution such that, at $t=0$,

$$\begin{pmatrix}x\\y\end{pmatrix}(0) = \alpha\begin{pmatrix}1\\1\end{pmatrix}$$

thus, by (xxx), we have

$$\alpha\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}x\\y\end{pmatrix}(0) = \alpha_1 e^{2\cdot 0}\begin{pmatrix}1\\1\end{pmatrix} + \alpha_2 e^{-2\cdot 0}\begin{pmatrix}0\\1\end{pmatrix}$$

meaning: $\alpha_1 = \alpha$ and $\alpha_2 = 0$. thus the solution is, for all other t,

$$\begin{pmatrix}x\\y\end{pmatrix}(t) = \alpha e^{2t}\begin{pmatrix}1\\1\end{pmatrix} \qquad (\dagger)$$

which remain colinear with $\begin{pmatrix}1\\1\end{pmatrix}$: the eigenspace associated with $\lambda_+ = 2$ is invariant. the same holds for the eigenspace associated with $\lambda_- = -2$.

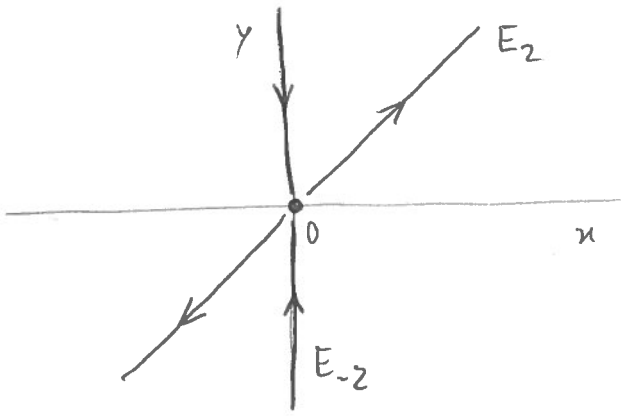
Note that, from $(\dagger)$, we have that

$$\begin{pmatrix}x\\y\end{pmatrix}(t) \longrightarrow 0 \quad \text{as} \quad t \to -\infty$$

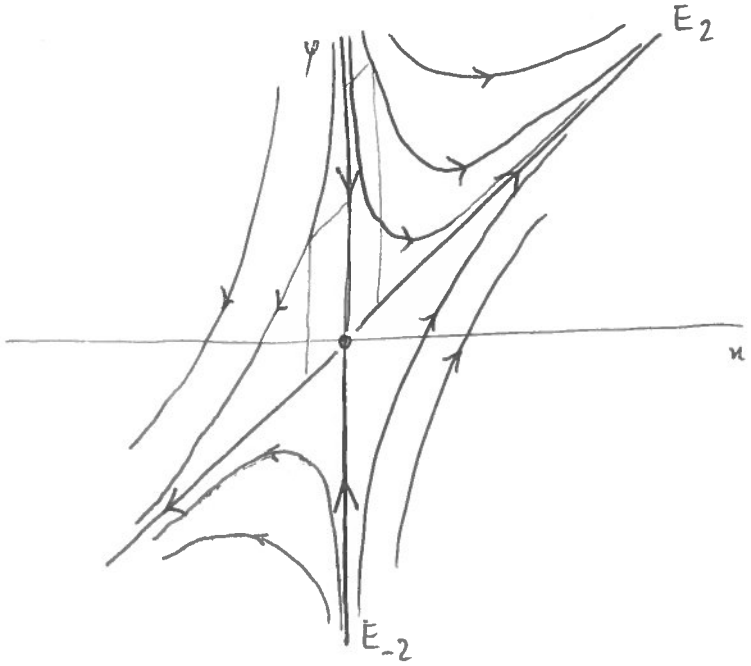
and $\left\| \begin{pmatrix}x\\y\end{pmatrix}(t) \right\| \longrightarrow +\infty \quad \text{as} \quad t \to +\infty.$

thus the arrows indicating the increasing time variable are as follows for the orbits on E_2 (with an analogous

— but opposite — behaviour for E_{-2})



Now the decomposition of a general solution into components in E_2 and in E_{-2} given by (xxx) allow us to get the following phase portrait



Example 2

Consider the system $\begin{pmatrix} x \\ y \end{pmatrix}' = A \begin{pmatrix} x \\ y \end{pmatrix}$ where the 2×2 real matrix A has eigenpairs $\left(-1, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)$, $\left(-2, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right)$.

Let us draw the phase portrait of this system.

Starting by the equilibria: as the two eigenvalues of A are not zero, the matrix A is invertible and so

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} \iff A^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \iff \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

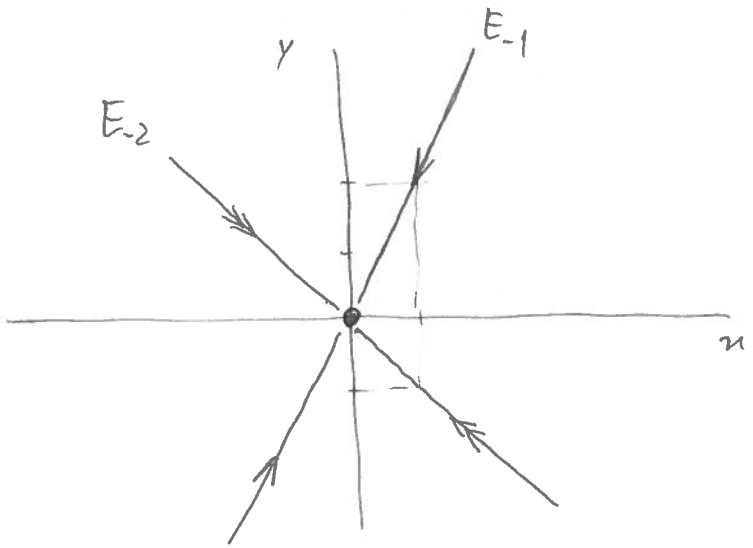
implying the only equilibrium solution is $(x,y) = (0,0)$.

We know that all solutions can be written as

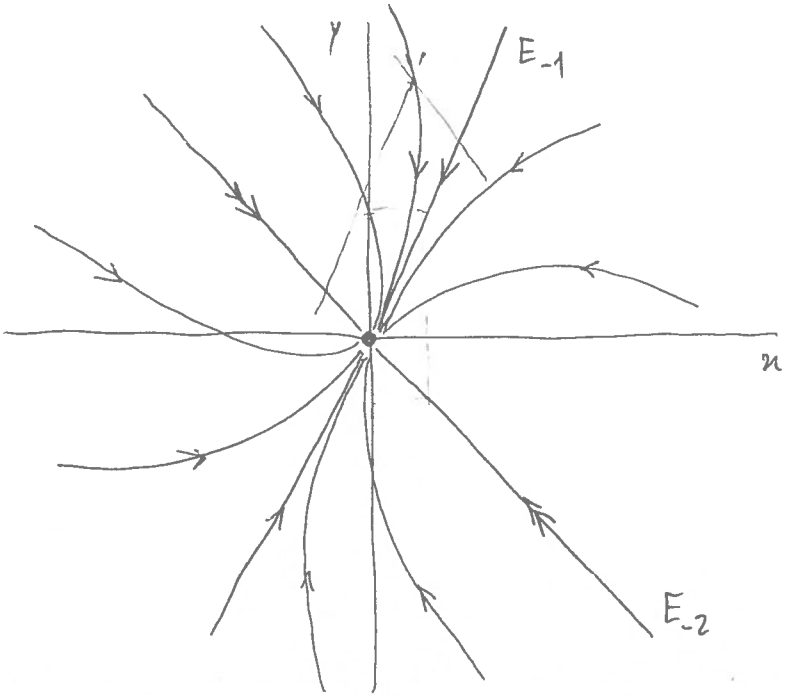
$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = \left(\begin{array}{c|c} 1 & 1 \\ 2 & -1 \end{array}\right) \begin{pmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$(\ddagger) \qquad = \alpha e^{-t} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta e^{-2t} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

As before, in Example 1, the eigenspaces are invariant; however, now solutions in either E_{-1} and E_{-2} will converge to the equilibrium point $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ as $t \to +\infty$. Thus, the situation is as shown next



For initial conditions not in the eigenspaces we do as before: decompose the initial point into components on the eigenspaces; each component evolve with time as any other point in the eigenspace and we reconstruct the orbit adding the components at later times using (*). Note that the component in E_{-2} converges faster to $\begin{pmatrix}0\\0\end{pmatrix}$ than that in E_{-1}. The end result of these observations is the following phase portrait:



We now turn to the problem of what happens when the real matrix A has complex eigenvalues (and thus also complex eigenvectors). Clearly, solutions to $\underline{x}' = A\underline{x}$ with $\underline{x}(t_0) = \underline{x}_0 \in \mathbb{R}^n$ are real valued vectors since

$$\underline{x}(t) = e^{At}\underline{x}_0$$

and e^{At} is a real matrix. However, if we write

$$\underline{x}(t) = e^{At}\underline{x}_0 = Q\begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix} Q^{-1}\underline{x}_0$$

$$= Q\begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{bmatrix} \hat{\underline{x}}_0$$

where, now, Q, $\operatorname{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t})$, and $Q^{-1}\underline{x}_0 =: \hat{\underline{x}}_0$, are complex, we cannot use the previous approach directly to draw the phase portrait in $\mathbb{R}^n$. For this we need to explore a little more carefully the relation between real and complex solutions to a system of ODEs with real matrices $A \in M_{n\times n}(\mathbb{R})$.

We first recall that if a real matrix A has a complex (and non-real) eigenvalue λ, then the complex conjugate $\bar{\lambda}$ is also an eigenvalue, and the eigenvectors are also conjugate:

$$Av = \lambda v \iff \overline{Av} = \overline{\lambda v} \iff \bar{A}\bar{v} = \bar{\lambda}\bar{v} \iff A\bar{v} = \bar{\lambda}\bar{v}$$

Also, recall from Linear Algebra that if $(\lambda, v), (\bar{\lambda}, \bar{v})$ are two eigenpairs of $A \in M_{n\times n}(\mathbb{R})$, then v_R and v_I, the real and imaginary parts of v, are linearly independent

Let $v_R = \operatorname{Re} v = \dfrac{v+\bar{v}}{2}$, $v_I = \operatorname{Im} v = \dfrac{v-\bar{v}}{2i}$

Suppose v_R and v_I are linearly dependent. then, there exists $a, b \in \mathbb{R}$ and different from zero such that

$$\begin{aligned} 0 &= a v_R + b v_I \\ &= a \frac{v+\bar{v}}{2} + b \frac{v-\bar{v}}{2i} \\ &= \frac{1}{2}(a-ib)\, v + \frac{1}{2}(a+ib)\, \bar{v} \end{aligned} \qquad (\star)$$

but, since $a, b \neq 0$, we have that the complex numbers $\frac{1}{2}(a-ib)$ and $\frac{1}{2}(a+ib)$ are also not zero, and $(\star)$ means that v and $\bar{v}$ are linearly dependent, which is a contradiction because a general Linear Algebra result tells us that eigenvectors associated with different eigenvalues are always linearly independent.

Now let's get back to the ODE $\underline{x}' = A\underline{x}$ and relate complex and real solutions: Let $\underline{x}(t) = \underline{x}_R(t) + i\underline{x}_I(t)$ be a complex solution with $\underline{x}_R(t) = \operatorname{Re} \underline{x}(t)$, $\underline{x}_I(t) = \operatorname{Im} \underline{x}(t)$.

then $\underline{x}' = A\underline{x} \iff (x_R + ix_I)' = A(x_R + ix_I)$

$$\iff x_R' + ix_I' = Ax_R + iAx_I$$

and since $A \in M_{n\times n}(\mathbb{R})$ the vectors Ax_R and Ax_I are real (i.e., vectors in $\mathbb{R}^n$) and thus the above equality is equivalent to

$$x_R' = Ax_R \quad \text{and} \quad x_I' = Ax_I$$

So, if x is a complex solution of $x' = Ax$, then $x_R = \mathrm{Re}\, x$ and $x_I = \mathrm{Im}\, x$ are real valued solutions of the same equation.

All this allow us to extract information from representing a solution using eigenvalues and eigenvectors of A even when these are complex valued. We now will see how with a two dimensional example:

Let $\begin{pmatrix} x \\ y \end{pmatrix}' = A\begin{pmatrix} x \\ y \end{pmatrix}$ be a 2-dimensional ODE system with $A \in M_{2\times 2}(\mathbb{R})$ having a pair of complex eigenvalues λ and $\bar{\lambda}$. Let v_λ, $v_{\bar{\lambda}}$ be the corresponding associated eigenvectors.

From what we did before we know that all solutions (real or complex valued) can be written as

$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = c_1 e^{\lambda t} v_\lambda + c_2 e^{\bar{\lambda} t} v_{\bar{\lambda}}$$

$\lambda = a+ib$, $v_{\bar{\lambda}} = \bar{v}_\lambda$

$$= c_1 e^{at} e^{ibt} v_\lambda + c_2 e^{at} e^{-ibt} \bar{v}_\lambda$$

$$= e^{at}\left(c_1 e^{ibt} v_\lambda + c_2 \overline{e^{ibt}} \bar{v}_\lambda \right)$$

$$= e^{at}\left(c_1 e^{ibt} v_\lambda + c_2 \overline{\left(e^{ibt} v_\lambda\right)} \right)$$

and for this function to be real valued we need to choose constants such that $c_2 = \bar{c}_1$. In this case

$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = e^{at}\left(c_1 e^{ibt} v_\lambda + \overline{\left(c_1 e^{ibt} v_\lambda\right)} \right)$$

$$= 2e^{at}\, \mathrm{Re}\left(c_1 e^{ibt} v_\lambda\right)$$

$$= 2e^{at}\, \mathrm{Re}\left((c_R + ic_I)(\cos bt + i \sin bt)(v_R + i v_I)\right)$$

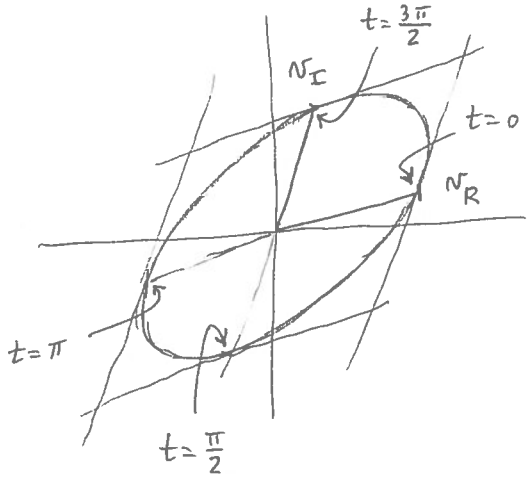
$$(\sharp) \qquad = 2e^{at}\left[c_R\left(v_R \cos bt - v_I \sin bt\right) - c_I\left(v_R \sin bt + v_I \cos bt\right)\right]$$

Since we saw that v_R and v_I are linearly independent vectors, the functions

$e^{at}\left(v_R \cos bt - v_I \sin bt\right)$ and $e^{at}\left(v_R \sin bt + v_I \cos bt\right)$

are linearly independent functions which (because the ODE system is linear) are solutions of $x' = Ax$.

Note that the terms in $[\cdots]$ in (♯) trace an ellipse as t evolves in $\mathbb{R}$: consider, to simplify, $c_R = 1$, $c_I = 0$, then, because v_R and v_I are linearly independent



For others c_R and c_I the result would be similar, with the ellipse bigger or smaller, depending on being $c_R^2 + c_I^2$ bigger or smaller than 1, respectively, and with the points along the ellipse attained at different values of t (i.e., the "movement" done with a different phase but with equal frequency b).

this, together with information comming from e^{at}, allow us to draw the phase portrait in these cases, as illustrated in the next example:

Example 3

Consider the damped pendulum described by the differential equation

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

Defining $y := x'$ we obtain the 1st order system

$$\begin{cases} x' = y \\ y' = -2x - 2y \end{cases}$$

which can be written as

$$\begin{pmatrix} x \\ y \end{pmatrix}' = \underbrace{\begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix}}_{A \in M_{2\times 2}(\mathbb{R})} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The eigenvalues of A are the solutions of

$$0 = \det(A - \lambda I_2) = \det \begin{pmatrix} -\lambda & 1 \\ -2 & -2-\lambda \end{pmatrix} = \lambda(\lambda+2) + 2$$

$$= \lambda^2 + 2\lambda + 2$$

which are $\lambda_{\pm} = -1 \pm i$

Computing the eigenvectors we obtain the eigenpairs

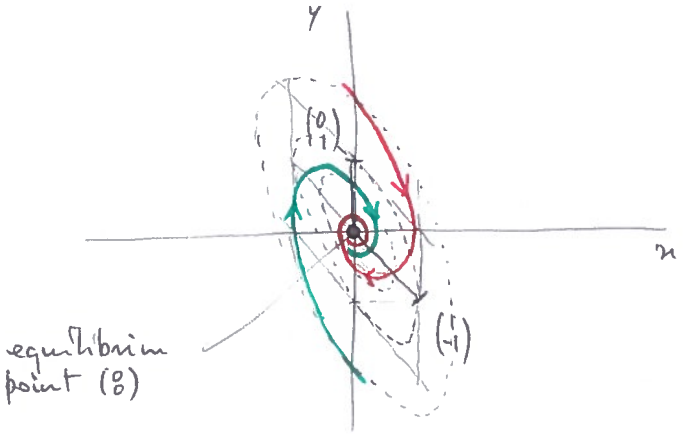
$$\left(-1+i, \begin{pmatrix} 1 \\ -1+i \end{pmatrix}\right), \quad \left(-1-i, \begin{pmatrix} 1 \\ -1-i \end{pmatrix}\right)$$

From what was done before, all real valued solutions

are given by

$$\begin{pmatrix} x \\ y \end{pmatrix}(t) = 2e^t \left[c_R \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} \cos t - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \sin t \right) - c_I \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} \sin t + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cos t \right) \right]$$

and the phase portrait is (two orbits and the equilibrium are drawn)



Note that from the first equation, $x' = y$, we know that $x' > 0$ when $y > 0$, which means that $x(t)$ is increasing when the orbit is in the half-plan $\{y > 0\}$ [and decreasing when in $\{y < 0\}$ because $x' < 0$ if $y < 0$.]

All that was done before relies on the possibility of diagonalizing the system matrix, i.e., when the matrix $A \in M_{n \times n}(\mathbb{R})$ has n linearly independent eigenvectors (i.e., the number of linearly independent eigenvalues is equal to the dimension of the phase space)

What happens when this is not the case?

When there is not a change of basis that allow us to diagonalize A we can still transform it in a simples "quasi diagonal" matrix called a Jordan canonical form. We have no time to deal in depth with this Linear Algebra topic so we will only describe what is possible to do without going into details on how to effectively implement it from a computational viewpoint as it is appreciably more difficult that the diagonalization process.

The Jordan Canonical form theorem states that every matrix can be transformed in a similar matrix of Jordan type, namely

$$A = R J R^{-1},$$

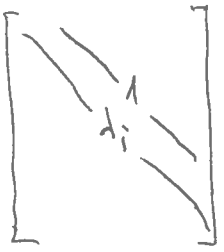
where J has the structure of a block diagonal matrix

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{bmatrix}$$

with the Jordan blocks are either

$J_i = [\lambda_i]$ (with dimension 1)

or

$J_i =$ 

(with dimension > 1)

and λ_i are the eigenvalues of the matrix A.
The matrix R has columns which are the <u>eigenvectors</u> and the <u>generalized eigenvectors</u> of A.

[An eigenvector of A is a $w \neq 0$ such that $(A - \lambda I)w = 0$.
A generalized eigenvector of A is a vector v such that $(A - \lambda I)v = w$, where w is either an eigenvector or another generalized eigenvector. this stops because, by the Cayley-Hamilton theorem, $(A - \lambda I)^n = 0$, and so the number of generalized eigenvectors is finite: actually its number is such that, together with the eigenvectors, they form a basis of $\mathbb{R}^n$: the basis for which A can be transformed into J.]

this result is almost as good as being able to diagonalize the matrix A because the exponential of a Jordan matrix J is very easy to compute:

$$e^{Jt} = e^{\operatorname{diag}(J_1,\dots,J_k)t} = \operatorname{diag}\left(e^{J_1 t},\dots,e^{J_k t}\right)$$

and $J_i = [\lambda_i] \Rightarrow e^{J_i t} = e^{\lambda_i t}$

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} = \lambda_i I + \underbrace{\begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}}_{\text{nilpotent matrix } N}$$

$$\Rightarrow e^{J_i t} = e^{\lambda_i I t + N t} = e^{\lambda_i t} I\, e^{Nt} = e^{\lambda_i t} e^{Nt}$$

and e^{Nt} is easy to compute (see Example 3 of session 4) because a nilpotent matrix is zero for some power m: in the case of the matrix N above m = dimension of N. So

$$e^{Nt} = I + Nt + \frac{1}{2!}N^2t^2 + \cdots + \frac{1}{(m-1)!}N^{m-1}t^{m-1}$$

$$= \begin{bmatrix} 1 & t & t^2/2! & \cdots & t^{m-1}/(m-1)! \\ 0 & 1 & t & \cdots & t^{m-2}/(m-2)! \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

GMIT

Differential Equations and Applications

Session 6 : Linearization about equilibria. Conservative systems.

What can we do if the system of ODE we need to study is not linear ?

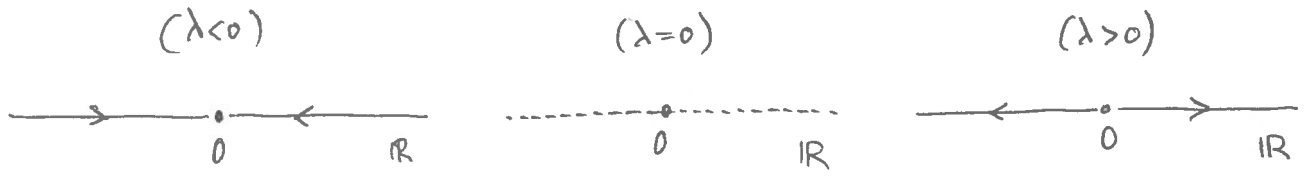
As the study of linear systems is (relatively) easy, it is natural to try to use what we just study about these systems in the nonlinear case.

Let us start to see what can happen in dimension 1.

In this case the linear system is the single equation

$$(*) \qquad x' = \lambda x$$

where $\lambda \in \mathbb{R}$ is a constant. Its phase portrait depends on the sign of λ:



(all points are equilibria)

Now consider the nonlinear differential equation

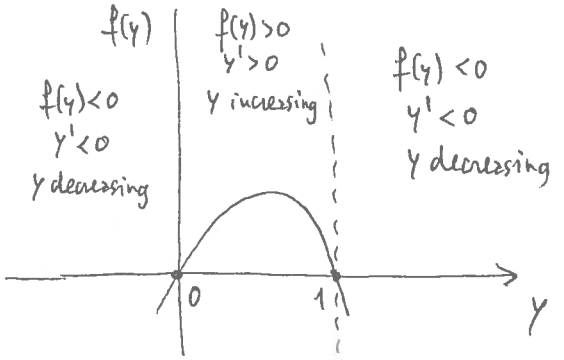
$$(**) \qquad y' = y - y^2$$

this equation has two equilibria; $0 = y - y^2 = y(1-y)$:

$$y_0 = 0 \quad \text{and} \quad y_1 = 1$$

the phase portrait is very easy to draw using the graph of $f(y) := y - y^2$ and observing that $f(y) > 0 \iff y' > 0 \iff$ $\iff y(t)$ is increasing, and correspondingly if $f(y) < 0$.

then:



$f(y) = y - y^2$

Now look at this in a different way:

Since f is differentiable we know that, for every point $a \in \mathbb{R}$, f is well approximated by its tangent straight line at $(a, f(a))$, which is a straight line passing through that point and with slope equal to $f'(a)$.

So, close to the equilibrium point $y_0 = 0$ we can approximate $f(y) = y - y^2$ by the straight line with equation

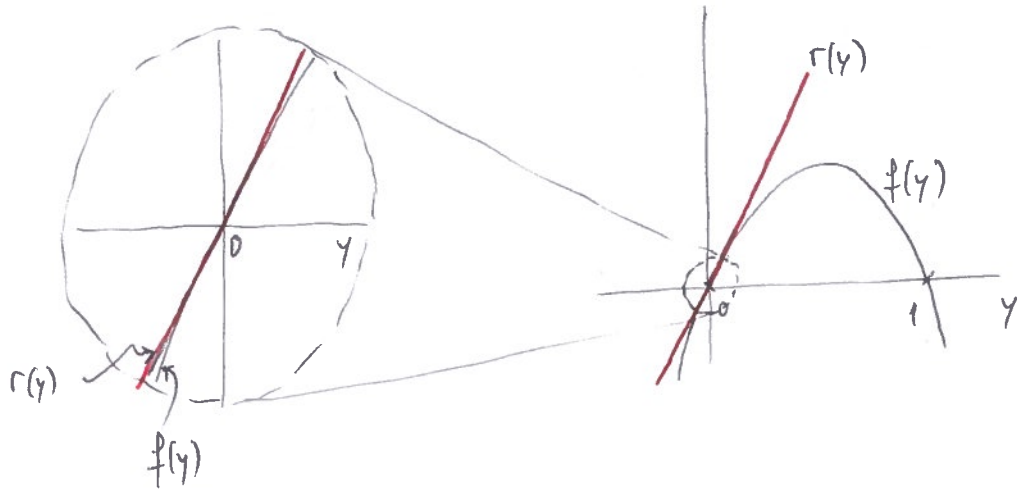
$$r(y) = f(y_0) + f'(y_0)(y - y_0)$$

$\downarrow y_0 = 0$

$$= f(0) + f'(0)(y - 0)$$

$\downarrow f(0) = 0 - 0^2 = 0$, $f'(0) = 1$

$$= y$$



So, close to the equilibrium point $y_0 = 0$ the differential equation $y' = y - y^2$ seems to be "well approximated" by the linear differential equation $y' = y$, which is (*) with $\lambda = 1$, and the phase portraits of both these equations indeed confirm the solutions behave qualitatively in the same way.

Now, close to the equilibrium $y_1 = 1$, the function $f(y) = y - y^2$ is well approximated by its tangent straight line at $(1, f(1)) = (1, 0)$, which has equation

$$s(y) = f(1) + f'(1)(y - 1)$$
$$= 0 - (y - 1)$$

$f(1) = 0$
$f'(1) = -1$

Now, $y - 1$ measures the displacement from the equilibrium point; call it $x \overset{\text{def}}{:=} y - 1$. Then $x' = (y-1)' = y'$ and the equation (**) about $y \approx 1$ (i.e., $x \approx 0$) can

be approximated by

$$x' = -x$$

about $x \approx 0$. We can check that the phase portraits are indeed qualitatively alike.

thus, in this example the behaviour of solutions close to the equilibria can be studied by studying the linear differential equations obtained by substituting the function on the right-hand side of the equation by its tangent straight line in those points.

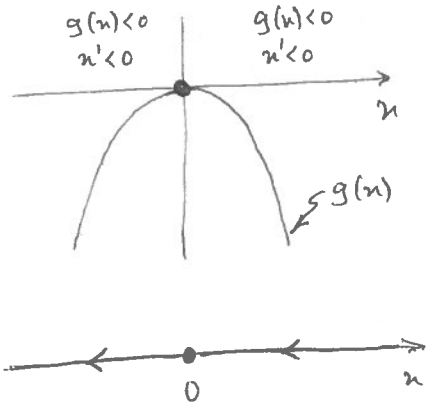
This is an instance of the <u>linearization procedure</u>: a useful and very common technique but one that it is by no means universally valid as the next example shows: consider the equation $x' = -x^2$; its (unique) equilibrium is $\hat{x} = 0$. the tangent line of $g(x) = -x^2$ at $\hat{x}$ has equation

$$u(x) = g(\hat{x}) + g'(\hat{x})(x - \hat{x})$$

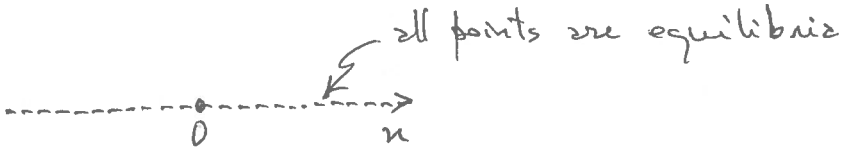
$$= 0, \quad \forall x \in \mathbb{R}$$

$g(0) = 0$
$g'(0) = 0$

the phase portrait of $x' = -x^2$ is also easy to plot:



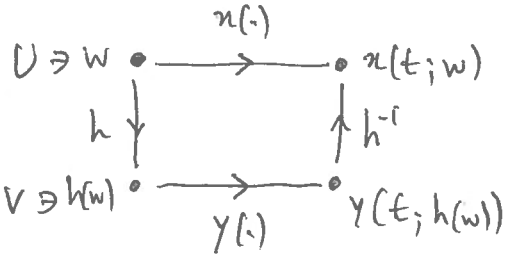
However, the phase portrait of the linearized equation about $\hat{x} = 0$ is



because the equation is $x' = 0$ (since the tangent line is $u(x) = 0 \; \forall x$). Clearly, the two phase portraits are not qualitatively analogous and we cannot use the linear equation to gain information about the behaviour of solutions to the nonlinear one.

We cannot go very deeply in this course about the relation about linear and non-linear systems and the linearization about equilibria, as it would require a more advanced course. We can, at least, state a general (hand-waving) idea. The rigorous result is the object of two results called the Hartman-Grobman and the Hadamard-Perron theorems

the result (which we will not state rigorously...) states that if $f: \mathbb{R}^n \to \mathbb{R}^n$ is a C^1 vector field and the jacobian matrix at the zero x_0 of f, $Df(x_0)$, does not have eigenvalues with zero real parts, then there exists a local homeomorphism h (a continuous function with a continuous inverse, defined in a neighborhood of x_0) such that solutions $x(t;w)$ of $x' = f(x)$, $x(0) = w$, and those $y(t;z)$ of the linear system $y' = Df(x_0)\cdot y$, $y(0) = z = h(w)$ are related by the conjugation defined by h, i.e.,



the result can be reinforced if $f \in C^k$, then it can be concluded the existence of invariant manifolds of $x' = f(x)$ which are C^k and tangent to the eigenspaces of the linearized system $y' = Df(x_0)\cdot y$.

Essentially this means that if $Df(x_0)$ has no eigenvalues with zero real parts (we call $Df(x_0)$ an <u>hyperbolic matrix</u> in this case) we can use the linearization procedure

to study the phase portrait of the non-linear system $x' = f(x)$ around the equilibrium x_0 by using the information arising from the study of the phase portrait of the linear system $y' = Df(x_0) \cdot y$ around $y = 0$.

Example 1

Consider the ODE system

$$(*) \quad \begin{cases} x' = y \\ y' = x(x-1)\left(x - \frac{1}{4}\right) \end{cases}$$

the equilibria of $(*)$ are

$$\begin{cases} 0 = y \\ 0 = x(x-1)\left(x - \frac{1}{4}\right) \end{cases} \iff \begin{cases} y = 0 \\ x = 0 \ \vee \ x = 1 \ \vee \ x = \frac{1}{4} \end{cases}$$

which gives the points $(0,0)$, $\left(\frac{1}{4}, 0\right)$, $(1,0)$

the jacobian of $f(x,y) = \left(y, x(x-1)\left(x - \frac{1}{4}\right)\right)$ is

$$Df(x,y) = \begin{bmatrix} 0 & 1 \\ 3x^2 - \frac{5}{2}x + \frac{1}{4} & 0 \end{bmatrix}$$

and hence at the equilibria we have

$$Df(0,0) = \begin{bmatrix} 0 & 1 \\ 1/4 & 0 \end{bmatrix} \longrightarrow \lambda_{\pm} = \pm \frac{1}{2} \quad , \quad \mathrm{Re}\,\lambda_{\pm} \neq 0$$

$$Df\left(\tfrac{1}{4},0\right) = \begin{bmatrix} 0 & 1 \\ -\frac{3}{16} & 0 \end{bmatrix} \longrightarrow \lambda_{\pm} = \pm i\sqrt{\tfrac{3}{16}} \quad , \quad \mathrm{Re}\,\lambda_{\pm} = 0$$

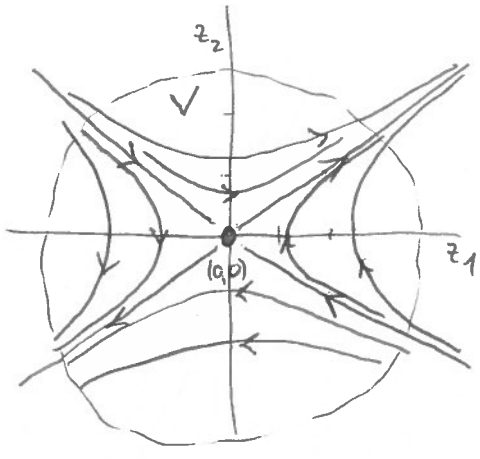
$$Df(1,0) = \begin{bmatrix} 0 & 1 \\ 3/4 & 0 \end{bmatrix} \longrightarrow \lambda_{\pm} = \pm \frac{\sqrt{3}}{2} \quad , \quad \mathrm{Re}\,\lambda_{\pm} \neq 0$$

So we can use linearization to study the behaviour of solutions of (*) around $(0,0)$ and $(1,0)$, but not around $(\frac{1}{4},0)$.

Let's consider what happens around $(0,0)$: the eigenvectors are $\lambda_+ = \frac{1}{2}$, $v_+ = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$; $\lambda_- = -\frac{1}{2}$, $v_- = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

So we have

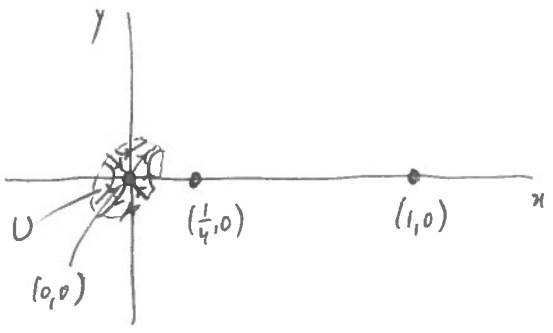
(linear system $\underline{z}' = \begin{bmatrix} 0 & 1 \\ 1/4 & 0 \end{bmatrix} \underline{z}$)



h

h^{-1}

(nonlinear (*) around $(0,0)$)



An analogous study can be done around $(1,0)$, but <u>not</u> around $(\frac{1}{4},0)$.

A problem remain: how to analyse the behaviour of solutions away from hyperbolic equilibria or when the equilibria is not hyperbolic? these are the most common cases but there is not a single technique or result that works in all cases: we will present next some tools that are useful in many situations, again without rigorous statements or details in some cases.

We start by the so called <u>conservative systems.</u>

An ODE system is called <u>conservative</u> if there exists a function $E : \mathbb{R}^n \to \mathbb{R}$, nonconstant in open sets of the phase space $\mathbb{R}^n$, such that E is constant when computed along solutions of the ODE.

this typically occurs in physical systems without friction or forcing, in which case E is the total energy of the system.

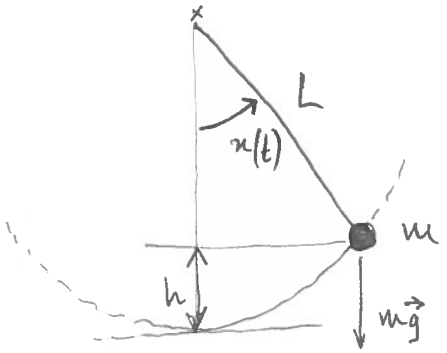
Usually E is a differentiable function and so the checking of the behaviour of E along solutions can be done by computing the value of $\frac{d}{dt} E(\underline{x}(t))$ and verifying if this quantity is, or not, constant for all t

If a function $E: \mathbb{R}^n \to \mathbb{R}$ is constant along solutions of $x' = f(x)$ then we conclude that the orbits of this ODE are (subsets of) level sets of E and so the computation of these level sets $\{x \in \mathbb{R}^n : E(x) = \text{constant}\}$ provides precious information about the qualitative behaviour of solutions of $x' = f(x)$.

Let us illustrate this with an example.

Example 2

Consider the gravitational pendulum already introduced in the first session : $x'' + \omega_0^2 \sin x = 0$, where $\omega_0^2 = \frac{g}{L}$



We know that this physical system has a mechanical energy, $E = E_{kinetic} + E_{potential}$, where

$$E_{kinetic} = \frac{1}{2} m (L x'(t))^2$$

$$E_{potential} = mgh(t) = mg(L - L\cos x(t))$$

Thus, we can write the pendulum equation as

$$(\dagger) \quad \begin{cases} x' = y \\ y' = -\omega_0^2 \sin x \end{cases}$$

and the energy

$$(\ddagger) \qquad E(x,y) := \frac{1}{2} m (Ly)^2 + mgL(1-\cos x).$$

Let $(x(t), y(t))$ be any solution of $(\dagger)$. Computing $\frac{d}{dt}E$ on this solution we have, for every t,

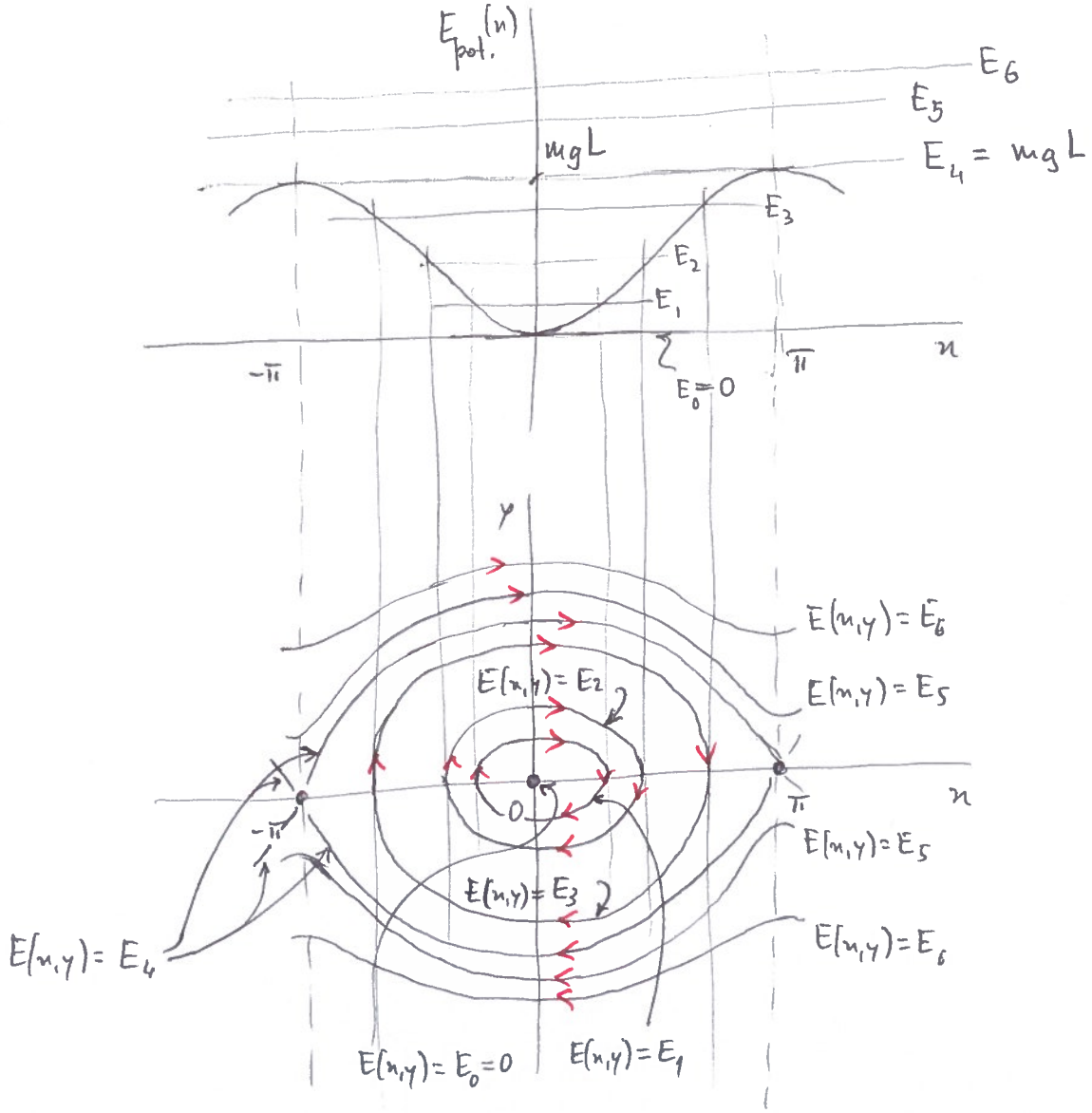
$$\begin{aligned} \frac{d}{dt} E(x(t), y(t)) &= \frac{1}{2} m 2 (Ly) L y' + mgL(\sin x)\, x' \qquad \text{by } (\dagger) \\ &= mL^2 y (-\omega_0^2 \sin x) + mgL(\sin x) y \\ &= mL^2 y (\sin x)\Big(\underbrace{-\omega_0^2 + \frac{g}{L}}_{=0}\Big) \\ &= 0 \end{aligned} \qquad (\text{because } \omega_0^2 = \tfrac{g}{L})$$

Thus, $(\dagger)$ is a conservative system and its orbits are part of the level sets of the total energy

$$E(x,y) = E_{kin.}(y) + E_{pot.}(x)$$

where $E_{kin.}(y) = \underbrace{\frac{1}{2} mL^2 y^2}_{\geq 0}$ and $E_{pot.}(x) = mgL(1-\cos x)$

Usually the determination of level sets of multivariable functio can be demanding but in the case the total energy has the above structure (the sum of a function only dependent of x and another only dependent of y which is non-negative) the drawing of the level sets becomes simpler from the plot of only $E_{pot.}(x)$ because we know in these cases that $E(x,y) \geq E_{pot.}(x)$ and what is missing is essentially the square of y (or of $-y$). Let's exemplify:



(for the arrows see next page)

to determine how the orbits, which are on those level sets, are evolving as t increases we get the information simple from (†): the first equation, $x' = y$, tells us that $x(t)$ is increasing when $y > 0$ and decreasing when $y < 0$, hence the arrows in red in the previous page.

It is important to interpret the phase portrait just obtained in terms of the movement of the physical pendulum we are attempting to model mathematically.

Let us see another example, originating in ecology:
Let x denote the concentration of preys and y those of predators in a fixed region. the evolution of these populations with time can be modelled, in a first approximation, by the famous Lotka-Volterra predator-prey equations

$$(\times) \quad \begin{cases} x' = x(a - by) \\ y' = y(-c + dx) \end{cases}$$

where a, b, c, d are positive constants.

Since x and y are populations they must be nonnegative and so the phase space of (X) is $\mathbb{R}^{2+}$. The equilibria of (X) are

$$\begin{cases} 0 = x(a-by) \\ 0 = y(-c+dx) \end{cases} \Leftrightarrow \begin{cases} x=0 \quad \vee \quad y=\frac{a}{b} \\ y=0 \quad \vee \quad x=\frac{c}{d} \end{cases}$$

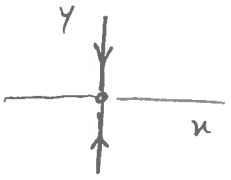
$(0,0)$, $\left(\frac{c}{d}, \frac{a}{b}\right)$.

From $x' = x(a-by)$ we conclude that if the initial data has $x=0$ then $x'=0$ and so x is a constant. this means that the line $\{x=0\}$ is invariant: if we start there the solution never leaves it.

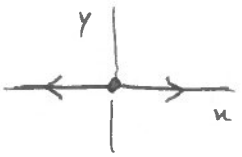
From $y' = y(-c+dx)$ the same conclusion is reached for the line $\{y=0\}$.

Furthermore, on these invariant lines we have

at $\{x=0\}$: $\begin{cases} x'=0 \\ y'=-cy \end{cases}$



at $\{y=0\}$: $\begin{cases} x'=ax \\ y'=0 \end{cases}$



From (x) we have following jacobian of the vector field $f(x,y) = (x(a-by), y(-c+dx))$:

$$Df(x,y) = \begin{bmatrix} a-by & -bx \\ dy & -c+dx \end{bmatrix}$$

and so, on the equilibrium $(0,0)$,

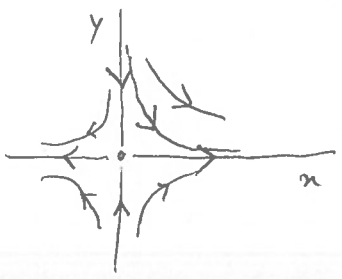
$$Df(0,0) = \begin{bmatrix} a & 0 \\ 0 & -c \end{bmatrix}$$

which has eigenvalues : $\lambda_+ = a > 0$, with eigenvector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$; and $\lambda_- = -c < 0$, with eigenvector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and on the equilibrium $\left(\frac{c}{d}, \frac{a}{b}\right)$,

$$Df\left(\frac{c}{d}, \frac{a}{b}\right) = \begin{bmatrix} 0 & -\frac{bc}{d} \\ \frac{ad}{b} & 0 \end{bmatrix}$$

which has eigenvalues : $\lambda_\pm = \pm i\sqrt{ac}$.

Thus, linearization is not applicable at $\left(\frac{c}{d}, \frac{a}{b}\right)$ but it is valid around $(0,0)$ and the phase portrait in a (small) neighborhood of this point is



Observe that by dividing the first equation in (x) from the second we obtain

$$\frac{dy}{dx} = \frac{y(-c+dx)}{x(a-by)} \iff \frac{dy}{dx} = \frac{-c+dx}{x} \cdot \frac{y}{a-by}$$

which is a separable equation that we easily integrate:

$$\frac{a-by}{y}\frac{dy}{dx} = \frac{-c+dx}{x} \iff \left(\frac{a}{y} - b\right)\frac{dy}{dx} = -\frac{c}{x} + d$$

$$\iff \int \left(\frac{a}{y} - b\right)\frac{dy}{dx}\,dx = \int \left(-\frac{c}{x} + d\right)dx$$

$$\iff a\log y - by + c\log x - dx = \alpha$$

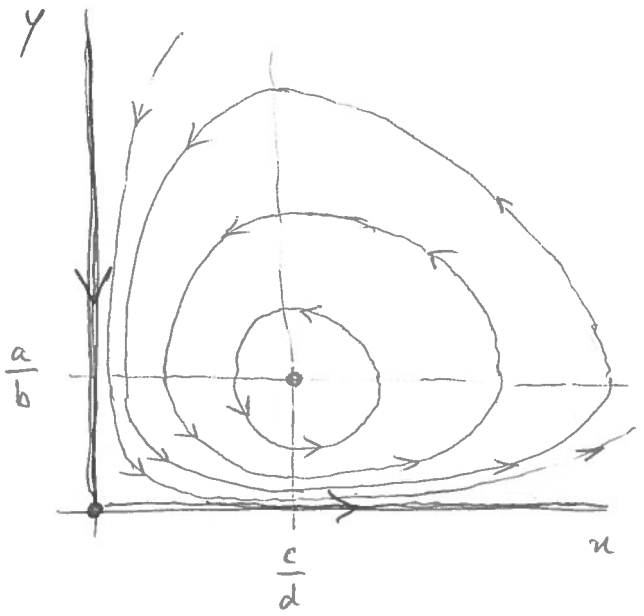
for some real valued constant α

This means that the function

$$E(x,y) := a\log y + c\log x - dx - by \tag{xx}$$

is a conserved quantity for the ODE system (x). Hence, the orbits of (x) are on level sets of $E(\cdot,\cdot)$. The trick used to plot these level sets in the case of the pendulum is not applicable in this case because the structure of E is not as before. Here

we need to plot the sets $\{(x,y) \in \mathbb{R}^{2+} : E(x,y) = \text{constant}\}$ by a careful analysis of (xx), or by using a computational help (such as Mathematica, Wolfram Alpha, Demos, etc.)



Now interpret this phase portrait in terms of the graphs of $x(t)$ and $y(t)$, and in terms of the population of the ecological species of preys (x) and predators (y).

GMIT

Differential Equations and Applications

Session 7: Lyapunov functions. Notions of stability of solutions. Relations between Lyapunov functions and stability of equilibria. Limit sets of orbits. Poincaré - Bendixon theorem (in $\mathbb{R}^2$).

It's clear that not all ODEs are linear or conservative. For example, consider the case of a gravitation pendulum with damping, modelled by the ODE

$$(*) \qquad x'' + 2\lambda x' + \omega_0^2 \sin x = 0 \quad , \quad \lambda > 0$$

As usual, transform this 2nd equation into a 1st order system by defining the auxiliary variable $y := x'$, thus getting

$$(**) \qquad \begin{cases} x' = y \\ y' = -\omega_0^2 \sin x - 2\lambda y \end{cases}$$

the phase space of $(**)$ is $\mathbb{R}^2$ and the equilibria are the solutions of

$$\begin{cases} 0 = y \\ 0 = -\omega_0^2 \sin x - 2\lambda y \end{cases} \iff \begin{cases} y = 0 \\ \sin x = 0 \end{cases} \iff (x, y) = (k\pi, 0)$$

as was the case of the undamped ($\lambda = 0$) pendulum.

System $(**)$ has also an associated Total (= kinetic + potential) energy

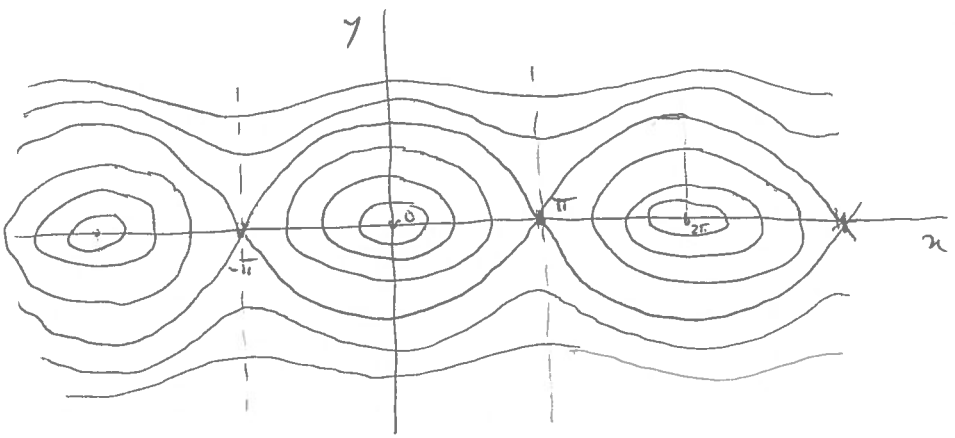
$$(***) \qquad E(x, y) = \frac{1}{2} m L^2 y^2 + mgL(1 - \cos x)$$

Computing the derivative of $E(x(t), y(t))$ where $(x(t), y(t))$ satisfy $(**)$ we get, remembering that $\omega_0^2 = \frac{g}{L}$

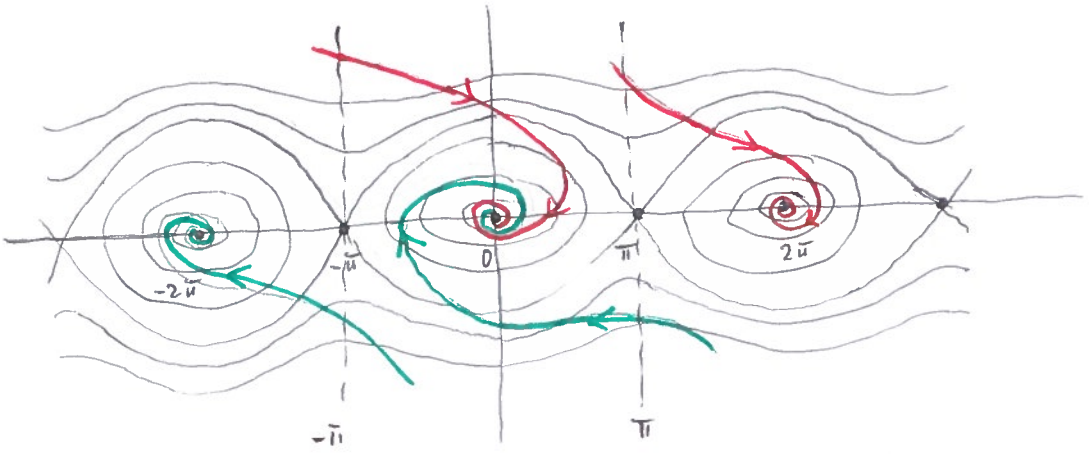
$$\frac{d}{dt}E(x(t),y(t)) = mL^2yy' + mgL(\sin x)x'$$
$$= mL^2y\left(-\frac{g}{L}\sin x - 2\lambda y\right) + mgL(\sin x)y$$
$$= -2mL^2\lambda y^2 \leq 0$$

So $E(x,y)$ is not constant along solutions: it is decreasing, and this means that orbits are directed throughwards level sets of $E(x,y)$ with lower and lower levels. this monotonicity is (almost) as good as having a conserved function because if we draw the level sets we know by $\frac{d}{dt}E(x(t),y(t)) \leq 0$ that orbits must decrease the E value which, together with information about the signs of x' and y' is, in many cases, sufficient to draw the phase portrait and understand the dynamics.

From the study of the level sets of E done in the study of the undamped pendulum we know these are



and knowing that $x(t)$ is increasing when $y>0$ (because $x'=y$) and decreasing when $y<0$, the information that along solutions $E(x,y)$ decreases with time t lead us to the phase portrait



A function like $E(x,y)$ in the above example, that is monotone (either decreasing or increasing) along solutions of a differential equation and has a local minimum at some equilibrium of the ODE is called a Lyapunov function for that equilibrium.

We observe from the previous example that solutions that start sufficiently close to the equilibria $(2k\pi, 0)$ with $k \in \mathbb{Z}$ will converge to that equilibria as $t \to +\infty$.

This behaviour means that the equilibrium points $(2k\pi, 0)$ are what is called asymptotically stable. Notions of <u>stability of equilibria</u> are central to the qualitative theory of ODEs and so we will concentrate a little more on them now.

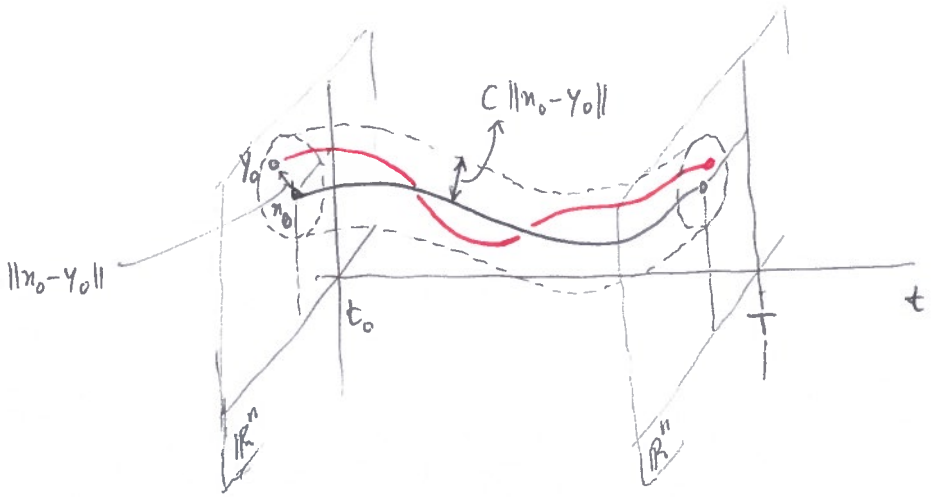
After the study of the Picard-Lindelöf theorem (on existance and uniqueness of solutions to initial value problems) we refer to a result about the continuous dependence of solutions on the initial data. We did not state the result then because a precise statement was not needed at that point. Here it convenient to do so. Under the assumptions of the Picard-Lindelöf theorem ($f:\mathbb{R}^n \to \mathbb{R}^n$ locally Lipschitz) the solutions of the initial value problem

$$\begin{cases} x' = f(x) \\ x(t_0) = x_0 \end{cases},$$

denoted by $x(t;t_0,x_0)$ depend on x_0 in a Lipschitz way; i.e.

$$\forall_{T>t_0} \exists_{C=C(T)} : \| x(t;t_0,x_0) - x(t;t_0,y_0) \| \leq C \| x_0 - y_0 \|, \quad \forall_{t\in[t_0,T]}$$

Schematically we have,



This conveys a kind of "stability" in bounded time intervals: if an initial condition y_0 is not too far from another x_0 then the corresponding solutions will also remain close to each other, at a distance not more than $C\|x_0 - y_0\|$.

The concepts of stability of solutions have to do with the cases of what happens in unbounded time intervals $[t_0, +\infty)$.

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz, and let the initial value problem for $x' = f(x)$ have solutions defined in time intervals $[t_0, +\infty)$.

A solution $x(t; t_0, \hat{x}_0)$ is <u>stable</u>, or <u>Lyapunov stable</u> if

$$\forall_{\varepsilon>0} \forall_{\tau \geq t_0} \exists_{\delta = \delta(\varepsilon,\tau)} : \forall_{t \geq \tau}, \ \|x_0 - \hat{x}_0\| < \delta \Rightarrow$$

$$\Rightarrow \|x(t; t_0, x_0) - x(t; t_0, \hat{x}_0)\| < \varepsilon$$

A solution $x(t; t_0, \hat{x}_0)$ is <u>asymptotically stable</u> if it is stable and also

$$\exists_{b = b(t_0)} : \|x_0 - \hat{x}_0\| < b \Rightarrow \|x(t; t_0, x_0) - x(t; t_0, \hat{x}_0)\| \to 0$$

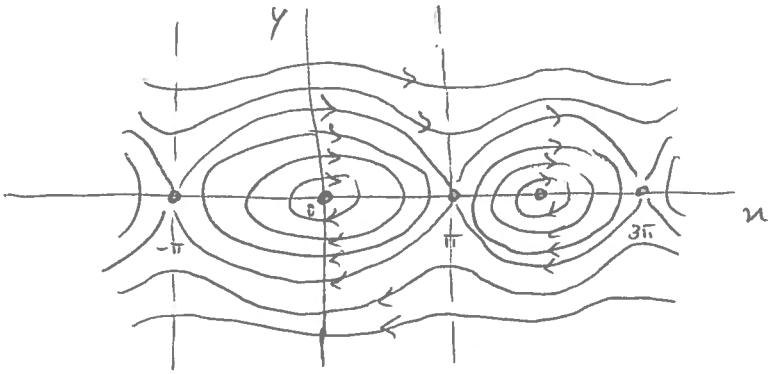
as $t \to +\infty$.

A solution $x(t; t_0, \hat{x}_0)$ is <u>unstable</u> if it is not stable.

Some examples are useful here:

Example 1

Consider the undamped pendulum $x'' + \sin x = 0$.
We saw that the phase portrait is



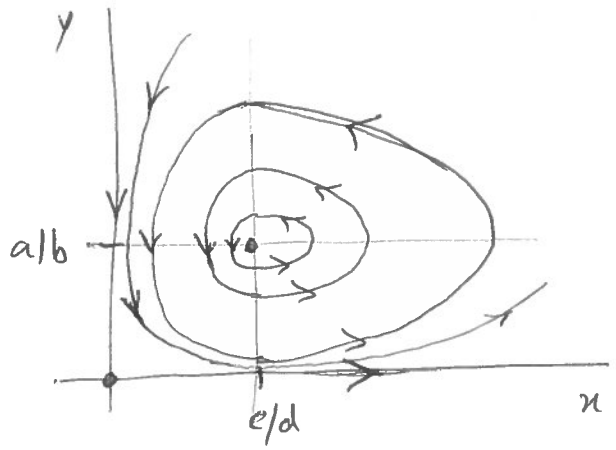
and it is clear from this that the equilibria $(2k\pi, 0)$, $k \in \mathbb{Z}$, are all stable, whereas the equilibria $((2k+1)\pi, 0)$, $k \in \mathbb{Z}$, are all unstable

Example 2

In the Lotka-Volterra predator-prey system

$$\begin{cases} x' = x(a - by) \\ y' = y(-c + dx) \end{cases}$$

the phase portrait was seen to be



and it is also clear that $(0,0)$ is an unstable equilibrium whereas $\left(\frac{c}{d}, \frac{a}{b}\right)$ is a stable one.

Example 3

For the damped pendulum $x'' + 2\lambda x' + \sin x = 0$ we saw at the start of this session that most solutions will converge to the equilibria $(2k\pi, 0)$, $k \in \mathbb{Z}$, which are, then, asymptotically stable (their stability must be checked before we can conclude the asymptotic stability, but this is not difficult using the behaviour of the energy function E along solutions and the level curves of this function — in particular that $(2k\pi, 0)$ are local minima of E.)

Likewise it is clear from the phase portrait that the equilibria $((2k+1)\pi, 0)$ are unstable.

these examples seem to exhibit a relation between the existence of a conserved function (in examples 1 and 2) or a Lyapunov function (in example 3) and the stability or asymptotical stability of certain equilibria.

In fact, the following result holds:

Theorem

Let $f: D \subset \mathbb{R}^n \to \mathbb{R}^n$ be C^1 on the open set D.
Let $x_* \in D$ be an equilibrium of $x' = f(x)$.
then (i) if there exists a Lyapunov function for x_* then x_* is stable,
(ii) if there exists a strict Lyapunov function for x_* then x_* is asymptotically stable.

Obs.:

To fix the notion, a function E is a Lyapunov function for x_* if $\exists U \subset D$, open, such that $E(x_*) = 0$, $E(x) > 0 \ \forall x \in U \setminus \{x_*\}$ and $\frac{d}{dt} E(x(t)) \leq 0$ for $x(t) \in U$. The Lyapunov function is called strict Lyapunov if the last inequality holds with $<$ instead of $\leq$.

In the study of stability we are interested in the behaviour of orbits as $t \to +\infty$. To describe what are the orbits approaching when $|t|$ is very large the following concepts of α-limit and ω-limit sets are introduced:

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be locally Lipschitz and denote by $\varphi_t(x_0)$ or by $x(t;0,x_0)$ the unique solution of $x' = f(x)$ with $x(0) = x_0$. Suppose its maximal interval of definition $I_{max}(x_0)$ is $\mathbb{R}$.

Remember that the <u>orbit</u> of $x_0 \in \mathbb{R}^n$ is

$$\gamma(x_0) = \left\{ \varphi_t(x_0) : t \in I_{max}(x_0) \right\}$$

the <u>positive semi-orbit</u> is

$$\gamma^+(x_0) := \left\{ \varphi_t(x_0) : t \in I_{max}(x_0) \cap \mathbb{R}^+ \right\}$$

and the <u>negative semi-orbit</u> is

$$\gamma^-(x_0) := \left\{ \varphi_t(x_0) : t : I_{max}(x_0) \cap \mathbb{R}^- \right\}$$

the limit sets of an orbit are

$$\alpha(x_0) := \bigcap_{y \in \gamma(x_0)} \overline{\gamma^-(y)}$$

$$\omega(x_0) := \bigcap_{y \in \gamma(x_0)} \overline{\gamma^+(y)}$$

Let us see some examples:

Example 4

For the harmonic pendulum (i.e., the undamped pendulum with x restricted to very small oscillations) $x'' + x = 0$ the 1st order system is

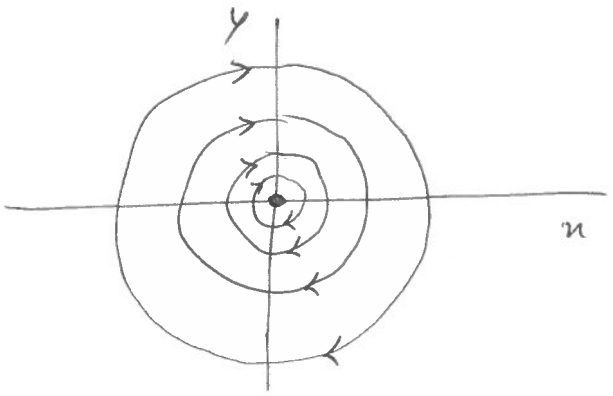
$$\begin{cases} x' = y \\ y' = -x \end{cases}$$

and it is clear by multiplying the first equation by x and the second by y and summing them that we get

$$\frac{d}{dt}\left(x(t)^2 + y(t)^2\right) = 0 \quad , \quad \forall t$$

As the level sets of the function $E(x,y) = x^2 + y^2$ are circunferences we conclude that the phase portrait

is



and thus it is clear to conclude that,

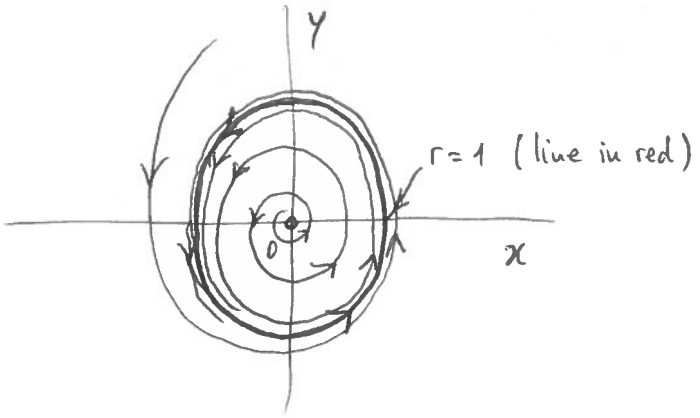
$\forall x_0 \in \mathbb{R}^2, \quad \alpha(x_0) = \omega(x_0) = \gamma^+(x_0) = \gamma^-(x_0) = \gamma(x_0)$

and the orbits are either the equilibrium point $(0,0)$ or circumferences centered at the origin corresponding to periodic solution of the equation: by this obvious reason they are called periodic orbits.

Example 5

Consider the following ODE written in polar coordinates $\begin{cases} r' = r(1-r) \\ \theta' = 1 \end{cases}$, and let $(x,y) = (0,0)$ be an equilibrium.

From this ODE we get the following phase portrait



Let $S_1 = \{(x,y) \in \mathbb{R}^2 : x^2+y^2=1\}$.

Then we have

$$P_0=(x_0,y_0) \in \{(0,0)\} \cup S_1 \Rightarrow \alpha(P_0) = \omega(P_0) = \gamma^+(P_0) = \gamma^-(P_0) = \gamma(P_0)$$

$$0 < \|(x_0,y_0)\| < 1 \Rightarrow \alpha(P_0) = \{(0,0)\}$$
$$\omega(P_0) = S_1$$

$$\|(x_0,y_0)\| > 1 \Rightarrow \alpha(P_0) = \phi$$
$$\omega(P_0) = S_1$$

As we saw in this last example, the α-limit set of an orbit, as well as its ω-limit set, can be empty. However if the corresponding semi-orbit is bounded it is easy to conclude (using the fact, from topology, that the intersection of closed sets is closed) that the limit set is non-empty, compact, and connected.

We saw above examples in which the α-limit or ω-limit sets are equilibria and periodic orbits. In general they

can be extremely complicated sets if the phase space has dimension $n \geq 3$ (and this is related to the existence of "deterministic chaos" and "strange attractors")

Clearly, for $n=1$ nothing more complicated than single equilibria can happen as α- and ω-limit sets of orbits.

For dimension $n=2$ the situation is also relatively simple, as a consequence of the Jordan curve theorem we have the following important result

Theorem (Poincaré-Bendixon)

Let $f: \mathbb{R}^2 \to \mathbb{R}^2$ be a C^1 function

If the semi-orbit $\gamma^+(x)$ of $x' = f(x)$ is bounded and $\omega(x)$ does not have critical points,

then $\omega(x)$ is a periodic orbit.

[Same holds by changing γ^+ to γ^- and ω- to α-limit.]

A consequence of this theorem is that, in 2-dimensional phase spaces, if $\gamma^+(x)$ is contained in a compact set with just a finite number of equilibria of $x' = f(x)$

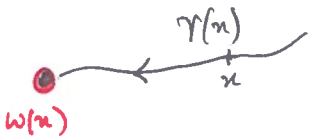
then either (i) $\omega(x)$ is an equilibrium

(ii) $\omega(x)$ is a periodic orbit

(iii) $\omega(x)$ is the union of a finite number of equilibrium points and orbits connecting them.

To illustrate the situation :

(i)



(ii)

(iii)

or

(Obs.: an orbit like the one in red in  is called an homoclinic orbit : it is such that its α- and ω-limit sets are the same equilibrium point; if the situation is like  then the orbit in red is called heteroclinic: its α- and ω-limit sets are two different equilibria.)

For systems of ODEs in 2-dimensional phase spaces it is a good idea to have these results in mind when analysing the behaviour of solutions. We illustrate its use in the following example:

Example 6

Consider the ODE

$$(*) \qquad \begin{cases} x' = x(x^2+y^2-3x-1) - y \\ y' = y(x^2+y^2-3x-1) + x \end{cases}$$

We are interested is knowing if $(*)$ has some periodic solution.

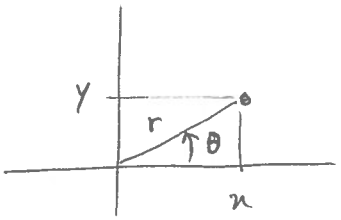
The phase space of $(*)$ is $\mathbb{R}^2$ and it is easy to see that $(0,0)$ is an equilibrium. To check the existence of other equilibria let us study the system

$$\begin{cases} 0 = x(x^2+y^2-3x-1) - y \\ 0 = y(x^2+y^2-3x-1) + x \end{cases} \iff \begin{cases} x(x^2+y^2-3x-1) = y \\ y(x^2+y^2-3x-1) = -x \end{cases}$$

multiplying the first equation by y and the second by $-x$ and adding the results we get $0 = y^2 + x^2$ which has as its only solution $(x,y) = (0,0)$. Thus, the origin is the only equilibrium of $(*)$.

Now, if there exists a periodic orbit of (*) it must have $(0,0)$ in its interior (why? use Poincaré-Bendixon theorem!). this suggests using polar coordinates (which are centered at $(0,0)$) to investigate the issue.

Let $x = r\cos\theta$, $y = r\sin\theta$



then, the evolution of r and θ with t comes from these and from (*) as follows: from the change of coordinates

$$x' = r'\cos\theta - r\theta'\sin\theta$$

$$y' = r'\sin\theta + r\theta'\cos\theta$$

and substituting into (*) we get

$$\begin{cases} r'\cos\theta - r\theta'\sin\theta = r\cos\theta\left(r^2 - 3r\cos\theta - 1\right) - r\sin\theta \\ r'\sin\theta + r\theta'\cos\theta = r\sin\theta\left(r^2 - 3r\cos\theta - 1\right) + r\cos\theta \end{cases}$$

multiplying the first equation by $\cos\theta$, the second by $\sin\theta$ and summing both we get

$$r' = r\left(r^2 - 3r\cos\theta - 1\right)$$

and multiplying the first equation by $-\sin\theta$, the second by $\cos\theta$ and adding the results we get

$$r\theta' = r$$

Thus, we obtain (*) written in polar coordinates ($r>0, \theta\in\mathbb{R}$)

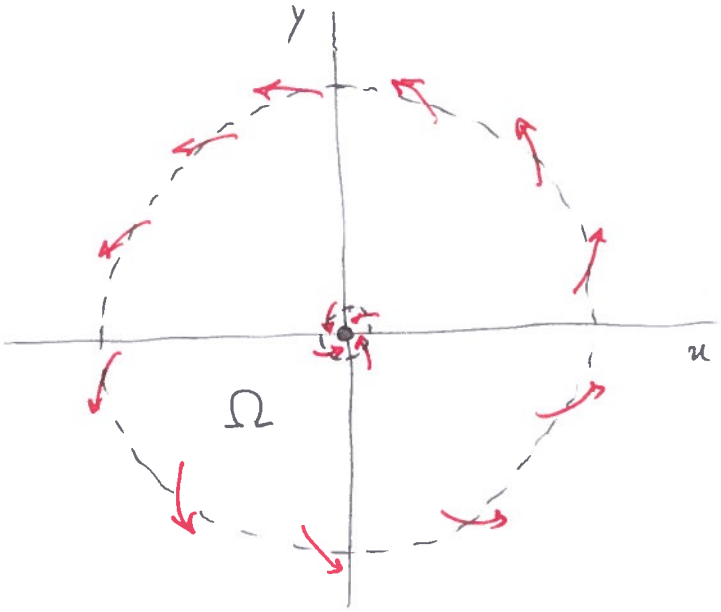
$$(**) \quad \begin{cases} r' = r(r^2 - 3r\cos\theta - 1) \\ \theta' = 1 \end{cases}$$

(From this we also conclude that there are no equilibria with $r>0$ because the equation for θ is $\theta'=1$, so this component of the vector field is never 0.)

Now consider a ball centered at $(0,0)$ with a very small radius, say $r = 10^{-2}$. Then, from the first equation in (**) we have that on those points $r'<0$, which means that the orbits of (**) on those points move to the direction of smaller values of r, i.e., to points closer to $(0,0)$.

If we take now a ball with a sufficiently large r, say $r = 10^2$, we again conclude from (**) that $r'>0$

on those points and so the points on those orbits move to points with larger r when t increases. the situation is, thus, like the one plotted next ($\theta' = 1 > 0 \Rightarrow \theta \uparrow \Rightarrow$ ↻):



Now in the region $\Omega = \{(x,y) : 10^{-2} \leq \|(x,y)\| \leq 10^{2}\}$ between those two balls there are no equilibria (we proved that the only equilibrium is $(0,0)$) and orbits with an initial point $(x_0, y_0) \in \Omega$ have its α-limit set in Ω (because the orbits can only leave Ω forward in time, so $\gamma^-(x_0,y_0) \in \Omega$ and thus $\alpha(x_0,y_0) \in \Omega$ also. But then, the Poincaré-Bendixon theorem implies that $\alpha(x_0,y_0)$ must be a periodic orbit.

Can (*) have more than one periodic orbit? Does it have? How could one check this (computationally)?

GMIT

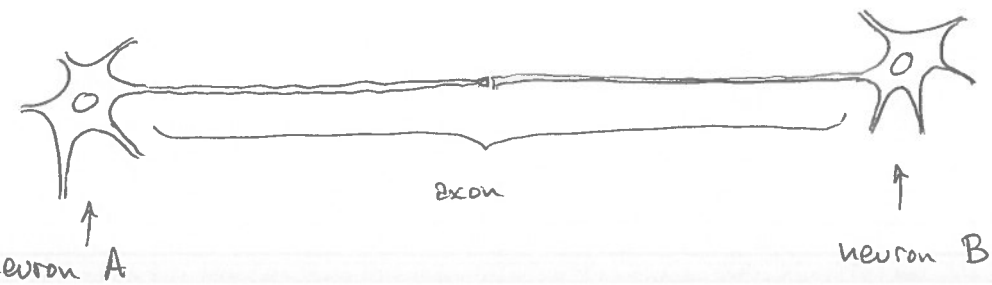
Differential Equations and Applications

Session 8: Applications of Differential Equations: study of a model for the propagation of a nerve impulse in axons (Hodgkin-Huxley model.)

the transmission of nerve impulses in a axon connecting two neurons (brain cells) is a complicated physiological process whose understanding is crucial to understand brain activity. In 1952 two scientists, A.L. Hodgkins and A.F. Huxley published an article about the mathematical modelling of this process in J. Physiol. 117 (4), 500-544 (1952), with the title "A quantitative description of membran current and its application to conduction and excitation in nerve". this article, not a mathematical article but one using mathematics in a nontrivial way became instantaneously famous and the authors recieved the Nobel prize in physiology and medicine, in 1963, largely due to this study.

A simplified version of the situation is the following:

We want to model the propagation of a nerve impulse (an electric depolarization wave) along an axon connecting two nerve cells



the axon is several orders of magnitude longer than the size of the neurons and so it is a reasonable modelling assumption to consider it has an infinite length and modelling it as the real line $\mathbb{R}$

(real situation)

(modelling situation)

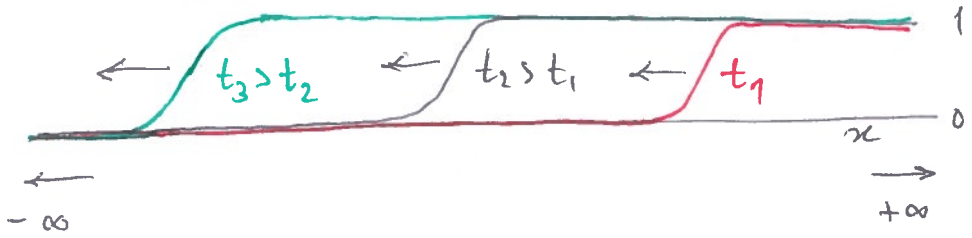
the electrical charge along the (mathematical) axon is, in a simplified version, a function $u = u(x,t)$ that satisfies the partial differential equation that after some rescaling and fixing of the parameters can be written as

$$(*) \qquad \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(u)$$

where $f(u) = -u(u-1)(u-\frac{1}{4})$.

We are interested in studying the possibility of (*) supporting a solution corresponding to the propagation of a signal from neuron B (now at $x = +\infty$) to neuron A (now at $x = -\infty$) with constant speed V, let us say: satisfying the "limit conditions"

$$(**) \qquad \begin{cases} u(x,t) \longrightarrow 0 & \text{as } x \longrightarrow -\infty \\ u(x,t) \longrightarrow 1 & \text{as } x \longrightarrow +\infty \end{cases}$$



Equation (*) is a partial differential equation (not an ODE...) but we are looking for a particular form of solution : a function $u(x,t)$ that corresponds to the propagation of a "signal" from $x=+\infty$ to $x=-\infty$ at constant speed, say V.

Thus, we are interested in a solution with the form

$$(***) \qquad u(x,t) = \varphi(x+Vt)$$

for some function φ and constant $V>0$.

To know if such solutions exist we need to get an equation that would be satisfied by φ and study it.

From (***) we have

$$\frac{\partial u}{\partial t} = \frac{\partial \varphi}{\partial t}\underbrace{(x+Vt)}_{s} = \frac{d\varphi}{ds}\frac{\partial s}{\partial t} = \frac{d\varphi}{ds}\cdot V$$

$$\frac{\partial u}{\partial x} = \frac{\partial \varphi}{\partial x}\underbrace{(x+Vt)}_{s} = \frac{d\varphi}{ds}\frac{\partial s}{\partial x} = \frac{d\varphi}{ds}\cdot 1$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{d\varphi}{ds}\right) = \frac{d^2\varphi}{ds^2}\frac{\partial s}{\partial t} = \frac{d^2\varphi}{ds^2}\cdot 1$$

$$f(u) = f[u(x,t)] = f(\varphi(x+Vt)) = f(\varphi(s)) = f(\varphi)$$

and so (*) is transformed into

$$\frac{d\varphi}{ds} V = \frac{d^2\varphi}{ds^2} + f(\varphi)$$

which can be written as

$$(*) \qquad \varphi'' - V\varphi' + f(\varphi) = 0$$

where $(\cdot)' = \frac{d}{ds}$.

Observe that (*) is like a pendulum equation although the term $-V\varphi'$ is not a damping (because $-V < 0$) but a forcing, and the function f is far from the function comming from a gravitational potential.

However, we can try to construct a "total energy" similar to what was done in the case of the pendulum, namely

$$E(\varphi, \varphi') := \underbrace{\frac{1}{2}(\varphi')^2}_{=: E_{kin}(\varphi')} + \underbrace{\int_0^{\varphi} f(\nu)\, d\nu}_{=: E_{pot}(\varphi)}.$$

Let us see what information this provides about the solutions of (*).

the limit conditions (**) become

$$(**) \qquad \begin{cases} \varphi(s) \longrightarrow 0 & \text{as } s \to -\infty \\ \varphi(s) \longrightarrow 1 & \text{as } s \to +\infty. \end{cases}$$

transforming (*)-(**) into a problem for a first order ODE we obtain, with $Y_1 = \varphi$, $Y_2 = Y_1' = \varphi'$,

$$(+) \qquad \begin{cases} Y_1' = Y_2 \\ Y_2' = V Y_2 - f(Y_1) \end{cases}$$

and we are looking for an heteroclinic solution that will satisfy $(Y_1, Y_2)(s) \longrightarrow (0,0)$ as $s \to -\infty$ and $(Y_1, Y_2)(s) \to (1,0$ as $s \to +\infty$.

Let us then try the "energy" E introduced before computing its behaviour on solutions of (+):

$$\frac{d}{dt} E(Y_1(t), Y_2(t)) = \frac{\partial E}{\partial Y_1} Y_1' + \frac{\partial E}{\partial Y_2} Y_2'$$

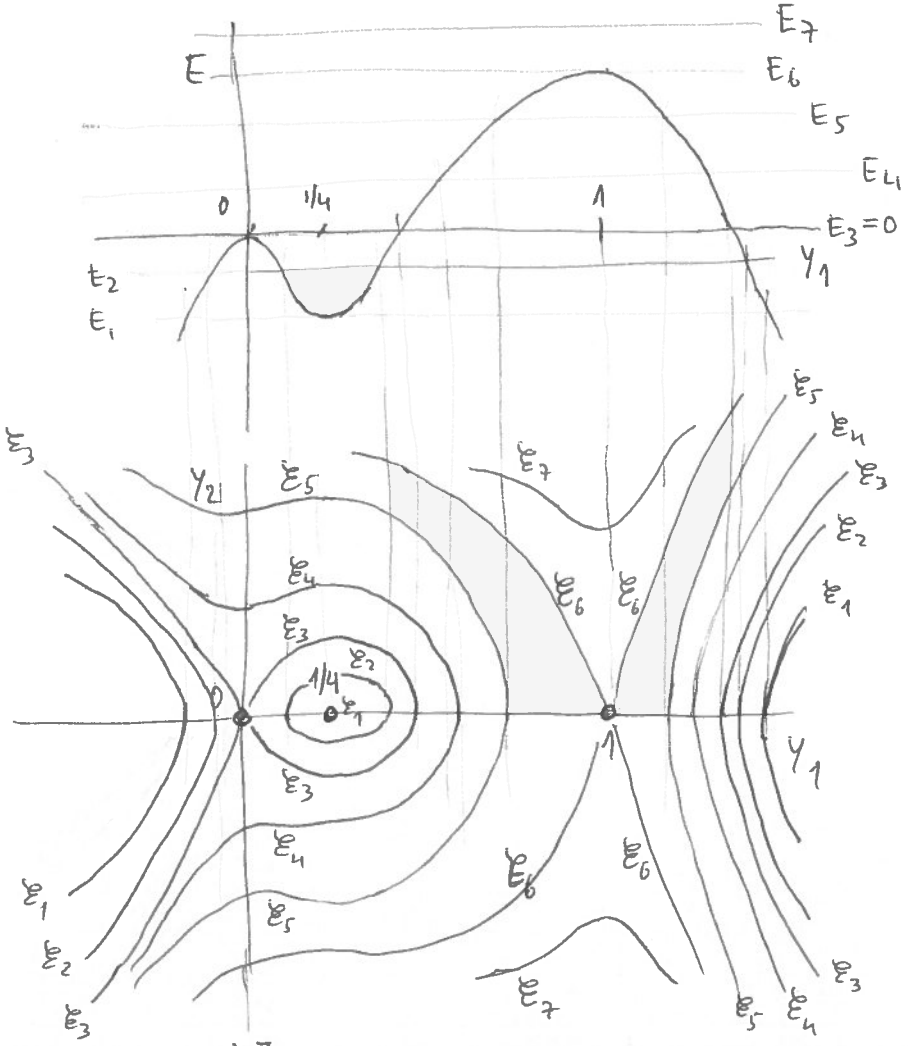
$$= f(Y_1) Y_2 + Y_2 \cdot (V Y_2 - f(Y_1))$$

$$= V Y_2^2 \geq 0$$

So, if $V = 0$ (which corresponds to a "standing wave" of the equation for u) the "energy" E is a conserved

quantity. for $V>0$ (as we want) the quantity $-E$ is a Lyapunov function for the equilibria for which E has a local minimum. In any case it is useful to draw the level sets of E in order to understand what happens when $V>0$.

In order to do this observe that $E(y_1,y_2)$ has the same structure of a mechanical energy: the sum of a "kinetic" term depending only on y_2 and nonnegative, and a "potential" contribution dependent of y_1 only. Thus, the level sets are



$[\mathcal{E}_i = \{(y_1,y_2) : E(y_1,y_2) = E_i\}]$

If $V>0$ the orbits are directed in the direction of increasing values of E so, the question now is: are there any orbits that "leaving" from $(0,0)$ when $s\to-\infty$, will increase its "energy" just enough to get to $(1,0)$ when $s\to+\infty$?.

How do solutions "leave" a neighborhood of $(0,0)$? to answer this let us try to see if linearization about $(0,0)$ can be used.

the jacobian of the vector field of (†) at a generic point (y_1,y_2) is

$$A(y_1,y_2)=\begin{bmatrix}0 & 1\\ -f'(y_1) & V\end{bmatrix}=\begin{bmatrix}0 & 1\\ 3y_1^2-\frac{5}{2}y_1+\frac{1}{4} & V\end{bmatrix}$$

thus

$$A(0,0)=\begin{bmatrix}0 & 1\\ \frac{1}{4} & V\end{bmatrix}$$

and its eigenvalues are $\lambda(V-\lambda)-\frac{1}{4}=0 \iff \lambda_{\pm}=\dfrac{V\pm\sqrt{V^2+1}}{2}$;

thus, $\lambda_-<0<\lambda_+$. the eigenvector associated to these eigenvalues are

$$\lambda_+ = \frac{1}{2}\left(v + \sqrt{v^2+1}\right) \quad \rightsquigarrow \quad \underline{0} = \left(A(0,0) - \lambda_+ I_2\right)\begin{pmatrix} v_1^+ \\ v_2^+ \end{pmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}\left(v+\sqrt{v^2+1}\right) & 1 \\ \frac{1}{4} & \frac{1}{2}\left(v-\sqrt{v^2+1}\right) \end{bmatrix}\begin{bmatrix} v_1^+ \\ v_2^+ \end{bmatrix}$$

thus $v_2^+ = \frac{1}{2}\left(v + \sqrt{v^2+1}\right)v_1^+$ and

$$v^+ = \begin{pmatrix} 1 \\ \frac{1}{2}\left(v+\sqrt{v^2+1}\right) \end{pmatrix} v_1^+$$

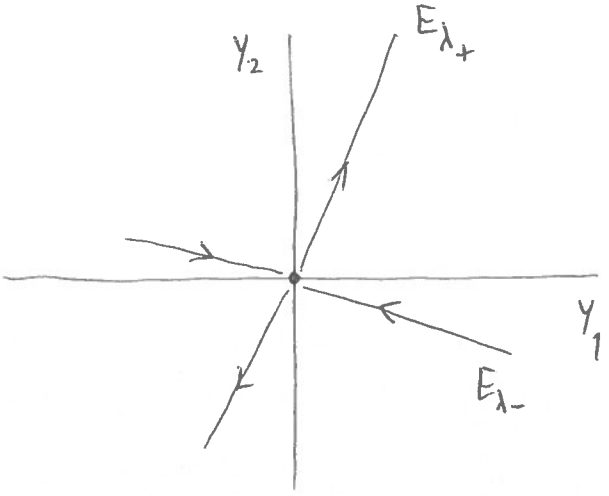
$$\lambda_- = \frac{1}{2}\left(v - \sqrt{v^2+1}\right) \quad \rightsquigarrow \quad 0 = \left(A(0,0) - \lambda_- I_2\right)\begin{pmatrix} v_1^- \\ v_2^- \end{pmatrix}$$

$$= \begin{bmatrix} -\frac{1}{2}\left(v-\sqrt{v^2-1}\right. & 1 \\ \frac{1}{4} & \frac{1}{2}\left(v+\sqrt{v^2+1}\right) \end{bmatrix}\begin{bmatrix} v_1^- \\ v_2^- \end{bmatrix}$$

thus $v_2^- = \frac{1}{2}\left(v - \sqrt{v^2-1}\right)v_1^-$ and

$$v^- = \begin{pmatrix} 1 \\ \frac{1}{2}\left(v-\sqrt{v^2+1}\right) \end{pmatrix} v_1^-$$

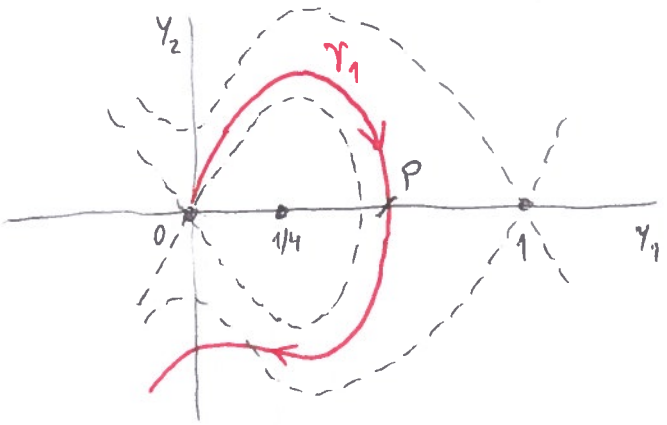
So, the eigenspaces associated with the eigenvectors are



$(0 < v < +\infty)$

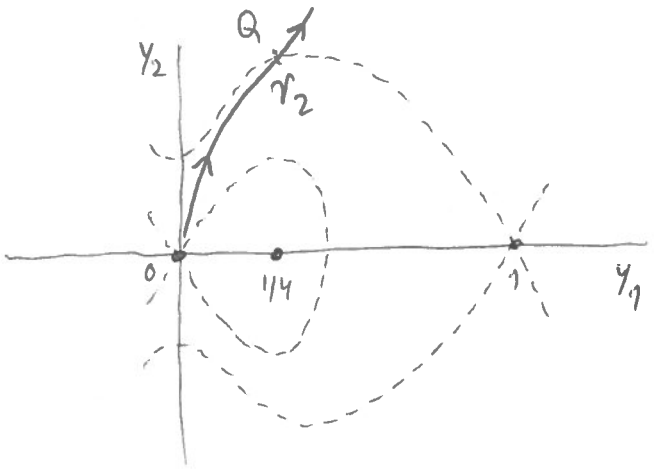
Note that the slope of E_{λ_+} is equal to $\frac{1}{2}(V+\sqrt{V^2+1})$ and so it is equal to $\frac{1}{2}$ when $V=0$ and increases monotonically to $+\infty$ when $V \to +\infty$. Likewise, the slope of E_{λ_-} is equal to $\frac{1}{2}(V-\sqrt{V^2+1})$, and so is equal to $-\frac{1}{2}$ when $V=0$ and increases to 0 when $V \to +\infty$.

Now suppose we have $V>0$ very small and consider an initial point $(\gamma_1(0), \gamma_2(0))$ in a close neighborhood of $(0,0)$ such that $(\gamma_1(s), \gamma_2(s)) \longrightarrow (0,0)$ as $s \to -\infty$. This point exists due to the Hadamard-Perron theorem (i.e., it is a consequence of linearization!). By the continuous dependence of the solution relative to the initial conditions and parameters the orbit γ_1 corresponding to this initial point has a behaviour as presented next



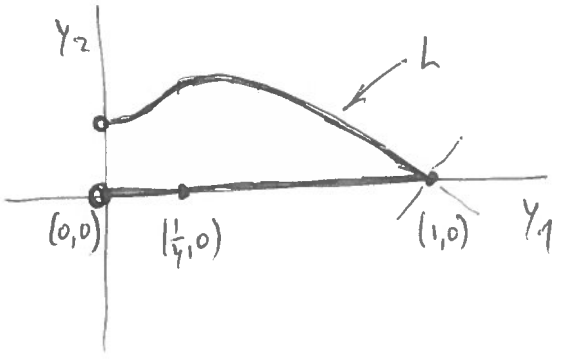
Consider now $V=V_2>V_1>0$ sufficiently large. For an initial point as before, i.e., such that the orbit will

converge to $(0,0)$ as $s \to -\infty$ we now have something like



Consider now the function $\Gamma : \mathbb{R}^+ \to \mathbb{R}^2$ defined as follows: for each $V>0$, let $\Gamma(V)$ be the point of intersection of the closure of the orbit of $(+)$ that satisfy $(y_1(s), y_2(s)) \longrightarrow$ $\longrightarrow (0,0)^+$ as $s \to -\infty$, with the line L defined by

$$L = \{(y_1, y_2) : y_2 = 0 \wedge y_1 \in (0,1)\} \cup \{\mathcal{E}_0 \cap \{(y_1, y_2) : y_2 > 0 \wedge y_1 \in (0,1)\}\}$$



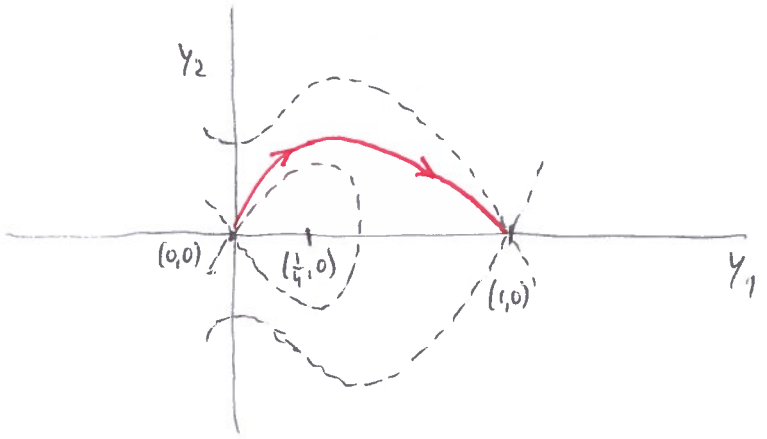
Let $g : L \to \mathbb{R}_0^+$ the length of the arc of L between $(0,0)$ and a point $X \in L$ and let $\tilde{l} = g(L)$ the total length of L. g is a bijection between L and $(0, \tilde{l})$. By uniqueness of

of solutions and by the continuous dependence of solutions relative to initial conditions and parameters the function $g \circ \Gamma : (0, +\infty) \ni V \longrightarrow g(\Gamma(V)) \in (0, \tilde{l})$ is a continuous and strictly monotonic function of V.

We saw that

$$(g \circ \Gamma)(V_1) = g(P) < 1 < g(Q) = (g \circ \Gamma)(V_2)$$

and so, by the intermediate value theorem it must exist a $V \in (V_1, V_2)$ such that $(g \circ \Gamma)(V) = 1$. But by definition of g and Γ the only point where g is 1 is $(1,0)$. Hence, we proved that there exists a value $V > 0$ such that the corresponding solution satisfies



this means that there is a unique traveling wave speed $V > 0$ and a wave profile φ such that (*)-(**) has a solution $u(x,t) = \varphi(x + Vt)$, and so the Hodgkin-Huxley equations (*) support the existence of a nerve impulse propagating at constant speed: a nice application on Mathematics to Biology!

GMIT

Differential Equations and Applications

Session 9: Applications of Differential Equations:
Diffusion equations: balance laws, constitutive laws; continuous limit of a 1-dim random walk; some examples of linear and nonlinear diffusion equations.

We now turn our attention to partial differential equations, i.e., differential equations whose unknown are functions of two or more independent variables.

We already saw one such function in the previous session the Hodgkin-Huxley partial differential equations, but we could solve our problem with that equation using techniques of ordinary differential equations; this is far from being the usual case: partial differential equations is a very rich and active mathematical field, also much harder than ordinary differential equations, and having techniques and tools much more sophisticated. We will touch some of them in the second half of this course illustrating how the easiest of those differential equations (the linear diffusion and the linear wave equations) can be studied.

We start by deducing how the diffusion equation appear as a model for the spread of particles in a line, or the diffusion of heat in a wire.

Suppose we have particles in a (essentially) one dimensional medium, i.e., a line modelled by the real line $\mathbb{R}$. Consider an interval $[a,b]$ and let us see what processes can lead to a change in the number of particles if they cannot be created or destroyed. In this case the only possibility for the change in the number of particles in $[a,b]$ is that they can enter or leave this interval at $x=a$ and $x=b$.

Let $c(x,t)$ denote de concentration of particles at location x and time t. Then, the total amount of particles in $[a,b]$ is given by $\int_a^b c(x,t)\,dx$.

Let $J(x,t)$ be the net number of particles that pass a point x from left to right, at time t, per unit time. Then, without creation or destruction of particles we have the following balance law (or conservation law)

$$\begin{pmatrix}\text{rate of change}\\ \text{of particles in}\\ \text{interval } [a,b]\end{pmatrix} = \begin{pmatrix}\text{flux of particles going into } [a,b]\\ \text{minus the flux of particles}\\ \text{going out of } [a,b]\end{pmatrix}$$

$$\frac{d}{dt}\int_a^b c(x,t)\,dx = J(a,t) - J(b,t)$$

Assuming the flux J is a C^1 function we can use the Fundamental Theorem of Calculus to write

$$J(a,t) - J(b,t) = -\int_a^b \frac{\partial J}{\partial x}\, dx$$

and this means that the balance law can be written as

$$\int_a^b \frac{\partial c}{\partial t}\, dx + \int_a^b \frac{\partial J}{\partial x}\, dx = 0$$

assuming $c(x,t)$ is smooth enough so that we can exchange the integral and the derivative. So, in this case we have

$$\int_a^b \left(\frac{\partial c}{\partial t} + \frac{\partial J}{\partial x}\right) dx = 0$$

and if this is valid for all intervals $[a,b]$ in the real line (and considering we are interested in modelling physical processes where the functions inside the integral are considered continuous) we can finally write the <u>balance law</u> as follows:

$$(*) \qquad \frac{\partial c}{\partial t} = -\frac{\partial J}{\partial x}.$$

this equation is not yet a differential equation that we can study, because we do not know what relates c and J. Establishing the relation between these two functions depends

on the assumptions being considered in the modelling process, and in some cases can be justified by deeper considerations. On a more basic level those constitutive laws providing a relation between J and c are obtained phenomenologically by measurement on some observational or experimental situation. In the present case of particles diffusing on a 1-dimensional medium a commonly used constitutive law is Fick's law of diffusion

$$J = -D\frac{\partial c}{\partial x} \qquad (**)$$

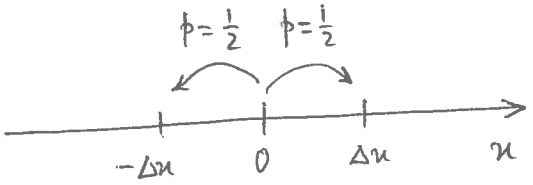
where D is a positive constant, called the diffusion coefficient. (If we are modelling diffusion of heat in a wire, c would be the temperature, J the flow of heat, and (**) is called the Fourier law of cooling.)

Using (**) in (*) we finally get the diffusion equation (or the heat equation) governing the evolution of the concentration of particles at location x and time t:

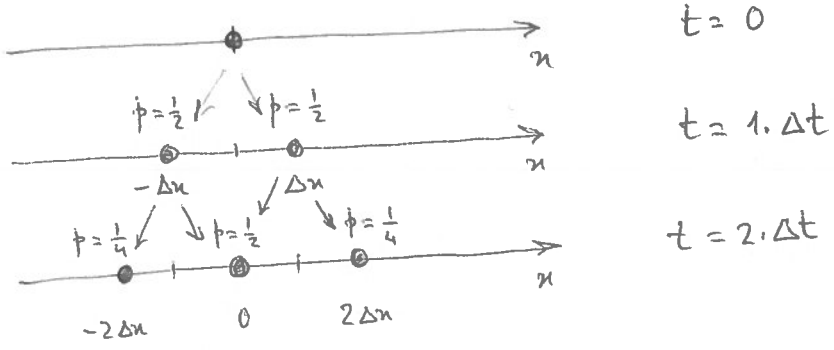
$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \qquad (*)$$

It is very interesting to observe that the diffusion equation can be obtained as the continuous limit of a stochastic process, modelling the random movement of particles in a straight line, known as a random walk.

Consider a system of particles that can move left and right on a straight line. Suppose that at time $t=0$ there is a particle at position $x=0$ and after a time step Δt the particle moves a distance Δx that can be either to the left or to the right, with equal probability.



At time Δt the particle, that is now at either Δx or $-\Delta x$, decides to move again a distance Δx or $-\Delta x$ so at time $2\Delta t$ the situation is now the following:

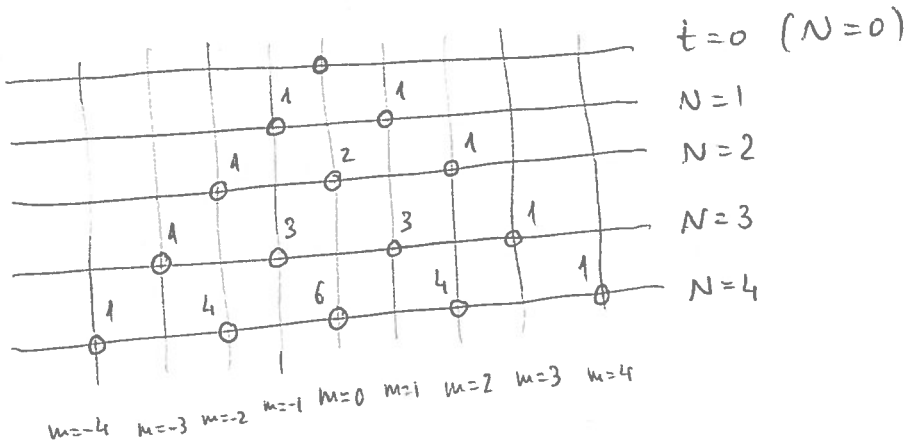


And so on: after N time steps (i.e., at $t = N\Delta t$) the particle that started at $x=0$ is now somewhere on the real line and

we want to describe the probability that it is at the position $x = m\Delta x$. Let $p(m,N)$ be the probability that a particle initially at $x=0$ be at $x = m\Delta x$ after N time steps, i.e., at time $t = N\Delta t$. Clearly

$$p(m,N) = \frac{\text{number of paths from } (0,0) \text{ to } (m\Delta x, N\Delta t)}{\text{total number of paths from } t=0 \text{ to } t=N\Delta t},$$

and so it is instructive to repeat our previous schematic evolution of the positions of the particle including now the number of paths leading to each location:



We infer from here that:

(i) for each time level N, $p(m,N)$ is nonzero for $m = -N, -N+2, \cdots, N-2, N$, and it is zero for all other m

(ii) the number of paths to reach $x = m\Delta x$ at time $t = N\Delta t$ is equal to the sum of the number of paths to reach $x = (m-1)\Delta x$ at $t = (N-1)\Delta t$, with the number of paths to reach $x = (m+1)\Delta x$ at that same time

(iii) the total number of paths from $t=0$ to $t = N\Delta t$ is 2^N

Suppose now that we fix the time interval, $0 \le t \le T$ but let the number of time steps increase, i.e., make Δt smaller.

From what we saw we can write

$$\frac{\text{paths for } (m,N)}{2^N} = \frac{\text{paths for } (m-1,N-1)}{2^N} + \frac{\text{paths for } (m+1,N-1)}{2^N}$$

$$= \frac{1}{2}\,\frac{\text{paths for } (m-1,N-1)}{2^{N-1}} + \frac{1}{2}\,\frac{\text{paths for } (m+1,N-1)}{2^{N-1}}$$

and so

$$\phi(m,N) = \tfrac{1}{2}\phi(m-1,N-1) + \tfrac{1}{2}\phi(m+1,N-1)$$

writing $u(x,t) = \phi(m,N)$, where $x = m\Delta x$ and $t = N\Delta t$, we get

$$u(x,t) = \tfrac{1}{2}u(x-\Delta x, t-\Delta t) + \tfrac{1}{2}u(x+\Delta x, t-\Delta t) \qquad (+)$$

Assume now that $u(\cdot,\cdot)$ is a sufficiently smooth function so that we can use Taylor's series in the right-hand side when Δx and Δt are small. In this case

$$u(x-\Delta x, t-\Delta t) = u(x,t) - u_x(x,t)\Delta x - u_t(x,t)\Delta t + \tfrac{1}{2}\left(u_{xx}(x,t)\Delta x^2 + 2u_{xt}(x,t)\Delta x\Delta t + u_{tt}(x,t)\Delta t^2\right) + \cdots$$

and analogously for the other term.

Then, substituting into (+) we get

$$u(x,t) = u(x,t) - u_t(x,t)\Delta t + \frac{1}{2}u_{xx}(x,t)\Delta x^2 + \frac{1}{2}u_{tt}(x,t)\Delta t^2 + \cdots$$

and thus we have, with all functions u_{xx}, u_t, etc. computed at (x,t),

$$u_t = \frac{(\Delta x)^2}{2\Delta t}u_{xx} + \frac{\Delta t}{2}u_{tt} + \cdots = 0$$

Letting $\Delta t \to 0$ and reducing Δx simultaneously so that $\frac{(\Delta x)^2}{2\Delta t} =: D \neq 0$ (sometimes called the distinguished limit) we obtain the diffusion equation

$$u_t = D u_{xx}.$$

Why is D defined in this way, i.e., as $D = \frac{(\Delta x)^2}{2\Delta t}$, a "diffusion" coefficient? In a certain sense the bigger the value of D is, the larger Δx is in a single time step Δt, and this is something to be expected from a parameter measuring diffusion.

In Physics, in particular in gas dynamics, D is defined in a similar way: if λ is the average distance the gas molecules travel between collisions, and τ is the average time between collisions, then $D = \lambda^2/2\tau$, which is called the Einstein-Smoluchowski equation.

the diffusion, or heat, equation (*) is only the simplest of the equations modelling diffusion of particles or conduction of heat. A similar, and similarly simple equation is obtained in higher dimensions: most relevant from the applications point of view is the diffusion equation in space dimension 2 and 3 where the equation is now

$$(**) \qquad \frac{\partial u}{\partial t} = D \Delta u$$

where $\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}$ is the Laplace operator in n-dim.

Although in higher dimensions, these equations do not differ in any substantial way from the 1-dim case (*).

One case that starts to differ from (*) in a more pronounced way is when the particle motion is not just a random Brownian motion but has also a drift term in response to some external force (like water on a river moving due to the force of gravity.) In these cases the resulting constitutive law has to be modified to accout for the drift: if the velocity due to the drift is v_d then a simple modification is to write the constitutive law

as $J = -D\frac{\partial c}{\partial x} + v_d c$ and the diffusion equation for the evolution of $c(x,t)$ becomes

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - v_d c \tag{$\S$}$$

which is the diffusion - drift equation.

The extra term $-v_d c$ in the right-side of ($\S$), being linear, do not complicate too much its study. But when we have diffusion equations with nonlinear terms the situation can, in some cases, become more complicated and more interesting.

A case when nothing very noticible occurs is when we model the predator-prey in a region Ω where the predators (with concentration $u(x,t)$) and the preys (with concentration $v(x,t)$) can move randomly in Ω. In this case the Lotka-Volterra system is generalized to

$$\begin{cases} \frac{\partial v}{\partial t} = D_1 \frac{\partial^2 v}{\partial x^2} + v(a - bu) \\ \frac{\partial u}{\partial t} = D_2 \frac{\partial^2 u}{\partial x^2} + u(-c + dv) \end{cases}$$

Much more interesting models of systems of diffusion equations are those modelling the apperence of paterns in animal skins due to the diffusion and reaction of chemical species in the cells of embrionic animals (see, for example: H. Meinhardt: "Models of Biological Pattern Formation", Academic Press, 1982; or B. Perthame: "Parabolic Equations in Biology", Springer, 2015.)

Diffusion equations can become substantially more difficult to study if the constitutive law is not linear, which happen, usually, at very high or very low values of $c(x,t)$.

An important such case is when we study the flow of gas in a porous medium. Being $\rho = \rho(x,t)$ the density of the gas, the conservation (of mass) equation is

$$\frac{\partial \rho}{\partial t} + \nabla \cdot J = 0$$

where J is the mass flux. A constitutive law analogous to Fick's or Fourier laws apply in this case and is called Darcy's law:

$$J = -\rho \frac{k}{\mu} \nabla p$$

where $k>0$ is the permeability, μ the gas viscosity, and p is the pressure that driven the gas flow. Comple-

menting Darcy's law with a (constitutive) law relating p and ρ is needed so that we can have a properly set of equations. the simplest such relation is the perfect gas law, $p = \rho R\theta$, where R is a constant and θ is the absolute temperature. So, in an isothermal situation the evolution of the gas density is described by the equation

$$(\#) \qquad \frac{\partial \rho}{\partial t} = \frac{kR\theta}{\mu} \nabla\cdot(\rho \nabla \rho).$$

Note that if instead of $\nabla\cdot(\rho\nabla\rho)$ we would have $\nabla\cdot(\tilde{d}\nabla\rho)$ for some constant $\tilde{d}$ then this equation is equal to $(**)$ with $D = \frac{kR\theta\tilde{d}}{\mu}$, because $\nabla\cdot(\nabla\rho) = \operatorname{div}(\nabla\rho) = \Delta\rho$. So, in $(\#)$ is like having a diffusion equation where the diffusion coefficient is not constant (and, actually, in this case goes to zero as $\rho \to 0$.)

Equation $(\#)$ is an example of a nonlinear diffusion equation, called the <u>porous medium equation</u>, and is extremely important in some applications (see J.L. Vazquez. "the Porous Medium Equation: Mathematical theory", Oxford University Press, 2006)

In this introductory course we will focus our attention

to linear diffusion equations like (*), (**) or (§), mainly on a bounded interval of $\mathbb{R}$ or $\mathbb{R}^n$ (i.e., a segment of straight line or a rectangle).

GMIT

Differential Equations and Applications

Session 10: Linear diffusion equation on a bounded interval: separation of variables and Fourier series.

Let us consider a wire of length L and negligible width which is isolated along its full length but not at the ends where it is kept a a constant ambient temperature $u = 0$.

The wire has been heated such that at the end of the heating process its temperature at the position x is given by a function $f(x)$.

The evolution of the temperature $u = u(x,t)$ of a point x of the wire and at the time t after the instant $t=0$ when the heating was turned off is governed by the following initial and boundary value problem for the heat (or diffusion) equation:

$$(*)_1 \qquad \frac{\partial u}{\partial t} = K \frac{\partial^2 u}{\partial x^2} \quad , \text{ for } (x,t) \in (0,L) \times \mathbb{R}^+,$$

$$(*)_2 \qquad u(x,0) = f(x) \quad , \text{ for } x \in [0,L],$$

$$(*)_3 \qquad u(0,t) = u(L,t) = 0 \, , \text{ for } t \geq 0.$$

We recognize here the heat equation in $(*)_1$, where K is the thermal diffusivity; condition $(*)_2$ is the initial condition stating that the initial (i.e., at $t=0$) temperature is fixed by the datum $f(x)$; and $(*)_3$ is the boundary condition

fixing the temperature at the end points (the boundary) of the wire to be the given ambient temperature zero.

We will try to solve this linear partial differential equation problem using the so called Fourier method, which consists in applying a separation of variables that allow the transformation of the partial differential equation into a set (in this case two) ordinary differential equations and the build of the solution to $(*)_1 - (*)_3$ from the solutions of the ODEs.

Let us write $u(x,t)$ in the form of a product of a function of x with another function of t and let us see if we can determine such functions (that are dependent on a single variable and so, if they satisfy some differential equation, it must be ODEs. So, take

$$u(x,t) = F(x)\,G(t)$$

Substituting this into $(*)_1$ we obtain

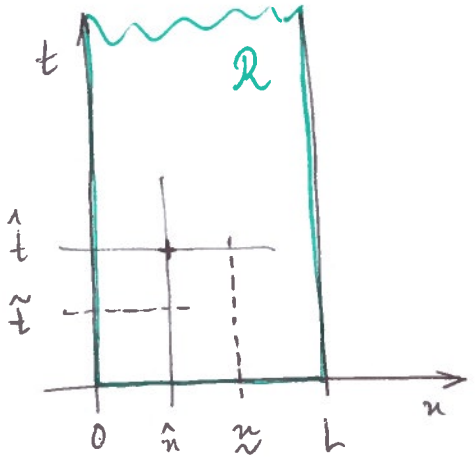
$$F(x)\,G'(t) = K\,F''(x)\,G(t)$$

If F and G are never zero (we'll get to this later...)

then we can divide the last equation by KFG and get

$$(**)_1 \qquad \frac{1}{K}\frac{G'(t)}{G(t)} = \frac{F''(x)}{F(x)} \qquad (x,t) \in (0,L) \times \mathbb{R}^+$$

Now observe that the left-hand side of $(**)_1$ is a function only of t, the right-hand side is a function only of x, and the equality $(**)_1$ has to be satisfied in the open set $\mathcal{R} := (0,L) \times \mathbb{R}^+$.



Pick any point $(\hat{x},\hat{t})$ in $\mathcal{R}$. By $(**)_1$ we know that

$$\frac{1}{K}\frac{G'(\hat{t})}{G(\hat{t})} = \frac{F''(\hat{x})}{F(\hat{x})}$$

Now move t from $\hat{t}$ to any other $\tilde{t}$, keeping $x = \hat{x}$ fixed, then the right-hand side of $(**)_1$ does not change (because it only depends on x) and so

$$\frac{1}{K}\frac{G'(\tilde{t})}{G(\tilde{t})} = \frac{F''(\hat{x})}{F(\hat{x})} = \frac{1}{K}\frac{G'(\hat{t})}{G(\hat{t})}$$

which means that the left-hand side has also kept the same

value. the same happens if we change $\hat{x}$ to $\underset{\sim}{x}$ keeping t fixed. But, then, this means that there must exist a constant σ such that

$$(**)_2 \qquad \frac{1}{k}\frac{G'(t)}{G(t)} = \sigma = \frac{F''(x)}{F(x)} \qquad (x,t) \in \mathcal{R}$$

which are two ordinary differential equations. Observe that the boundary condition $(*)_3$ becomes

$$F(0)G(t) = F(L)G(t) = 0$$

Now, if $G(t) \not\equiv 0$ we need to have $F(0) = F(L) = 0$ and so we get the following <u>boundary value problem</u> for the ordinary differential equation for F:

$$(***) \qquad \begin{cases} F''(x) - \sigma F(x) = 0, & 0 < x < L \\ F(0) = F(L) = 0. \end{cases}$$

Observe that we do not know what σ is apart that it is a constant. Let us try to see what kind of constants are possible.

- suppose $\sigma = 0$; then the differential equation in $(***)$ becomes $F''(x) = 0$ whose solutions are $F(x) = \alpha x + \beta$

and using the boundary condition in (***) becomes

$$\begin{cases} \alpha\cdot 0+\beta = F(0) = 0 \\ \alpha L+\beta = F(L) = 0 \end{cases} \Leftrightarrow \begin{cases} \beta = 0 \\ \alpha = 0 \end{cases}$$

and so the only solution of (***) is $F(x) = 0$, $\forall x$. this gives $u(x,t) = 0$, $\forall (x,t)\in\Omega$, which clearly is a solution of $(*)_1$: an uninteresting one that we could have discovered just by inspection, and furthermore it will not satisfy the initial condition $(*)_3$ unless $f(x)\equiv 0$ (i.e., if we had not heated the wire...)

- So, let us consider now $\sigma\neq 0$, stating with the assumption that $\sigma>0$. In this case we need to solve the differential equation in (***). Transforming the second order equation into a system of first order ODEs we have

$$F''(x) = \sigma F(x) \underset{(F_1=F)}{\Leftrightarrow} \begin{cases} F_1' = F_2 \\ F_2' = +\sigma F_1 \end{cases} \Leftrightarrow \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}' = \underbrace{\begin{pmatrix} 0 & 1 \\ +\sigma & 0 \end{pmatrix}}_{A}\begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$$

the eigenvalues of A are

$$0 = \det(A-\lambda I_2) = \det\begin{pmatrix} -\lambda & 1 \\ -\sigma & -\lambda \end{pmatrix} = \lambda^2-\sigma \Leftrightarrow \lambda_{\pm} = \pm\sqrt{\sigma}$$

and the eigenvectors are $v^{\pm} = (v_1^{\pm}, v_2^{\pm})^T$, solutions of

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mp\sqrt{\sigma} & 1 \\ -\sigma & \mp\sqrt{\sigma} \end{pmatrix}\begin{pmatrix} v_1^{\pm} \\ v_2^{\pm} \end{pmatrix} \Leftrightarrow v^{\pm} = \begin{pmatrix} 1 \\ \pm\sqrt{\sigma} \end{pmatrix}$$

So, the solution of the 1st order system is

$$\begin{pmatrix} F_1 \\ F_2 \end{pmatrix}(x) = \alpha \begin{pmatrix} 1 \\ \sqrt{\sigma} \end{pmatrix} e^{\sqrt{\sigma}x} + \beta \begin{pmatrix} 1 \\ -\sqrt{\sigma} \end{pmatrix} e^{-\sqrt{\sigma}x}$$

and so, recalling that $F = F_1$, the solutions of $F'' = \sigma F$ are

$$F(x) = \alpha e^{\sqrt{\sigma}x} + \beta e^{-\sqrt{\sigma}x}$$

Using the boundary conditions for F we get

$$\begin{cases} 0 = F(0) = \alpha + \beta \\ 0 = F(L) = \alpha e^{\sqrt{\sigma}L} + \beta e^{-\sqrt{\sigma}L} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{pmatrix} 1 & 1 \\ e^{\sqrt{\sigma}L} & e^{-\sqrt{\sigma}L} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

because $\det \begin{pmatrix} 1 & 1 \\ e^{\sqrt{\sigma}L} & e^{-\sqrt{\sigma}L} \end{pmatrix} \neq 0$

$$\Leftrightarrow \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ e^{\sqrt{\sigma}L} & e^{-\sqrt{\sigma}L} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So again in the case $\sigma > 0$ the only solutions to (***) is the trivial solution $F(x) \equiv 0$.

- It remains to see what (or if) $\sigma < 0$ gets us.

  Let us write $\sigma = -\mu^2$ with $\mu > 0$.

  From the computations in the last case we have that the matrix A has now two complex conjugate eigenvectors

$$\lambda_{\pm} = \pm\sqrt{\sigma} = \pm\sqrt{-\mu^2} = \pm\mu i = 0 \pm \mu i$$

the corresponding eigenvectors are

$$v^{\pm} = \begin{pmatrix} 1 \\ \pm\mu i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \pm i \begin{pmatrix} 0 \\ \mu \end{pmatrix}$$

and thus, from what we saw in session 5, the real valued solutions of $F'' = -\mu^2 F$ are

$$\begin{aligned} f(x) &= 2e^{0x}\,\mathrm{Re}\left((\alpha+i\beta)(\cos\mu x + i\sin\mu x)\cdot(1+i0)\right) \\ &= 2\alpha\cos\mu x - 2\beta\sin\mu x \end{aligned}$$

Using the boundary conditions for F we get

$$\begin{cases} 0 = f(0) = 2\alpha \\ 0 = f(L) = 2\alpha\cos\mu L - 2\beta\sin\mu L \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \alpha = 0 \\ \beta\sin\mu L = 0 \end{cases}$$

Thus, we still have $(\alpha,\beta) = (0,0)$ as a solution, but additionally we can solve the second equation in the last system by choosing $\mu > 0$ such that $\sin\mu L = 0$, that is, by taking $\mu L = k\pi$ with $k \in \mathbb{N}^+$, which means

$$\mu = \mu_k := \frac{k\pi}{L}, \quad k \in \mathbb{N}^+.$$

So, non-identically zero solutions of (***) are

$$F_k(x) = \sin \frac{k\pi}{L} x \,, \quad k \in \mathbb{N}^+.$$

Note that, because the differential equation in (***) is linear, linear combinations of functions F_k, for different values of k, are also solutions of (***).

Now let us consider the equation for G in $(**)_2$. We already know that $\sigma = -\left(\frac{k\pi}{L}\right)^2$ with $k \in \mathbb{N}^+$, so the equation is

$$\frac{1}{k} \frac{G'(t)}{G(t)} = -\left(\frac{k\pi}{L}\right)^2$$

which we recognize as the first order linear ODE

$$G' = -k\left(\frac{k\pi}{L}\right)^2 G$$

whose solutions are functions multiple of

$$G(t) = e^{-k\left(\frac{k\pi}{L}\right)^2 t}.$$

From what was done up to this point we conclude that the boundary value problem $(*)_1$, $(*)_3$ has a solutions finite linear combinations of

$$e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin \frac{k\pi}{L} x \,,$$

that is

$$(+) \qquad u(x,t) = \sum_{k=1}^{N} a_k e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L}x\right)$$

for any fixed $N \in \mathbb{N}^+$ and constants $a_k \in \mathbb{R}$.

However, our problem also included an initial condition $u(x,0) = f(x)$, where f was a given function. If this function is a linear combination of functions $\sin \frac{k\pi}{L}x$ then all is well and we can determine the parameters in $(+)$ that guarantees the fitting $u(x,0) = f(x)$. For example, if $f(x) = 2\,\text{sen}\,\frac{\pi x}{L} + \frac{1}{4}\sin\frac{5\pi x}{L}$, then the initial condition becomes

$$2\,\text{sen}\,\frac{\pi x}{L} + \frac{1}{4}\sin\frac{5\pi x}{L} = \sum_{k=1}^{N} a_k e^{-0} \sin\frac{k\pi x}{L}$$

which means that $a_1 = 2$, $a_5 = \frac{1}{4}$, and all other a_k are zero giving the solution

$$u(x,t) = 2e^{-k\left(\frac{\pi}{L}\right)^2 t} \sin\frac{\pi x}{L} + \frac{1}{4} e^{-k\left(\frac{5\pi}{L}\right)^2 t} \sin\frac{5\pi x}{L}.$$

Clearly these functions f, linear combinations of a <u>finite</u> number of functions $\sin\frac{k\pi}{L}x$, are not of interest for applications, nor for the theory. So a natural idea is the

following: can a function f, with reasonable characteristics, be written as a linear combination of an <u>infinite</u> number of functions $\sin\frac{k\pi}{L}x$? If it can, i.e., if we can write

$$(\neq) \qquad f(x) = \sum_{k=1}^{\infty} c_k \sin\frac{k\pi}{L}x \quad , \quad 0 \le x \le L$$

then, by $u(x,0) = f(x)$, we could get a "candidate" to a solution of $(*)_1 - (*)_3$ as

$$(\neq\!\!\neq) \qquad u(x,t) = \sum_{k=1}^{\infty} c_k e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L}x\right)$$

with the same constants c_k used to express f as a series of sines in $(\neq)$.

We call the function $u(x,t)$ defined by $(\neq\!\!\neq)$ a "candidate" for a solution because although each term of the series is a differentiable function and satisfy $(*)_1$ [and $(*)_3$] it is not obvious a priori that the series in $(\neq\!\!\neq)$ is <u>convergent</u> and neither that, if it is convergent, its sum is a <u>differentiable</u> function. We will call such an expression a <u>formal solution</u> of $(*)_1 - (*)_3$, meaning that it would be indeed a solutions if these problems of convergence and differentiability were successfully tackled.

the strategy presented in the last page, first explicitly considered by Fourier (without the worries about convergence and differentiability), in the early 19th Century, depends on the possibility of writing a function f as a series (#), of sines, or of cosines. These are called trigonometric series and if the coefficients are computed in a certain way (to be presented shortly) the series are called Fourier series.

The study of Fourier series, and of trigonometric series in general, is part of Harmonic Analysis, an extremely active and useful part of Mathematical Analysis with applications in very many other parts of mathematics (from differential equations to number theory, and even to statistics) and in other sciences and engineering (signal processing, remote sensing, etc.)

In what follows we will just present the very basic notions and results of Fourier series for their application in the context of the linear diffusion equation (and, later, of the linear wave equation.)

GMIT

Differential Equations and Applications

Session 11: Fourier series: definition, convergence theorems, examples.

We shall now make a very brief visit to the topic of Fourier series centering our attention in aspects that are relevant for their use in the study of partial differential equations.

We are interested in knowing which functions $f : \mathbb{R} \to \mathbb{R}$ can be written as

$$(*)_1 \qquad f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} \left(a_k \cos \frac{k\pi x}{L} + b_k \sin \frac{k\pi x}{L} \right).$$

[the reason of writing $\frac{1}{2}a_0$ instead of a_0 will become clear soon.]

Clearly, if the equality $(*)_1$ holds then the coefficients a_k and b_k will be related to f and one of the first things to look at is what is that relation. Before this, however, let us notice that if the equality $(*)_1$ is to hold then, because the functions in the right-hand side are periodic, the function f will also be periodic.

Let us briefly recall this concept.

A function $f : \mathbb{R} \to \mathbb{R}$ is periodic with period T, or T-periodic, if $\forall x \in \mathbb{R}$, $f(x+T) = f(x)$. We say that T is the minimal period, or <u>fundamental period</u>, if T is the smallest positive number for which $f(x+T) = f(x)$ holds.

If T is a fundamental period of f, then $2T, 3T, \dots$ are also periods (but not fundamental periods) of f.

Just to illustrate the concept consider the function

$$x \longmapsto \sin \frac{k\pi x}{L}.$$

For the fundamental period T of this function we have

$$\sin \frac{k\pi(x+T)}{L} = \sin \frac{k\pi x}{L} \quad , \quad \forall x \in \mathbb{R}$$

and T is the smallest positive value for which this holds. Then, for all $x \in \mathbb{R}$ it must be true the equality

$$\sin \frac{k\pi x}{L} \cos \frac{k\pi T}{L} + \cos \frac{k\pi x}{L} \sin \frac{k\pi T}{L} = \sin \frac{k\pi x}{L}.$$

In particular, making $x = \frac{L}{2k}$ we get

$$\sin \frac{\pi}{2} \cos \frac{k\pi T}{L} + \cos \frac{\pi}{2} \sin \frac{k\pi T}{L} = \sin \frac{\pi}{2} \iff$$

$$\iff \cos \frac{k\pi T}{L} = 1$$

and the smallest positive T for which this holds is

$$\frac{k\pi T}{L} = 2\pi \iff T = T_k := \frac{2L}{k}.$$

Getting back to the series $(*)_1$, the above result means that the period of the function in the right-hand side (if the series is convergent) is $2L$, which is the largest fundamental period of the terms in the series. (We did not study the fundamental period of $x \mapsto \cos\frac{k\pi x}{L}$, but it is exactly equal to what was done above for the sine.)

So, in an expression like $(*)_1$, the function $f : \mathbb{R} \to \mathbb{R}$ must be $2L$-periodic.

Suppose the series in the right-hand side is uniformly convergent and that the equality in $(*)_1$ is satisfied for all $x \in \mathbb{R}$.

The uniform convergence of the series and the continuity of the functions sine and cosine imply that the function f given by $(*)_1$ is also continuous [we later recall the theorem that guarantees this.] hence it is integrable in $[-L, L]$. So, let us integrate $(*)_1$ in $[-L, L]$ and use the uniform convergence of the series to change $\int_{-L}^{L} \sum_{k=1}^{\infty}$ to $\sum_{k=1}^{\infty} \int_{-L}^{L}$ in the right-hand side:

$$\int_{-L}^{L} f(x)\,dx = \frac{1}{2}a_0\int_{-L}^{L} dx + \int_{-L}^{L}\sum_{k=1}^{\infty}\left(a_k \cos\frac{k\pi x}{L} + b_k \sin\frac{k\pi x}{L}\right)dx$$

$$= a_0 L + \sum_{k=1}^{\infty}\left(a_k\int_{-L}^{L}\cos\frac{k\pi x}{L}\,dx + b_k\int_{-L}^{L}\sin\frac{k\pi x}{L}\,dx\right)$$

But we know that

$$\int_{-L}^{L}\cos\frac{k\pi x}{L}\,dx = 0 = \int_{-L}^{L}\sin\frac{k\pi x}{L}\,dx$$

and so the equality above reduces to the following formula for a_0:

$$a_0 = \frac{1}{L}\int_{-L}^{L} f(x)\,dx.$$

This simple computation gives us the idea of how to compute the other coefficients a_k and b_k once we remember that sines and cosines satisfy the following so called <u>orthogonality relations</u>:

$$\int_{-L}^{L}\cos\frac{k\pi x}{L}\sin\frac{\hat{k}\pi x}{L}\,dx = 0 \quad , \quad \forall k,\hat{k}\geq 1$$

$$\int_{-L}^{L}\cos\frac{k\pi x}{L}\cos\frac{\hat{k}\pi x}{L}\,dx = \int_{-L}^{L}\sin\frac{k\pi x}{L}\sin\frac{\hat{k}\pi x}{L}\,dx = \begin{cases} L & \text{if } k=\hat{k} \\ 0 & \text{if } k\neq\hat{k}. \end{cases}$$

Multiply $(*)_1$ by $\cos \frac{n\pi x}{L}$ and integrate the result in $[-L, L]$. Using the orthogonality relations we get

$$\int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} \, dx = a_n L$$

and if we multiply $(*)_1$ by $\sin \frac{n\pi x}{L}$ and do the same

$$\int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} \, dx = b_n L$$

these are the same as

$$(*)_2 \qquad a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} \, dx \quad , \quad n \geq 0 ,$$

$$(*)_3 \qquad b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} \, dx \quad , \quad n \geq 1 .$$

The expressions $(*)_2$ and $(*)_3$ are called the <u>Fourier coefficients</u> of f, and the trigonometric series $(*)_1$ with a_k and b_k given by $(*)_2$ and $(*)_3$, respectively, is called the <u>Fourier series</u> of f.

So, given a function $f: \mathbb{R} \to \mathbb{R}$, $2L$-periodic, integrable and absolutely integrable in $[-L, L]$, we define the <u>Fourier series</u> of f by the right-hand side of $(*)_1$ with the coefficients given by $(*)_2$ and $(*)_3$. This is represented by notation

$$(*)_4 \qquad f(x) \sim \frac{1}{2}a_0 + \sum_{k=1}^{\infty} \left(a_k \cos \frac{k\pi x}{L} + b_k \sin \frac{k\pi x}{L} \right).$$

What we would like is that the relation between the function f and its Fourier series (loosely denoted by $\sim$ above) to be an equality. Unfortunately this is not always the case: there are examples of continuous functions f for which the Fourier series diverges (and so $\sim$ is not an equality since the right-hand side does not make sense!). So, to get a reasonable relation between f and its Fourier series we need to make stronger assumptions about the regularity of the function f. One of these is that f is piecewise differentiable. Let us introduce this concept now:

A function f is <u>piecewise continuous</u> if in every bounded interval $[a,b]$ there are points $a_1 < \cdots < a_n$ such

that f is continuous in (a_j, a_{j+1}), $j = 1, \dots, n-1$, and the limits

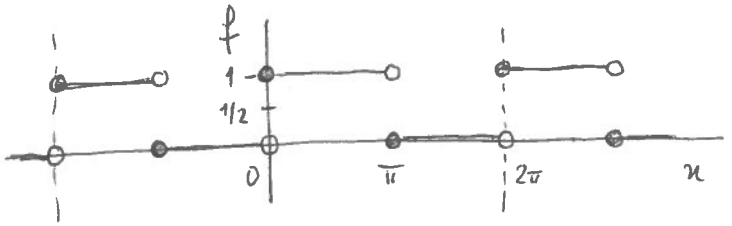
$$f(a_j^+) := \lim_{x \to a_j^+} f(x) \quad , \quad f(a_j^-) := \lim_{x \to a_j^-} f(x)$$

exist and are finite.

A function is <u>piecewise differentiable</u> if it is piecewise continuous and its derivative is also piecewise continuous.

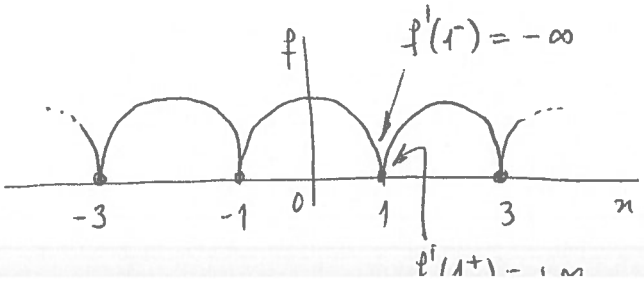
<u>Example 1</u>

the function $f(x) = \begin{cases} 1, \text{ if } 0 \le x < \pi \\ 0, \text{ if } \pi \le x < 2\pi \\ 2\pi\text{-periodic} \end{cases}$ is piecewise differentiable.



<u>Example 2</u>

the function $f(x) = \begin{cases} \sqrt{1-x^2}\,, \text{ if } |x| \le 1 \\ 2\text{-periodic} \end{cases}$ is <u>not</u> piecewise differentiable: it is continuous, but the limits of its derivative at the points $2k+1$, $k \in \mathbb{Z}$, are not finite.



the notion of piecewise differentiability is the required concept for the following theorem:

Theorem 1 (Fourier) (pointwise convergence)

Let $f : \mathbb{R} \to \mathbb{R}$ be a $2L$-periodic and piecewise differentiable function.

then, the Fourier series $(*)_4$ is (pointwise) convergent in each x and the following equality holds:

$$\frac{f(x^+) + f(x^-)}{2} = \frac{1}{2}a_0 + \sum_{k=1}^{\infty}\left(a_k \cos\frac{k\pi x}{L} + b_k \sin\frac{k\pi x}{L}\right).$$

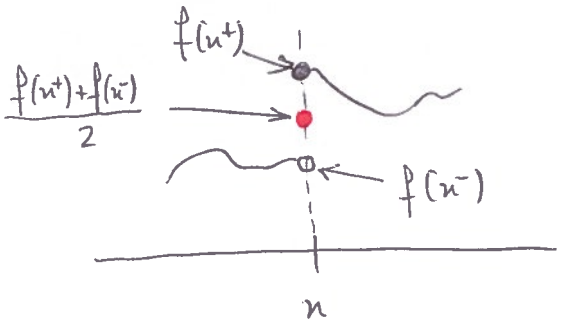
Obs.: Note that the quantity $\frac{f(x^+) + f(x^-)}{2}$ is equal to $f(x)$ if f is continuous at the point x: in fact, in that case

$$f(x^\pm) := \lim_{\hat{x} \to x^\pm} f(\hat{x}) = f(x)$$

and thus

$$\frac{f(x^+) + f(x^-)}{2} = \frac{2f(x)}{2} = f(x)$$

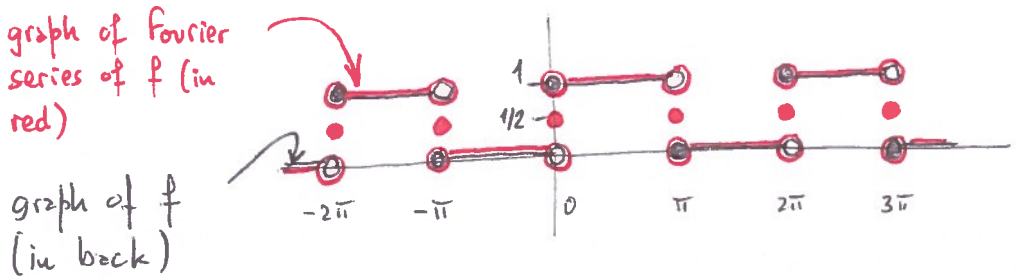
If f is discontinuous at x then $\frac{f(x^+) + f(x^-)}{2}$ is the mean of the values of the limits from the right and from the left at x which, graphically, is something like



Hence, fourier theorem tells us that if f is piecewise differencible and continuous the sum of its Fourier series is equal to $f(x)$, for every x.

If f is piecewise differentiable but has discontinuities, then the sum of its fourier series always exist but at those x where the function f is discontinuous its value is not $f(x)$: it is $\frac{f(x^+)+f(x^-)}{2}$.

This means that we can use Fourier theorem to plot the sum of a Fourier series almost without computations. For example: since the function in Example 1 is piecewise differentiable we know that its fourier series is convergent at every point $x \in \mathbb{R}$ and its sum is a function of x whose graph is represented in red in the following plot:



Note that the convergence guaranteed by Fourier's theorem is pointwise convergence and the pointwise convergence of a sequence of continuous functions need not be continuous, as the illustration with the function of Example 1 clearly shows.

Given that we want to use Fourier series to represent solutions to partial differential equations we are interested that they are not only continuous but even differentiable (even twice differentiable!). One way to have this regularity result is to have uniform convergence of the Fourier series. In fact the following theorems hold:

Theorem 2 (1st theorem on uniform convergence)

Let $f: \mathbb{R} \to \mathbb{R}$ be $2L$-periodic, continuous, with f' and $(f')^2$ integrable in $[-L, L]$.

then the Fourier series of f is uniformly convergent to f.

theorem 3 (2nd theorem on uniform convergence)

Let $f: \mathbb{R} \to \mathbb{R}$ be $2L$-periodic, piecewise continuous with f' and $|f'|$ integrable in $[-L, L]$.

Then the Fourier series of f is uniformly convergent to f in every closed interval that do not contain discontinuous points of f.

We will not prove theorems 1-3 (although we will write some remarks about the proofs later on). Now we will use them and $(*)_1$-$(*)_3$ to compute some Fourier series.

Example 3

Let us pick Example 1 again: we want to compute the Fourier series of $f(x) = \begin{cases} 1 & \text{if } 0 \le x < \pi \\ 0 & \text{if } \pi \le x < 2\pi \\ 2\pi\text{-periodic} \end{cases}$.

From $(*)_2$ and $(*)_3$ we have $a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f = \frac{1}{\pi}\int_0^{\pi} 1 = 1$,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx = \frac{1}{\pi}\int_0^{\pi} \cos nx\,dx = \frac{1}{\pi}\frac{\sin nx}{n}\Big|_0^{\pi} = 0,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx = \frac{1}{\pi}\int_0^{\pi} \sin nx\,dx = \frac{1}{\pi}\frac{-\cos nx}{n}\Big|_0^{\pi} =$$

$$= \frac{1}{n\pi}(1-\cos n\pi) = \frac{1}{n\pi}(1-(-1)^n)$$

which means that $b_{2k} = 0$ and $b_{2k-1} = \frac{2}{(2k-1)\pi}$, $k \in \mathbb{N}^+$

Since f is piecewise differentiable, Fourier's theorem ensure us that

$$\frac{f(x^+)+f(x^-)}{2} = \frac{1}{2} + \sum_{k=1}^{\infty} \frac{2}{(2k-1)\pi}\sin(2k-1)x$$

and the function in the left-hand side is

$$\frac{f(x^+)+f(x^-)}{2} = \begin{cases} 1 & \text{if } 0 < x < \pi \\ 0 & \text{if } \pi < x < 2\pi \\ 1/2 & \text{if } x = k\pi,\ k \in \mathbb{Z} \\ 2\pi\text{-periodic} \end{cases},$$

and the series is uniformly convergent in all intervals that do not contain the points of discontinuity of f, $x = k\pi$, $k \in \mathbb{Z}$.

It is interesting to observe that at $x = \frac{\pi}{2}$, a point where f is continuous and has value 1, we have

$$1 = \frac{1}{2} + \sum_{k=1}^{\infty} \frac{2}{(2k-1)\pi} \sin(2k-1)\frac{\pi}{2}$$

which can be rewritten as

$$\frac{\pi}{4} = \sum_{k=1}^{\infty} \frac{1}{2k-1} \sin(2k-1)\frac{\pi}{2}$$

$$= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2k-1}.$$

In computing Fourier series of functions it is usually convenient to check beforehand if the function is even or odd (with the eventual exception of a finite number of points). Recalling that a function is even if $f(x) = f(-x)$ for all $x \in \mathbb{R}$, and is odd if $f(x) = -f(-x)$, we observe that

- the sum of two odd functions is an odd function
- the sum of two even functions is an even function
- the product of two odd functions, or two even functions is an even function
- the product of two functions, one odd and another even is an odd function

As $x \mapsto \sin \frac{k\pi}{L} x$ is odd and $x \mapsto \cos \frac{k\pi}{L} x$ is even, the results just stated imply that if f is odd its Fourier series is a series of sines (i.e., all $a_k = 0$), and if f is even the corresponding fourier series is a series of cosines (i.e., all $b_k = 0$). This can be checked easily because, when f is odd then $f(x) \sin \frac{k\pi x}{L}$ is even and $f(x) \cos \frac{k\pi x}{L}$ is odd and we know that if a function g is even then

$$\int_{-L}^{L} g(x)\,dx = \int_{-L}^{0} g(x)\,dx + \int_{0}^{L} g(x)\,dx$$

$$= \int_{+L}^{0} g(-x)\,d(-x) + \int_{0}^{L} g(x)\,dx$$

$$= \int_{0}^{L} g(-x)\,dx + \int_{0}^{L} g(x)\,dx \qquad (g \text{ even})$$

$$= 2\int_{0}^{L} g(x)\,dx$$

and if g is odd

$$\int_{-L}^{L} g(x)\,dx = \cdots = \int_{0}^{L} g(-x)\,dx + \int_{0}^{L} g(x)\,dx$$

$$= -\int_{0}^{L} g(x)\,dx + \int_{0}^{L} g(x)\,dx \qquad (g \text{ odd})$$

$$= 0$$

So, if f is even (2L-periodic, integrable and absolutely integrable) we have

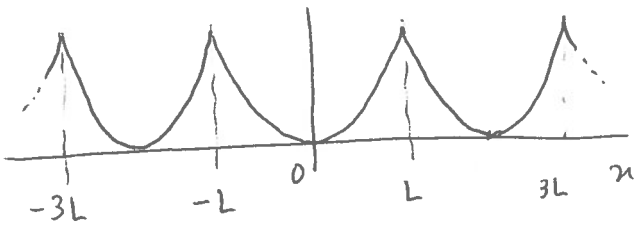
$$a_k = \frac{2}{L}\int_0^L f(x)\cos\frac{k\pi x}{L}\,dx \quad , \quad b_k = 0,$$

and if f is odd

$$a_k = 0 \quad , \quad b_k = \frac{2}{L}\int_0^L f(x)\sin\frac{k\pi x}{L}\,dx.$$

Example 4

Consider the function $f:\mathbb{R}\to\mathbb{R}$, 2L-periodic and equal to x^2 for $x\in[-L,L]$,



which is continuous (and hence integrable and absolutely integrable). It is also, obviously, an even function in $[-L,L]$ because $(-x)^2 = x^2$, and so, by periodicity, it is an even function in $\mathbb{R}$.

Its Fourier coefficients are given by

$$a_k = \frac{2}{L}\int_0^L x^2 \cos\frac{k\pi x}{L}\,dx \quad , \quad b_k = 0$$

and so

$$a_0 = \frac{2}{L}\int_0^L x^2\,dx = \frac{2}{L}\frac{L^3}{3} = \frac{2L^2}{3}$$

and for a_k we change variables $x \mapsto y = \frac{k\pi x}{L}$ to get

$$a_k = \frac{2L^2}{k^3\pi^3}\int_0^{k\pi} y^2 \cos y\,dy$$

and integrating by parts twice we get

$$a_k = \frac{4L^2}{k^2\pi^2}(-1)^k$$

from where we get the Fourier series for f

$$f(x) = \frac{L^2}{3} + \frac{4L^2}{\pi^2}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\cos\frac{k\pi x}{L} \tag{x}$$

and the equality is guaranteed by Fourier's theorem because f is continuous and piecewise differentiable.

As a curious note, observe that taking $x = L$ in (x) we get

$$L^2 = \frac{L^2}{3} + \frac{4L^2}{\pi^2}\sum_{k=1}^{\infty}\frac{(-1)^k}{k^2}\cos k\pi$$

which, noting that $\cos k\pi = (-1)^k$ and simplifying, gives

$$\frac{\pi^2}{6} = \sum_{k=1}^{\infty} \frac{1}{k^2} .$$

Recall that in the context of application to the diffusion equation in a bounded interval $[0,L]$ the function we are interested in expand in Fourier series, i.e., the initial temperature (or particles, or individuals) distribution $f(x)$ is given only in the interval $[0,L]$, which is the only region where it makes Physical (or Chemical, or Ecological, ...) sense.

Because to have a representation of f in a Fourier series we need to have f periodic, we need to extend f from $[0,L]$ to the whole real line in a periodic way and there are many different ways to do this. The choice of the extension to use will depend on the kind of Fourier series we want to get: for example, in the case we used as an illustration in last session we needed to have f written in terms of a series of sines, so, for that case, we will need to consider an extension of f to $\mathbb{R}$ as an odd function.

We exemplify this issue next

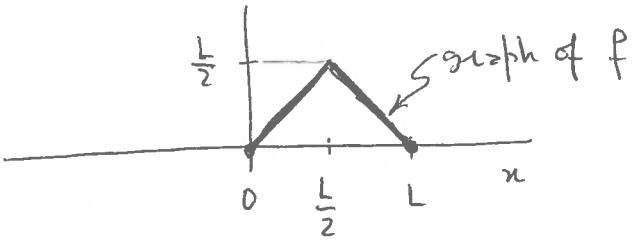
Example 5

Consider the function $f:[0,L]\longrightarrow \mathbb{R}$ defined by $f(x)=\frac{L}{2}-\left|x-\frac{L}{2}\right|$.

We are interested to get an expansion of f in Fourier series of sines.

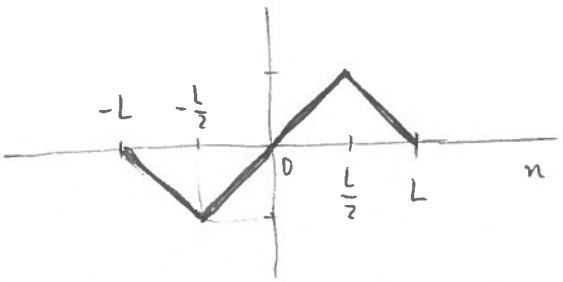
Starting by observing that f is continuous and piecewise differentiable in $[0,L]$ we need to first extend it to a function $\tilde{f}$ defined on $\mathbb{R}$ and odd, so that its Fourier series will be a series of sines. As the graph of f is

(1)



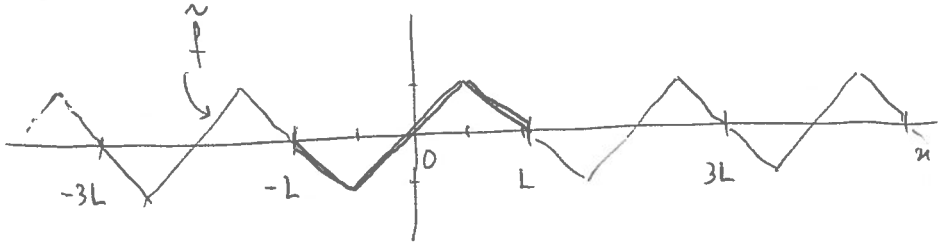
we need first to extend it as an odd function to $[-L,L]$

(2)



and now need to extend to $\mathbb{R}$ as a periodic function. If we don't have any more information about the period of the extension we can take it as $2L$ and just repeat what was obtained in $[-L,L]$:

(3)

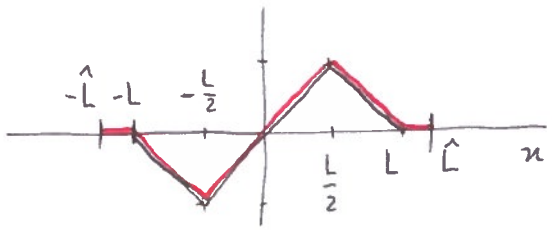


Clearly $\tilde{f}$ coincides with f on $[0,L]$ and to compute the Fourier coefficients b_k we need only to use what happens in this interval (the a_k are known to be zero because we are extending f as an odd function)

Observe that if we wanted the period of the Fourier expansion to be $2\hat{L}$ with $\hat{L} > L$ then we would need to extend f from $[0,L]$ to $\mathbb{R}$ as an odd $2\hat{L}$-periodic function. From f in $[0,L]$ to an odd function in $[-L,L]$ we have already done it in (1) $\rightsquigarrow$ (2), but now we need to have this expanded (as an odd function) to $[-\hat{L},\hat{L}]$ before extending periodically to $\mathbb{R}$; again there are many possible ways to do this, the

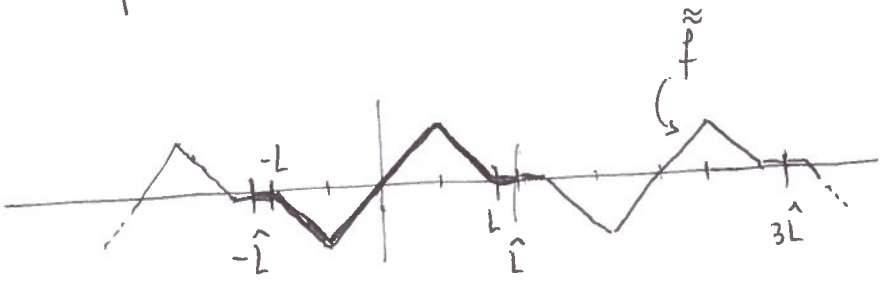
simplest is to put the function equal to zero in $[-\hat{L}, \hat{L}] \setminus [-L, L]$

(2')



and now extend to the whole line $\mathbb{R}$ as a $2\hat{L}$-periodic function

(3')



Note that $\tilde{\tilde{f}}$ coincides with f in $[0, L]$ and is zero in $(L, \hat{L}]$. Thus, the Fourier coefficients are given by

$$b_k = \frac{2}{\hat{L}} \int_0^{\hat{L}} \tilde{\tilde{f}}(x) \sin \frac{k\pi x}{\hat{L}} \, dx$$

$$= \frac{2}{\hat{L}} \int_0^{L} f(x) \sin \frac{k\pi x}{\hat{L}} \, dx$$

(the upper integration limit, L, is not equal to the half-period of the function $\hat{L}$)

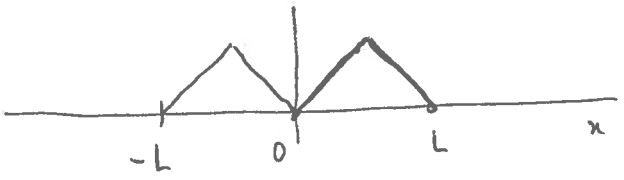
So, the series will be different from the

previous one (as it should: the periods are different!) but, by Fourier's theorem, we know that its sum is equal to the previous sum in $[0,L]$ and both are equal to the given function f.

It is left as an exercise to compute the Fourier series of $f(x) = \frac{L}{2} - \left|x - \frac{L}{2}\right|$ for a series of sines with period $2L$ and for a series of sines with period $2L+1$ (i.e., with period $2\hat{L}$ with $\hat{L} = L + \frac{1}{2}$).

Clearly, if we want a series of cosines for f we need first to extend f from $[0,L]$ to $[-L,L]$ as an even function

(2")



and then extend it periodically to $\mathbb{R}$.

GMIT

Differential Equations and Applications

Session 12 : Diffusion equation on bounded intervals : its solution using Fourier series.

Let us go back to the initial and boundary value problem for the heat equation we considered in session 10. We studied

$$(*)_1 \qquad \frac{\partial u}{\partial t} = K \frac{\partial^2 u}{\partial x^2} \quad , \qquad \text{for } (x,t) \in R := (0,L) \times \mathbb{R}^+$$

$$(*)_2 \qquad u(x,0) = f(x) \quad , \qquad \text{for } 0 \le x \le L$$

$$(*)_3 \qquad u(0,t) = u(L,t) = 0 \quad , \qquad \text{for } t \ge 0$$

using Fourier's method, i.e., by separation of variables and writing the solution using a Fourier series.

Let us now make a concrete assumption about the initial temperature distribution f. Assume:

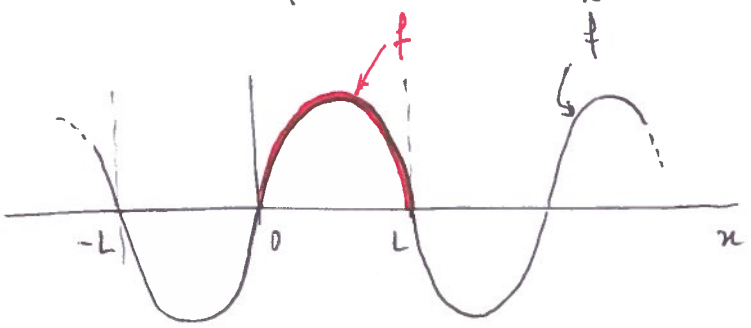
$$f(x) = x(L-x) \quad , \qquad \text{for } x \in [0,L].$$

We saw in session 10 that a formal solution to $(*)_1 - (*)_3$ is given by

$$(**) \qquad u(x,t) = \sum_{k=1}^{\infty} c_k \, e^{-K\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L} x\right)$$

where c_k are the Fourier coefficients of the expansion of f in a Fourier series of sines with period $2L$.

In this situation, f must be extended to $\mathbb{R}$ as a $2L$-periodic even function; call it $\tilde{f}$. Of course $\tilde{f} = f$ on $[0,L]$:



So, the Fourier coefficients are

$$c_k = \frac{2}{L}\int_0^L \tilde{f}(x)\sin\frac{k\pi x}{L}\,dx$$

$$= \frac{2}{L}\int_0^L f(x)\sin\frac{k\pi x}{L}\,dx$$

$$= \frac{2}{L}\int_0^L x(L-x)\sin\frac{k\pi x}{L}\,dx$$

$$= 2\int_0^L x\sin\frac{k\pi x}{L}\,dx - \frac{2}{L}\int_0^L x^2\sin\frac{k\pi x}{L}\,dx$$

↙ using integration by parts

$$= \frac{4L^2}{k\pi}(-1)^{k-1} + \frac{4L^2}{k^3\pi^3}\left(1-(-1)^k\right)$$

Plugging this into (**) we have the expression for the formal solution; but is it really a solution? What is natural do have so we can call a function $u(x,t)$ a solution?

Well, if $u(x,t)$ is to be a solution of $(*)_1 - (*)_3$ then it must be such that the derivatives in $(*)_1$ can be computed, the equalities are all satisfied, and what happens to $u(x,t)$ at the boundary $\{(0,t)\}$, $\{(L,t)\}$ and initially at $\{(x,0)\}$

glue together in a nice way. To be more precise we consider the following

Definition (classical solution 1)

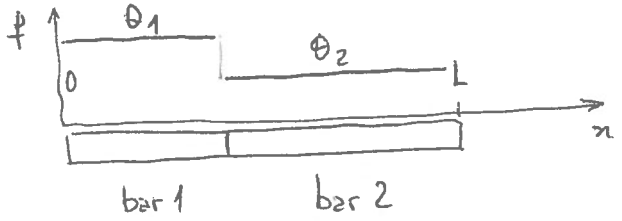
A function $u : \overline{R} = [0,L] \times [0,+\infty) \longrightarrow \mathbb{R}$ is a solution to the initial and boundary value problem $(*)_1 - (*)_3$ if

(i) $u \in C^0(\overline{R})$, i.e., u is continuous in $\overline{R}$

(ii) $\frac{\partial u}{\partial t}$ and $\frac{\partial^2 u}{\partial x^2}$ are defined in R

(iii) the three conditions $(*)_1 - (*)_3$ are satisfied

In some cases it is convenient to relax a bit the continuity assumption at $t=0$. For example, suppose we have two bars at temperatures θ_1, θ_2 with $\theta_1 \neq \theta_2$ and initially put them into contact with each other, then the initial condition will be



So f is a discontinuous function and it is not reasonable to impose that a solution must be continuous in $\overline{R}$, i.e., back to $t=0$!

To overcome this difficulty we consider a weaker kind of continuity at $t=0$, some kind of continuity

in the mean:

Definition (classical solution 2)

A function $u: \widehat{\Omega} = [0,L] \times (0,+\infty) \longrightarrow \mathbb{R}$ is a solution to the initial and boundary value problem $(*)_1 - (*)_3$ if the conditions in the previous definition are fulfilled with $(*)_2$ changed to

$$\lim_{t \to 0} \int_0^L u(x,t)\varphi(x)\,dx = \int_0^L f(x)\varphi(x)\,dx$$

for every piecewise continuous function $\varphi: [0,L] \to \mathbb{R}$.

This is something very common in partial differential equations: to weaken an a priori reasonable definition of solution to allow for a rigorous study of important situations that could not be tackled with the first, naive, definition. Checking the mathematical literature on PDEs we find many studies of "weak solutions", which have a clear technical meaning that must be checked in order to make sense of what is being really studied. They are usually solutions in some integrated sense, as in the case stated in the above definition or similar cases. We shall not proceed in this direction in this introductory course.

Let us return to the problem of checking if a formal solution is indeed a classical solution. We have $u(x,t)$ written as a Fourier expansion (like in the case above)

$$u(x,t) = \sum_{k=1}^{\infty} \underbrace{c_k e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L}x\right)}_{=: u_k(x,t)}.$$

Each term $u_k(x,t)$ in this series is a $C^\infty_{t,x}$-function, i.e., it has derivatives of all orders relative to t and x. So, in particular $u_k \in C^1_t \cap C^2_x$, and in particular $u_k \in C^0$. But a convergent series of continuous functions is not necessarily continuous. So we cannot automatically conclude the continuity of $u(x,t)$, and then neither the needed differentiability. The easiest way to guarantee that $u(x,t)$ is continuous is by checking if the series is not only (pointwise) convergent, but actually <u>uniform convergent</u>.

In fact, we have the following important result:

R1 | If $u_n(x,t)$ are continuous functions
If $\sum_n u_n(x,t)$ is uniformly convergent
then the sum $u(x,t) = \sum_{n=1}^{\infty} u_n(x,t)$ is a continuous function

Checking that a series is uniformly convergent can be hard. In many cases the following "Weierstrass test" is very useful:

> Let $u_n : I \to \mathbb{R}$, defined in an interval $I \subset \mathbb{R}$, be such that $|u_n(x,t)| \le M_n$, $\forall (x,t), n$, for some constants M_n
> Suppose the series of constants $\sum_n M_n$ is convergent
> then, the series of functions $\sum_n u_n(x,t)$ is absolutely an uniformly convergent in the interval I.

Note that this result is just a sufficient condition, but when it is applicable it is very useful.

For example, suppose we have obtained a formal solution given by

$$(X) \qquad u(x,t) = \sum_{k=1}^{\infty} \underbrace{\frac{A}{k^4} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L} x\right)}_{=: u_k(x,t)}, \qquad x \in \overline{R}$$

where A is some constant independent of k.

We can easily apply Weierstrass test: observing that

$$|u_k(x,t)| := \frac{|A|}{k^4} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \left|\sin\left(\frac{k\pi}{L} x\right)\right| \le \frac{|A|}{k^4}$$

We can consider $M_k = \frac{|A|}{k^4}$ and since the series $\sum_k \frac{1}{k^4}$ is convergent (the series $\sum_k \frac{1}{k^\alpha}$ converge if $\alpha > 1$ and diverge if $\alpha \leq 1$) Weierstrass' test tells us that the series (x) is uniformly convergent and result R1 allow us to conclude that $u(x,t)$ is continuous in $\overline{R}$.

What about differentiability and integrability ?

The integrability result is also very similar to R1:

R2 | If $u_n(x)$ are integrable functions in I
If $\sum_n u_n(x)$ is uniformly convergent
then the sum $u(x,t) = \sum_{n=1}^{\infty} u_n(x,t)$ is integrable
and
$$\int_I \left(\sum_{n=1}^{\infty} u_n(x,t) \right) dx = \sum_{n=1}^{\infty} \int_I u_n(x,t)\, dx$$

[this equality result means that the integral of the series (= the left-hand side) can be done term-by-term (= the right-hand side).]

The differentiability result requires an additional condition, easily satisfied in practice:

R3

If $u_n(x)$ are continuously differentiable in an interval I

If the series $\sum_n u_n'(x)$ is uniformly convergent

If there exists a $x_0 \in I$ such that $\sum_n u_n(x_0)$ is convergent

then, the sum $u(x) = \sum_{n=1}^{\infty} u_n(x)$ is continuously differentiable and

$$\frac{du}{dx}(x) = \frac{d}{dx}\left(\sum_{n=1}^{\infty} u_n(x)\right) = \sum_{n=1}^{\infty} u_n'(x)$$

[Like in the previous case this means that the series can be differentiate term-by-term]

Using these results, together with Weierstrass' test, is useful to check differentiability of formal solutions: suppose we want to check that $u(x,t)$ given by (x) is C^1_t:

Now

$$\left|\frac{\partial u_k}{\partial t}\right| = \left|\frac{\partial}{\partial t}\left(\frac{A}{k^4} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \sin\left(\frac{k\pi}{L}x\right)\right)\right|$$

$$= \left|\frac{A}{k^4}(-k)\left(\frac{k\pi}{L}\right)^2 e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L}x\right)\right|$$

$$\leq \frac{|A|k^2\pi^2/L^2}{k^2} = \frac{\text{constant}}{k^2} =: M_k$$

But the series $\sum_k M_k = \sum_k \frac{\text{constant}}{k^2}$ is convergent and so

the series of $\sum_k \left|\frac{\partial u_k}{\partial t}\right|$ is uniformly convergent.

As we know that $\sum_k u_k(x,t)$ is convergent in $\bar{\Omega}$ (by Fourier's theorem!) then aplying R3 we conclude that $u(x,t)$ given by (x) is continuously differentiable relative to the variable t.

To settle this type of argument let us apply it again to analyse what we can say about the second derivative with respect to the variable x, when $t \geq 0$,

$$\left|\frac{\partial u_k}{\partial x}\right| = \left|\frac{\partial}{\partial x}\left(\frac{A}{k^4} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \sin\left(\frac{k\pi}{L}x\right)\right)\right|$$

$$= \left|\frac{A}{k^4}\frac{k\pi}{L} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \cos\left(\frac{k\pi}{L}x\right)\right|$$

$$\leq \frac{|A|\pi/L}{k^3} = \frac{\text{constant}}{k^3} =: M_k$$

Again $\sum_k M_k = (\text{constant}) \sum_k \frac{1}{k^3}$ is convergent, so by Weierstrass test the series $\sum_k \frac{\partial u_k}{\partial x}$ is uniformly convergent. As we know that $\sum_k u_k(x,t)$ is convergent in $\bar{R}$ we know that, by R3, the function $u(x,t)$ is differentiable relative to the variable x and we have

$$\underbrace{\frac{\partial u}{\partial x}(x,t)}_{=: v(x,t)} = \sum_{k=1}^{\infty} \underbrace{\frac{A\pi/L}{k^3} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \cos\left(\frac{k\pi}{L}x\right)}_{=: v_k(x,t)}$$

Now applying the same procedure to the series $v(x,t) = \sum_k v_k(x,t)$ we have

$$\left|\frac{\partial v_k}{\partial x}\right| = \left|\frac{\partial}{\partial x}\left(\frac{A\pi/L}{k^3} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \cos\left(\frac{k\pi}{L}x\right)\right)\right|$$

$$= \left|-\frac{A\pi^2/L^2}{k^2} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \sin\left(\frac{k\pi}{L}x\right)\right|$$

$$\leq \frac{|A|\pi^2/L^2}{k^3} = \frac{\text{constant}}{k^3} =: M_k$$

Again for this sequence (M_k) we have $\sum_k M_k$ convergent and Weierstrass' test imply the series $\sum_k \frac{\partial v_k}{\partial x}(x,t)$ is uniformly convergent. As we know that $\sum_k v_k(x,t)$ is convergent whatever (x,t) may be (in $\mathcal{R}$) [actually it is even uniform convergent, as we saw above] we can apply R_3 to conclude that $v(x,t)$ is differentiable with respect to x and, in $\mathcal{R}$, it holds

$$\frac{\partial v}{\partial x} = \sum_{k=1}^{\infty} \frac{\partial v_k}{\partial x}(x,t)$$

but $\frac{\partial v}{\partial x} = \frac{\partial^2 u}{\partial x^2}$ and $\frac{\partial v_k}{\partial x} = \frac{\partial^2 u_k}{\partial x^2}$, and hence

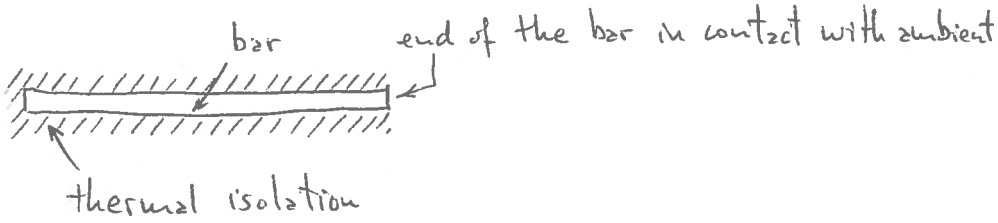
$$\frac{\partial^2 u}{\partial x^2}(x,t) = -\sum_{k=1}^{\infty} \frac{A\pi^2/L^2}{k^2} e^{-k\left(\frac{k\pi}{L}\right)^2 t} \cdot \cos\left(\frac{k\pi}{L}x\right)$$

and the series, due to the same arguments, is uniformly convergent and so $u \in C^2$.

We reinforce once more that conditions R1, R2 and R3, as well as Weierstrass' test are sufficient conditions. Even when their assumptions are not satisfied it may happen that their conclusions still hold. In those cases checking that a formal solution to the PDE is indeed a classical solution will require more sophisticated (and usually more difficult to apply) results. We will not focus our attention on those cases here.

Let's exemplify once more Fourier's method applied to the diffusion (or heat) equation.

Consider a bar of length 1 made of a material with termal diffusivity equal to 1 and such that it is thermally isolated along its length, also isolated at one end, and in contact with the ambient temperature $u=0$ at the other end. Schematically we have the following:



Assuming the non-isolated end is kept at the constant ambient temperature θ_0 the boundary value problem for

the evolution of the temperature $u(x,t)$ with time t along the bar assumed to be modelled by the interval $[0,1]$ is

$$(+)_1 \quad \begin{cases} u_t = u_{xx} & (x,t) \in (0,1) \times \mathbb{R}^+ =: \mathcal{R} \\ u_x(0,t) = 0 \,, \; u(1,t) = \theta_0 \,, & t \geq 0 \end{cases}$$

Note that we used Fourier's cooling law at $x=0$: if this end is isolated then the flow of heat there must be zero, and the flow of heat is proportional to the derivative of the temperature relative to x, as we saw in session 9. In the problem studied in session 10 it was crucial to have boundary conditions equal to zero (remember why?) so we transform $(+)$ into a problem with boundary conditions equal to zero by a change of variables: Let $\theta(x,t) = u(x,t) - \theta_0$ be the temperature relative to ambient temperature θ_0. Then, $\frac{\partial \theta}{\partial t} = \frac{\partial u}{\partial t}$, $\frac{\partial \theta}{\partial x} = \frac{\partial u}{\partial x}$, and $\frac{\partial^2 \theta}{\partial x^2} = \frac{\partial^2 u}{\partial x^2}$, and so

$$(+)_2 \quad \begin{cases} \theta_t = \theta_{xx} & , (x,t) \in (0,1) \times \mathbb{R}^+ =: \mathcal{R} \\ \theta_x(0,t) = \theta(1,t) = 0 \,, & t \geq 0 \end{cases}$$

Let us solve $(+)_2$ using Fourier method: let

$$\theta(x,t) = F(x)G(t)$$

(assuming F and G non-zero). then $\theta_t = FG'$ and $\theta_{xx} = F''G$. then

$$\theta_t = \theta_{xx} \Leftrightarrow FG' = F''G \Leftrightarrow \frac{F''}{F} = \frac{G'}{G}$$

(note that F' represents the derivative of F relative to its independent variable, x, whereas G' represent the derivative relative to its independent variable, t.)

Since

$$\frac{F''}{F}(x) = \frac{G'}{G}(t), \qquad (x,t) \in \mathcal{R}$$

is valid in the open set $\mathcal{R} = (0,1) \times \mathbb{R}^+$, we conclude (as we did before) that it must exist a constant σ such that

$$\frac{F''(x)}{F(x)} = \sigma = \frac{G'(t)}{G(t)}, \qquad (x,t) \in \mathcal{R}.$$

the boundary conditions are

$0 = u_x(0,t) = \theta_x(0,t) = F'(0)G(t) \Rightarrow F'(0) = 0$

$\theta_0 = u(1,t) \Leftrightarrow 0 = \theta(1,t) = F(1)G(t) \Rightarrow F(1) = 0$

Let us then start by solving the boundary value problem

$$(+)_3 \quad \begin{cases} F'' = \sigma F \\ F'(0) = F(1) = 0 \end{cases}$$

To determine the possible values of σ we proceed, as in session 10, by trying different signs of σ, because that corresponds to solutions of $(+)_3$ with different behaviours:

- if $\sigma = 0$ the solution is $F(x) = \alpha x + \beta$, and the boundary conditions tell us that $0 = F'(0) = \alpha$, $0 = F(1) = \alpha + \beta$, and so $\alpha = \beta = 1$
- if $\sigma > 0$ the solution is $F(x) = \alpha e^{\sqrt{\sigma} x} + \beta e^{-\sqrt{\sigma} x}$ from which we get $F'(x) = \alpha \sqrt{\sigma} e^{\sqrt{\sigma} x} - \beta \sqrt{\sigma} e^{-\sqrt{\sigma} x}$ and the boundary conditions get us to

$$\begin{cases} 0 = \alpha \sqrt{\sigma} - \beta \sqrt{\sigma} \\ 0 = \alpha e^{\sqrt{\sigma}} + \beta e^{-\sqrt{\sigma}} \end{cases} \iff \begin{pmatrix} \sqrt{\sigma} & -\sqrt{\sigma} \\ e^{\sqrt{\sigma}} & e^{-\sqrt{\sigma}} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and since the determinant of the matrix is $\sqrt{\sigma} e^{-\sqrt{\sigma}} - \sqrt{\sigma} e^{\sqrt{\sigma}} = \sqrt{\sigma} \left(e^{-\sqrt{\sigma}} - e^{\sqrt{\sigma}} \right) \neq 0$, the matrix is invertible and so the only solution of the equation is $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

- for $\sigma = -\mu^2 < 0$ we have, as before, solutions

$$F(x) = \alpha \cos \mu x + \beta \sin \mu x$$

and the boundary condition gives

$$\begin{cases} 0 = F'(0) = -\alpha \mu \sin(\mu \cdot 0) + \beta \mu \cos(\mu \cdot 0) \\ 0 = F(1) = \alpha \cos \mu + \beta \sin \mu \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} \beta \mu = 0 \\ \alpha \cos \mu + \beta \sin \mu = 0 \end{cases} \Leftrightarrow \begin{cases} \beta = 0 \\ \alpha \cos \mu = 0 \end{cases}$$

and so, for $\alpha \cos \mu = 0$ and the solution to the differential equation not to be identically zero we need to choose μ such that $\cos \mu = 0$, which means

$$\mu = \mu_k = k\pi - \frac{\pi}{2} = \frac{(2k-1)\pi}{2}, \quad k \in \mathbb{N}^+$$

for which the solution of the equation for F is

$$F_k(x) = \cos\left(\frac{(2k-1)\pi}{2} x\right), \quad k \in \mathbb{N}^+$$

then, going back to the equation for G, $G' = \sigma G$, with these $\sigma = -\mu_k^2 = -\left(\frac{(2k-1)\pi}{2}\right)^2$, we get the solution

$$G_k(t) = \exp\left(-\left(\frac{(2k-1)\pi}{2}\right)^2 t\right), \quad k \in \mathbb{N}^+$$

and the formal solution of the boundary value problem $(+)_2$ is

$$(+)_4 \qquad \theta(x,t) = \sum_{k=1}^{\infty} \alpha_k \cos\left(\frac{(2k-1)\pi}{2} x\right) e^{-(2k-1)^2\pi^2 t/4}.$$

Now suppose we are given an initial temperature distribution f, meaning that we know that

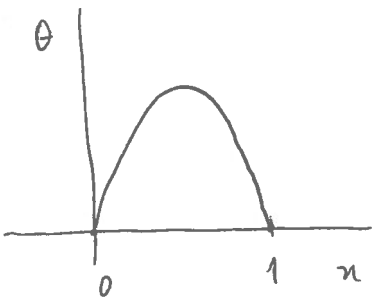
$$u(x,0) = f(x) \quad , \quad x \in [0,1],$$

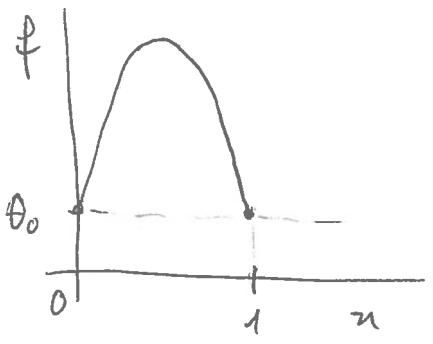
and that distribution is modelled by the function

$$f(x) = \theta_0 + 1 - x^2.$$

So, the initial condition for the relative temperature θ is $\theta(x,0) = \theta_{ini}(x)$ where

$$\theta_{ini.}(x) = f(x) - \theta_0 = 1 - x^2.$$

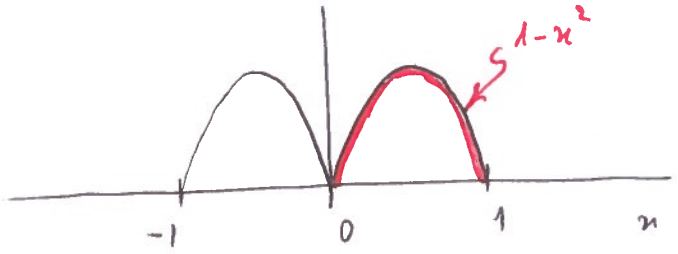




Using $(\dagger)_4$ the initial condition becomes

$$(\ast) \qquad 1-x^2 = \theta_{ini}(x) = \theta(x,0) = \sum_{k=1}^{\infty} \alpha_k \cos\left(\frac{(2k-1)\pi}{2} x\right).$$

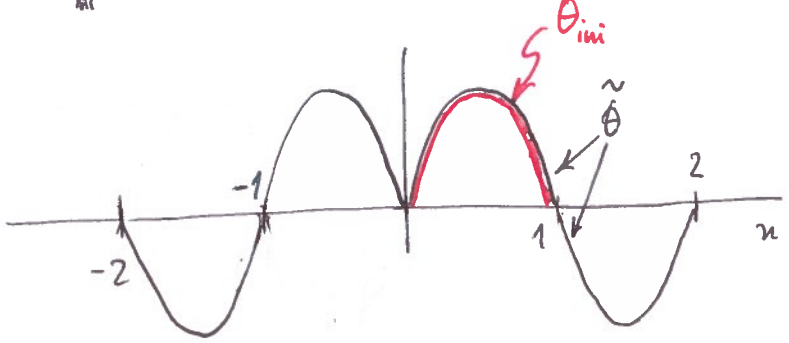
So, we need to extend the function $1-x^2$ to the whole real line as an even function, so that its Fourier series is a series of cosines. As before, the first step is to extend it as an even function to $[-1,1]$:



Now observe that the period of the right-hand side of $(\ast)$ is $2L$ with $L=2$, so the period of the extension of $1-x^2$ to $\mathbb{R}$ must be $2L=4$. To extend the above function defined in $[-1,1]$ as a 4-periodic function we need to first define the extension in $[-2,2]\setminus[-1,1]$. We can do this in an infinite number of ways. We can do it, as before, by defining it to be zero, or something else (it doesn't really matter in the end because all the extensions coincide with $1-x^2$ on $[0,1]$: although

the corresponding Fourier series will be different, their sum will be the same in the place of interest!)

Let us consider now a different extension from what we did before. Consider that, in $x \in [1,2]$, the function $\tilde{\theta}(x) = -\theta_{ini}(2-x)$, and then extended it to $[-2,2]$ as even.



And now extend this as a 4-periodic function $\tilde{\tilde{\theta}}$ in $\mathbb{R}$; then, the Fourier series of $\tilde{\tilde{\theta}}$ is a series of cosines (because $\tilde{\tilde{\theta}}$ is even) and the Fourier coefficients are

$$a_0 = \frac{2}{2}\int_0^2 \tilde{\tilde{\theta}}(x)\,dx = \int_0^1 \theta_{ini}(x)\,dx + \int_1^2 (-\theta_{ini}(2-x))\,dx$$

$$= \int_0^1 \theta_{ini}(x)\,dx - \int_0^1 \theta_{ini}(u)\,du$$

changing variables $x \mapsto u = 2-x$

$$= 0$$

and again by changing variables we can conclude that

$$a_{2k} = \frac{2}{2}\int_0^2 \tilde{\tilde{\theta}}(x)\cos\left(\frac{2k\pi}{2}x\right)dx = \dots = 0$$

$$a_{2k-1} = \frac{2}{2}\int_0^2 \tilde{\tilde{\theta}}(x)\cos\left(\frac{(2k-1)\pi}{2}x\right)dx = \dots = \frac{2(-1)^{k+1}}{(2k-1)\pi}\left(1 + \frac{8}{(2k-1)^3\pi^2}\right)$$

Hence, the formal solution to the initial-boundary value problem $(+)_2$ with initial condition $\theta_{ini}(x) = 1-x^2$ is

$$(+)_5 \qquad \theta(x,t) = \sum_{k=1}^{\infty} \frac{2(-1)^{k+1}}{(2k-1)\pi}\left(1+\frac{8}{(2k-1)^3\pi^3}\right)\cos\left(\frac{2k-1}{2}\pi x\right) \times e^{-(2k-1)^2\pi^2 t/4}$$

and the solution to the original problem for the temperature u is just

$$u(x,t) = \theta_0 + \theta(x,t).$$

Observe that the methods used before do not allow us to prove that $(+)_5$ is a classical solution because the best bound to the summand in $(+)_5$ on $\overline{R}$ is

$$\frac{2}{(2k-1)\pi}\left(1+\frac{8}{(2k-1)^3\pi^3}\right) =: M_k \text{ , and, since the series } \sum_k \frac{1}{k}$$

is divergent, we cannot apply Weierstrass' test as we did before; this is not uncommon but, as stated before, we cannot go further in this direction in this introductory course.

We can, however, do a little bit more in the study of $(+)_5$

Suppose we want to know the differentiability properties of $(t)_5$ only for $t>0$ and not on $\overline{R} = [0,1]\times[0,\infty)$. Consider a fixed $\tau>0$. If, in $(t)_5$, we have $t \geq \tau$ then

$$|\theta(x,t)| \leq \sum_{k=1}^{\infty} \left|\frac{2(-1)^{k+1}}{(2k-1)\pi}\right| \left(1 + \frac{8}{(2k-1)^3\pi^3}\right) \left|\cos\left(\frac{2k-1}{2}\pi x\right)\right| e^{-\left(\frac{2k-1}{2}\pi\right)^2 t}$$

$$\leq \sum_{k=1}^{\infty} \frac{4}{(2k-1)\pi} e^{-\left(\frac{2k-1}{2}\pi\right)^2 t}$$

$$\leq \sum_{k=1}^{\infty} \frac{4}{(2k-1)\pi} e^{-\left(\frac{2k-1}{2}\pi\right)^2 \tau} \qquad \text{(} t \geq \tau \text{)}$$

now the right-hand side does not depend on t (remember that $\tau>0$ is a fixed, constant, number) and so the terms inside the sum are elements of a sequence

$$M_k := \frac{4}{(2k-1)\pi} e^{-\left(\frac{2k-1}{2}\pi\right)^2 \tau}$$

$$\leq 2\, e^{-\left(\frac{2k-1}{2}\pi\right)^2 \tau} \qquad \text{(} e^{-\alpha_k} = \frac{1}{e^{\alpha_k}} \leq \frac{1}{1+\alpha_k} \text{)}$$

$$\leq \frac{2}{1+\left(\frac{2k-1}{2}\pi\right)^2 \tau}$$

and knowing that the series $\sum_k \frac{1}{k^2}$ is convergent we have the convergence of $\sum_k M_k$ and, by Weierstrass test, the uniform convergence of the series defining $\theta(x,t)$ and

hence its continuity for all $(x,t) \in [0,1] \times [\tau, \infty)$. Since $\tau > 0$ was fixed arbitrarily this means that we concluded continuity for (x,t) in $[0,1] \times (0,\infty)$.

Exactly the same argument can be used for the derivatives. Let's see on case: the derivative of the terms of the series relative to t gives

$$\frac{\partial \theta_k}{\partial t} = \frac{2(-1)^{k+1}}{(2k-1)\pi}\left(1 + \frac{8}{(2k-1)^3\pi}\right)\cos\left(\frac{2k-1}{2}\pi x\right)\cdot\left(-\left(\frac{2k-1}{2}\pi\right)^2\right)e^{-\left(\frac{2k-1}{2}\pi\right)^2 t}$$

Now, for $(x,t) \in [0,1] \times [\tau, \infty)$ with $\tau > 0$ fixed, we get

$$\left|\frac{\partial \theta_k}{\partial t}\right| \leq \frac{2k-1}{\pi^3} e^{-\left(\frac{2k-1}{2}\pi\right)^2 \tau} =: \tilde{M}_k$$

and, again, $\tilde{M}_k$ is independent of (x,t) and it is easy to see that the series $\sum_k \tilde{M}_k$ is convergent. Hence by Weierstrass test, $\frac{\partial}{\partial t}\theta(x,t)$ is uniformly convergent and, by R_3, $\theta \in C^1_t([0,1] \times [\tau, \infty))$ and hence in $[0,1] \times (0,\infty)$.

This can be repeated for derivatives of θ in t and in x of all orders! On will get always bounds of the type

$$\left|\frac{\partial^p \theta_k}{\partial t^q \partial x^{p-q}}\right| \leq (\text{polynomial in } k) \cdot e^{-\left(\frac{2k-1}{2}\pi\right)^2 \tau}$$

and thus the conclusion that $\theta \in C^\infty([0,1] \times (0,\infty))$, a somewhat surprising result!

OMIT

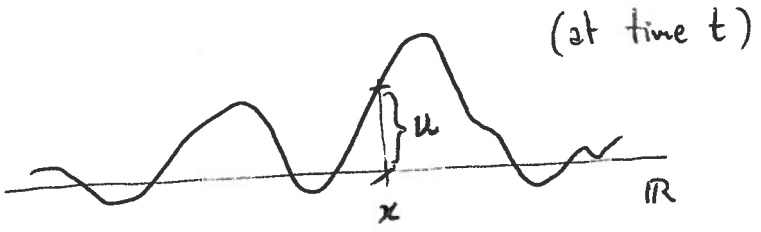
Differential Equations and Applications

Session 13: Applications of Differential Equations:
Wave equations: deduction and some examples
of linear and nonlinear wave equations.

We will now consider a different class of equations, called <u>wave equations</u>, or <u>hyperbolic laws</u> in more general forms. Like we did with the diffusion equation, we will obtain them as a balance law. We will see that, for the linear case, Fourier method of separation of variables and Fourier series is still applicable, although the solutions will exhibit a completely different behaviour.

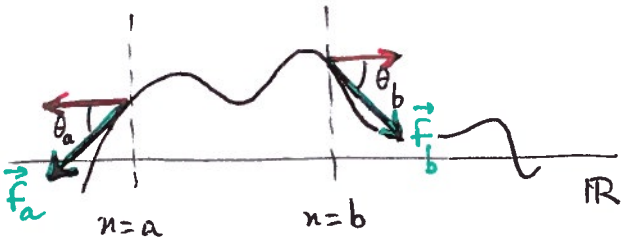
Let us consider a flexible rope vibrating along its rest horizontal position, identified as the real line $\mathbb{R}$.

Let $u = u(x,t)$ denote the transversal displacement of the point x of the rope at time t



To obtain a differential equation governing the evolution of $u = u(x,t)$ with time at the different positions x we use Newton's law $\vec{F} = m\vec{a}$, where $\vec{F}$ is the resultant of the forces applied to the rope, $\vec{a}$ its acceleration, and m its mass. Let us see how to use it in this context.

Let us consider the piece of rope between two arbitrary locations $x=a$ and $x=b$.



Let denote by $\rho(x)$ the density of the rope at point x. So, as the rope has displacement only on the transversal direction to x, the (liner) momentum of the rope at a point x is given by $\rho(x)\frac{\partial u}{\partial t}$ and so the momentum of the stretch of the rope between a and b is given by

$$M(t) := \int_a^b \rho(x)\frac{\partial u}{\partial t}(x,t)\,dx.$$

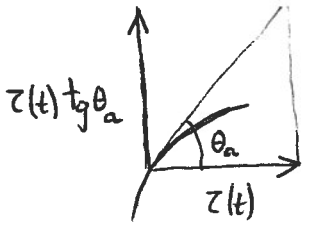
the forces that act on this piece of rope are of two types:

the tension forces $\vec{F}_a$ and $\vec{F}_b$, parallel to the rope at a and b, that were represented in the scheme above. Let $f(a,t) := \|\vec{F}_a(t)\|$ and $f(b,t) = \|\vec{F}_b(t)\|$. Since the points do not move in the x-direction (by assumption the vibrations are transversal!) the components of $\vec{F}_a$

and $\vec{F}_b$ in the x-direction must compensate each other (red arrows in the previous scheme); thus

$$f(a,t)\cos\theta_a = f(b,t)\cos\theta_b$$

where θ_a and θ_b are the angles of $\vec{F}_a$ and $\vec{F}_b$ with the horizontal x-axis. Thus, the horizontal component of the tension forces on the rope is independent of x (since the equality above is valid for all a and b) and depends only on time t; let us denote it by $\tau(t)$. But then, the vertical component of the tension



is given by $\tau(t)\,\mathrm{tg}\,\theta_a$ at $x=a$, and analogously for $x=b$. Remembering that the angle θ is, at each x, given by $\mathrm{tg}\,\theta = \frac{\partial u}{\partial x}$ at that x, the resultant of the forces of tension acting on the stretch of rope between $x=a$ and $x=b$ is

$$\tau(t)\,\mathrm{tg}\,\theta_b - \tau(t)\,\mathrm{tg}\,\theta_a = \tau(t)\,\frac{\partial u}{\partial x}(x,t)\Big|_{x=a}^{x=b}$$

$$= \int_a^b \tau(t)\,\frac{\partial^2 u}{\partial x^2}(x,t)\,dx$$

Additionally, there are external forces that can act on the rope affecting the movement, like gravitational forces, resistance of the medium where the rope is in, inner resistance to the movement, etc. If we denote the density of all these forces along the rope by a function $\tilde{h}(x,t,u)$, then Newton's law now gives

$$\int_a^b \tau(t)\frac{\partial^2 u}{\partial x^2}(x,t)\,dx + \int_a^b h_1(x,t,u(x,t))\,dx = \frac{\partial}{\partial t}\int_a^b \rho(x)\frac{\partial u}{\partial t}\,dx$$

Since a and b are arbitrary (and assuming $\frac{\partial^2 u}{\partial t^2}$ is continuous) we can write the equation for the evolution of the transversal displacement of the rope as

$$(*)_0 \qquad \rho(x)\frac{\partial^2 u}{\partial t^2} = \tau(t)\frac{\partial^2 u}{\partial x^2} + h_1(x,t,u)$$

which can be written as

$$(*)_1 \qquad \frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2} + h(x,t,u)$$

where $c(x,t)^2 = \frac{\tau(t)}{\rho(x)}$ and $h(x,t,u) = \frac{h_1(x,t,u)}{\rho(x)}$.

Equation (*) is known as the <u>wave equation.</u>

It is interesting to observe that an easy dimensional analysis argument give that c has the dimension of a velocity: in fact, $\tau(t)$ has the dimensions of a force, $[\tau] = MLT^{-2}$,

and $\rho(x)$ has the dimensions of a linear density, i.e., $[\rho] = ML^{-1}$; hence $c^2 = \frac{\tau}{\rho}$ gives $[c^2] = \frac{MLT^{-2}}{ML^{-1}} = L^2T^{-2}$ which means that $[c] = LT^{-1}$, indeed the dimensions of a velocity. We will see that this velocity has, in fact, a physical meaning.

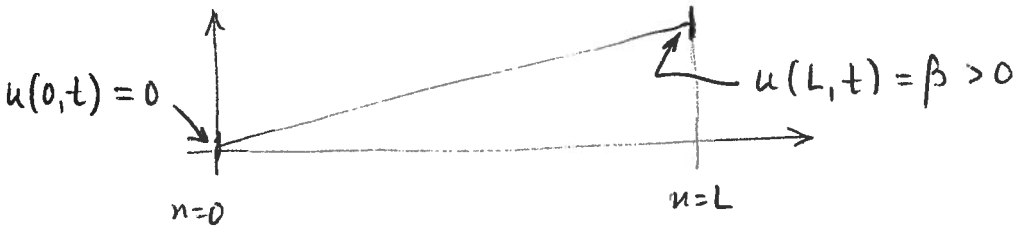
The simplest case of wave equation is when no external force h is acting, i.e., with $h \equiv 0$ in $(*)_1$. Other external forces are also important in the applications of the wave equation: typically the vibrating rope is immersed in a fluid (air, water, etc.) that damps its movement with a force opposing the movement and with intensity proportional to the velocity of the vibration, i.e., $h(x,t) = -b\frac{\partial u}{\partial t}$; another possible damping of the vibration is when the rope is equiped with some contraption producing a force leading to move the rope to its rest position and given by $h(x,t) = -au(x,t)$; an easier possibility is that the rope is subject to the force of gravity, which acts everywhere, independent of u and $\frac{\partial u}{\partial x}$ or $\frac{\partial u}{\partial t}$, in which case we have $h(x,t) = -g$.

As in the case of the diffusion equation we must complement the wave equation $(*)_1$ with appropriate additional conditions (initial, boundary, limit) which

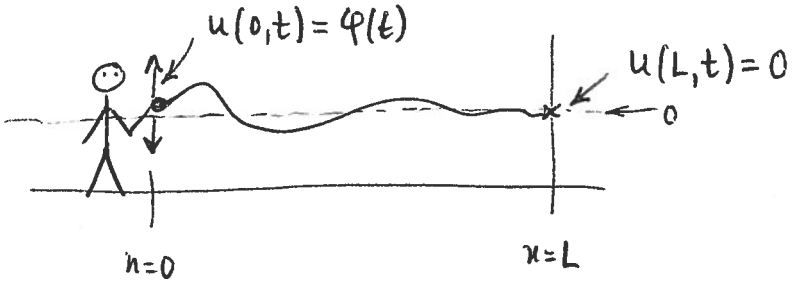
are used to obtain a unique solution of the equation.

If the rope has a finite length $L>0$, then we must prescribe the behaviour at the extremities of the rope. As in the diffusion equation we can have several boundary conditions: we can fix the positions of the rope at $x=0$ and $x=L$ independently of time, such as in the case

or

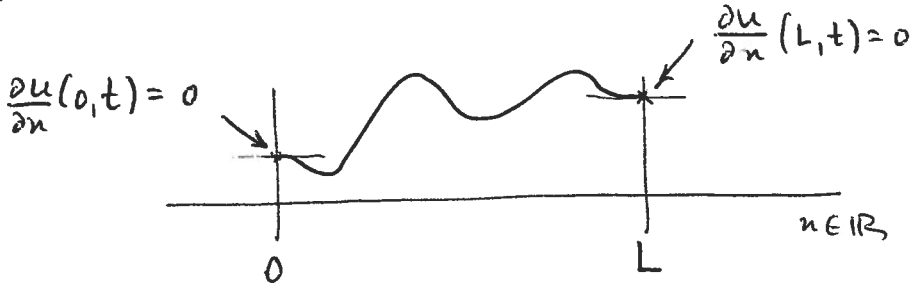


or even fixing positions at 0 or L that may vary with time (as when someone is making a rope vibrating):



Another possibility is that the rope has extremities at $x=0$ and $x=L$ that are free to move (in the transversal direction); in mathematical terms this is

represented by the boundary conditions $\frac{\partial u}{\partial n}(0,t) = 0$ and $\frac{\partial u}{\partial n}(L,t) = 0$.



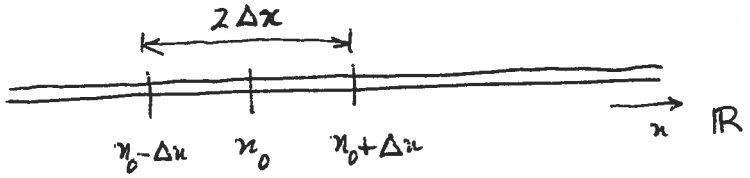
Other boundary conditions are possible, depending on the physical setting of the rope attachments.

On top of the boundary conditions we need to have <u>initial conditions</u>, i.e., additional conditions in the t variable at some fixed t. It is reasonable to expect that to fix the solution of the second-order in t differential equation $(*)_1$ we need to fix to conditions at a given time; physically it is reasonable to fix the initial position $u(n,0)$ and the initial velocity $\frac{\partial u}{\partial t}(n,0)$ of the rope at the initial time $t=0$.

So, in contradistinction to the diffusion equation (which is first-order in t ...) the initial condition for the wave equation $(*)_1$ requires that two function, φ and ψ, be given such that $u(n,0) = \varphi(n)$ and $\frac{\partial u}{\partial t}(n,0) = \psi(n)$, for all $n \in [0,L]$.

Another mathematical model that, although apparently very different from the wave equation, and occurring in a entirely different situation, turns out to be intimately related, is the equation that will appear in the modelling of the motion of objects along a one dimensional path: cars moving along a street (with no crossings) or cells moving in a blood vessel, or molecules along a nanotube, or elements of fluid along a pipe, etc.

Let us consider here a continuum model of traffic flow. In this type of modelling we ignore individual cars and, instead, focus our attention on macroscopic traffic variables, such as density of cars, ρ, their flux, J, and average speed, u. Clearly this macroscopic variables are related to relevant microscopic variables in some kind of continuum limit, as we saw in the relation of the random walk with the diffusion equation. In the present case, assuming a street, with no intersections, modelled by the real line $\mathbb{R}$, an approximate value of the density of cars at a point $x_0 \in \mathbb{R}$ in the instant t_0 is obtained by



$$(**)_1 \qquad \rho(x_0, t_0) \approx \frac{\text{number of cars from } x_0-\Delta x \text{ to } x_0+\Delta x \text{ at time } t=t_0}{2\Delta x}$$

and the flux of cars at x_0 in the instant t_0 can be approximate by counting the number of cars passing x_0 in a small time interval containing t_0, as in

$$J(x_0,t_0) \approx \frac{\text{net number of cars passing } x_0 \text{ from } t_0-\Delta t \text{ to } t_0+\Delta t}{2\Delta t}. \qquad (**)_2$$

(with a car moving to the right counting +1, and to the left counting -1.)

Now, the balance law for cars in the road is a conservation law (cars are not created nor destroyed on the road!):

$$\begin{aligned}
&(\#\text{ cars in } [x_0-\Delta x, x_0+\Delta x] \text{ at } t_0+\Delta t) - \\
&-(\#\text{ cars in } [x_0-\Delta x, x_0+\Delta x] \text{ at } t_0-\Delta t) = \\
&\quad = (\text{net } \#\text{ cars crossing } x_0-\Delta x \text{ in times } [t_0-\Delta t, t_0+\Delta t]) - \\
&\quad\quad -(\text{net } \#\text{ cars crossing } x_0+\Delta x \text{ in times } [t_0-\Delta t, t_0+\Delta t])
\end{aligned}$$

and using $(**)_1$ and $(**)_2$ we get

$$\begin{aligned}
&2\Delta x\big(\rho(x_0,t_0+\Delta t)-\rho(x_0,t_0-\Delta t)\big) = \\
&\quad = 2\Delta t\big(J(x_0-\Delta x,t_0)-J(x_0+\Delta x,t_0)\big)
\end{aligned}$$

Assuming ρ and J can be expanded in Taylor series about (x_0,t_0) we can write the last equality as

$$2\Delta x\Big(\rho + \Delta t\,\rho_t + \tfrac{1}{2!}(\Delta t)^2\rho_{tt} + \tfrac{1}{3!}(\Delta t)^3\rho_{ttt} + \cdots$$
$$-\rho - (-\Delta t)\rho_t - \tfrac{1}{2!}(-\Delta t)^2\rho_{tt} - \tfrac{1}{3!}(-\Delta t)^3\rho_{ttt} - \cdots\Big) =$$

$$= 2\Delta t\Big(J + (-\Delta x)J_x + \tfrac{1}{2!}(-\Delta x)^2 J_{xx} + \tfrac{1}{3!}(-\Delta x)^3 J_{xxx} + \cdots$$
$$- J - \Delta x\,J_x - \tfrac{1}{2!}(\Delta x)^2 J_{xx} - \tfrac{1}{3!}(\Delta x)^3 J_{xxx} - \cdots\Big)$$

with all functions ρ, J, ρ_t, J_x, etc... evaluated at (x_0, t_0). this equation can be written as

$$\rho_t + O((\Delta t)^2) = -J_x + O((\Delta x)^2)$$

and so, letting $\Delta x, \Delta t \to 0$ we have the macroscopic balance law

$(**)_3$
$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x}.$$

Note that $(**)_3$ is exactly equal to the conservation (or balance) law we deduced in session 9 (equation (*)) for the diffusion / heat equation.

What will distinguish this model from the diffusion/heat case is the relation between J and ρ. In the diffusion / heat model the relation was given by fick's / / fourier's law. Now a different relation is in operation:

In some cases (e.g.: for a uniform distribution of cars) we can define an average velocity at (x_0, t_0), $v = v(x_0, t_0)$ in a natural way and check that $J = \rho v$. In other (more complicated) cases this is used as the definition of the flux J, or of the average velocity v. With this relation equation $(**)_3$ becomes

$$(**)_4 \qquad \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(v\rho) = 0$$

In the context of traffic flow modelling this equation is called the Lighthill - Whitham - Richards equation, but it appears in many other circumstances (way before its use in traffic in the mid 1950s), in which it has different names: it is the (mass) continuity equation in fluid mechanics, or the current continuity equation in electrodynamics.

this is not yet what we wanted because we need to have a way to relate the density ρ with the average velocity v. In the context of traffic studies, as in many other situations, this have to be supplied by experiments or observations. the easiest relation, which

is typically valid for (very) small densities ρ is that v is equal to a constant a, $v = a$, and thus we get

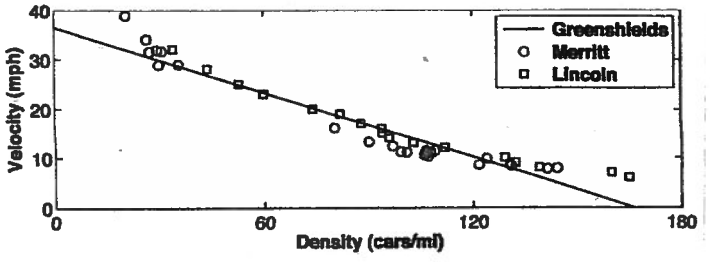
$$(**)_5 \qquad \frac{\partial \rho}{\partial t} + a \frac{\partial \rho}{\partial x} = 0$$

this is sometimes known as the advection equation, and it has a (at this point maybe surprising) relation with the linear wave equation $(*)_1$ (without outside forces acting, $h \equiv 0$) that we will explore in later sessions.

More realistic constitutive laws of v vs. ρ are valid for a fuller range of car densities on the road and, as pointed out before, are based on measurements in real setings. One such case is the linear law, or Greenshields model, given by

$$(x) \qquad v = v_M \left(1 - \frac{\rho}{\rho_M}\right)$$

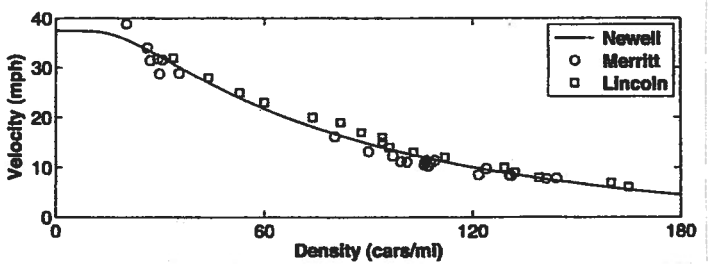
where v_M and ρ_M are constants. How this fits some data is shown next, with observational values in two different locations in New York (from Holmes' book, pg. 213).



A better fit to the shown data (and to other data sets) is obtained with the constitutive law proposed by Newell

$$v = v_M \left(1 - e^{-\lambda \left(\frac{1}{\rho} - \frac{1}{\rho_M} \right)} \right) \qquad (xx)$$

with $v_M, \rho_M, \lambda > 0$ constant parameters. With the same data as before the fitting is notoriously better:



However, the LWR-equation with (x) is a lot easier to study and its results easier to interpret and to apply than with (xx), so the use of a better fitting for the constitutive law has to be decided having the resulting harder analysis into consideration. This is, of course, the case in every instance where Mathematics is applied to Science and Technology!

GMIT

Differential Equations and Applications

Session 14: Linear wave equation in bounded intervals: use of separation of variables and Fourier series.

Let us consider the equation of a vibrating string of length L, attached at the end points at the same height and such that at the initial time (assumed to be $t=0$) it has a position and a velocity given by, respectively, $\varphi(x)$ and $\psi(x)$. The corresponding initial-boundary value problem is

$$(*)_1 \begin{cases} \dfrac{\partial^2 u}{\partial t^2} = c^2 \dfrac{\partial^2 u}{\partial x^2} \quad , \quad (x,t) \in (0,L) \times \mathbb{R}^+ \\ u(x,0) = \varphi(x) \, , \; \dfrac{\partial u}{\partial t}(x,0) = \psi(x) \quad , \quad x \in [0,L] \\ u(0,t) = 0 = u(L,t) \, , \quad t \geq 0 \end{cases}$$

Note that the part corresponding to the boundary value problem (i.e., the 1st and 3rd lines in $(*)_1$) is identical with what we had for the diffusion equation, namely: a linear differential equation and a linear and homogeneous boundary condition (i.e., a boundary condition equal to zero) these were the ingredients that allow us to use separation of variables; so it is natural to try to use the same method here!

Assume the solution $u = u(x,t)$ can be written as

$$u(x,t) = F(x)\,G(t).$$

then $\dfrac{\partial^2 u}{\partial t^2} = F(x)G'(t)$, $\dfrac{\partial^2 u}{\partial x^2} = F''(x)\,G(t)$, and the equation

becomes

$$F(x)G''(t) = c^2 F''(x)G(t) \qquad (x,t) \in \mathcal{R} := (0,L) \times \mathbb{R}^+$$

and assuming F and G are not zero we can write this in the following form with the independent variables x and t "separated":

$$\frac{F''}{F}(x) = \frac{1}{c^2}\frac{G''}{G}(t) \qquad (x,t) \in \mathcal{R} := (0,L) \times \mathbb{R}^+$$

As in the case of the diffusion equation, because we can change x and t independently from each other (a consequence of $\mathcal{R}$ being an open set) we conclude that it must exist a constant σ such that

$$\frac{F''}{F}(x) = \sigma = \frac{1}{c^2}\frac{G''}{G}(t) \qquad (x,t) \in \mathcal{R}$$

Now, the boundary conditions after writing $u = FG$ become

$$F(0) = 0 = F(L)$$

and so the boundary value problem for the function F is

$$(**) \quad \begin{cases} F''(x) - \sigma F(x) = 0 \\ F(0) = F(L) = 0 \end{cases}$$

which is exactly the equation we got for the diffusion equation.

Thus, the study of the possible values of σ that we presented in session 10 is applicable verbatim and we conclude that the possible values of σ are

$$\sigma = -\mu_k^2 = -\left(\frac{k\pi}{L}\right)^2 , \quad k \in \mathbb{N}^+$$

and the solutions of (**) are

$$F_k(x) = \sin\frac{k\pi x}{L} , \quad k \in \mathbb{N}^+.$$

Now the equation for G comming from the separation of variables is

$$G'' + c^2\left(\frac{k\pi}{L}\right)^2 G = 0 \qquad , k \in \mathbb{N}^+$$

and again the same argument used to solve the equation for F, but now without boundary conditions, give the result

$$G_k(t) = \alpha_k \cos\left(\frac{k\pi c}{L}t\right) + \beta_k \sin\left(\frac{k\pi c}{L}t\right) \qquad , k \in \mathbb{N}^+$$

for α_k, β_k arbitrary constants.

Thus, arbitrary solutions of the boundary value problem are obtained by linear combinations of $f_k(x)G_k(t)$, and we obtain a formal solution (in the same sense as we

used in the diffusion equation) if we allow infinite sums of these functions, namely

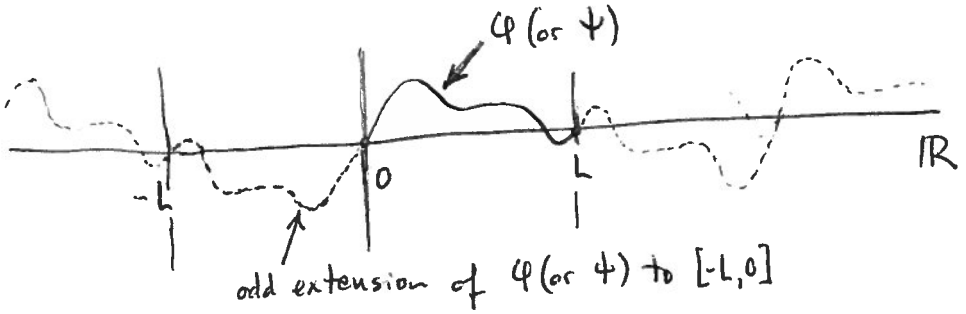
$$(*)_2 \qquad u(x,t) = \sum_{k=1}^{\infty} \left(\alpha_k \sin\left(\frac{k\pi}{L}x\right) \cos\left(\frac{k\pi c}{L}t\right) + \beta_k \sin\left(\frac{k\pi}{L}x\right) \sin\left(\frac{k\pi c}{L}t\right) \right).$$

The constants α_k and β_k are fixed by the initial conditions in a way similar to what was done in the diffusion case:

$$\varphi(x) = u(x,0) = \sum_{k=1}^{\infty} \alpha_k \sin\left(\frac{k\pi}{L}x\right),$$

$$\psi(x) = \frac{\partial u}{\partial t}(x,0) = \sum_{k=1}^{\infty} \beta_k \frac{k\pi c}{L} \sin\left(\frac{k\pi}{L}x\right),$$

and thus we need to expand φ and ψ in Fourier series of sines of period $2L$ to obtain the numerical values of the coefficients α_k and β_k: we need to extend φ and ψ to $[-L,L]$ as odd functions (because $\sin(\cdot)$ is an odd function, and the periodically to $\mathbb{R}$



From what we saw about Fourier series we then get

$$(*)_3 \qquad \alpha_k = \frac{2}{L}\int_0^L \varphi(x)\,\sin\frac{k\pi x}{L}\,dx$$

$$(*)_4 \qquad \beta_k = \frac{2}{k\pi c}\int_0^L \psi(x)\,\sin\frac{k\pi x}{L}\,dx$$

Again like what occurred with the diffusion equation the formal solution $(*)_2 - (*)_4$ is not automatically a true (classical) solution of $(*)_1$ and extra work has to be done to check if $(*)_2$ enjoys the needed differentiability properties in $\mathcal{R}$ and continuity in $\bar{\mathcal{R}}$, namely,

$$u \in C^0(\bar{\mathcal{R}}) \cap C_t^2(\mathcal{R}) \cap C_x^2(\mathcal{R})$$

Lets start by the continuity result: from $(*)_2$ we get

$$|u(x,t)| \le \sum_{k=1}^{\infty}\left(\left|\alpha_k \sin\left(\frac{k\pi}{L}x\right)\cos\left(\frac{k\pi c}{L}t\right)\right| + \left|\beta_k \sin\left(\frac{k\pi}{L}x\right)\sin\left(\frac{k\pi c}{L}t\right)\right|\right)$$

$$(\dagger) \qquad \le \sum_{k=1}^{\infty}\left(|\alpha_k| + |\beta_k|\right)$$

Now, to estimate this series we need to use $(*)_3$ and $(*)_4$ which relates it to the properties of the initial data φ and ψ.

Suppose φ is differentiable with $|\varphi'|$ integrable in $[0,L]$; then, integrating by parts,

$$|\alpha_k| = \left|\frac{2}{L}\int_0^L \varphi(x)\,\sin\frac{k\pi x}{L}\,dx\right|$$

$$(*) \qquad = \frac{2}{k\pi}\left|\int_0^L \varphi'(x)\,\cos\frac{k\pi x}{L}\,dx\right|$$

$$\leq \frac{2}{k\pi}\int_0^L |\varphi'| =: \frac{2M_1}{k\pi} = \frac{1}{k}(\text{constant})$$

and a similar result for β_k under the same conditions for φ. This is an interesting (and useful) result, showing that a regularity (differentiability) in the initial data translates into a faster decay of the Fourier coefficients.

If we assume φ'' exists and $|\varphi''|$ is integrable we can repeat the process once more, integrating by parts the integral in $(*)$ to get

$$|\alpha_k| \leq \frac{2M_2 L}{k^2\pi^2} = \frac{1}{k^2}(\text{constant})$$

where $M_2 := \int_0^L |\varphi''|$. And similarly for β_k. Using this in $(\dagger)$ we conclude that

$$|u(x,t)| \leq \sum_{k=1}^{\infty} (|\alpha_k| + |\beta_k|) \leq (\text{constant}) \sum_{k=1}^{\infty} \frac{1}{k^2}$$

and so, by Weierstrass' test, the series in the right-hand side of $(*)_2$ is uniformly convergent and since its terms are continuous functions we conclude that $u(x,t)$ is also continuous.

This result suggests that if we want u to be C^2 we need to impose better regularity assumptions on the data

In fact, using this integration by parts argument we need that $\varphi \in C^4$ and $\psi \in C^3$ to conclude that $u(x,t)$ is twice differentiable relative to t and x (in fact, by using an important result in Fourier series, called Bessel's inequality, which states that the Fourier coefficients of a function f such that $|f|, |f|^2$ are integrable satisfy $\frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \leq \frac{1}{L}\int_0^L |f|^2$, then the previously stated regularity results can be reduced to $\varphi \in C^3$ and $\psi \in C^2$.)

Clearly in many cases assuming the data is C^3 or C^4 is not very realistic. For instance, when we put into vibration a string in a musical instrument we usually pluck the string making its initial position much closer to a piecewise differentiable function than to a differentiable (C^3) one, like in the scheme

If we don't have enough regularity on the data we cannot use the argument above and, actually, the formal

solution may not be enough regular to be a true classical solution. In these cases we need to work with a weaker notion of solution, which essentially consists in working with the integrated version of the differential equation (the equation that was originally obtained from the balance law, before getting rid of the integral!) multiplied by a sufficiently regular "test function". To proceed in this direction it is important to understand present day research work in PDEs but would require a follow up course to the current one.

In the study of the wave equation we can define an <u>energy</u> that has analogies to what was done for the gravitational pendulum and, furthermore, is relevant both for theoretical and practical purposes.

Consider the equation for the transversal displacements of a rope (or wire) deduced in session 13

$$(*)_0 \qquad \rho(x)\frac{\partial^2 u}{\partial t^2} = \tau(t)\frac{\partial^2 u}{\partial x^2} + h_1(x,t,u)$$

and assume that the horizontal component of the tension forces does not depend on t, $\tau(t) \equiv \tau$, $\forall t$. Let us assume $u \in C^1(\bar{R}) \cap C^2(R)$. Multiplying $(*)_0$

by $\frac{\partial u}{\partial t}$ and integrating (in the x variable) in $[0,L]$ we get

$$\int_0^L \rho(x)\frac{\partial^2 u}{\partial t^2}\frac{\partial u}{\partial t}\,dx = \tau\int_0^L \frac{\partial^2 u}{\partial x^2}\frac{\partial u}{\partial t}\,dx + \int_0^L h_1(x,t,u)\frac{\partial u}{\partial t}\,dx.$$

Observe that $\frac{\partial^2 u}{\partial t^2}\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial t}\right)^2$, and so the left-hand side becomes equal to

$$\frac{1}{2}\frac{d}{dt}\int_0^L \rho(x)\left(\frac{\partial u}{\partial t}\right)^2 dx.$$

Furthermore, integrating by parts in the first integral of the right-hand side above we conclude it is equal to

$$\tau\left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial u}{\partial t}\right)\Big|_0^L - \int_0^L \tau\frac{\partial u}{\partial x}\frac{\partial^2 u}{\partial x\partial t}\,dx$$

and the integral here is equal to $\frac{1}{2}\frac{d}{dt}\int_0^L \tau\left(\frac{\partial u}{\partial x}\right)^2 dx$.

Hence, the above integral equality can be written as

$$(*)_2 \quad \begin{cases} \dfrac{d}{dt}\left[\dfrac{1}{2}\displaystyle\int_0^L \rho(x)\left(\frac{\partial u}{\partial t}\right)^2 dx + \frac{1}{2}\int_0^L \tau\left(\frac{\partial u}{\partial x}\right)^2 dx\right] = \\ \qquad = \tau\left(\dfrac{\partial u}{\partial x}\right)\left(\dfrac{\partial u}{\partial t}\right)\Big|_0^L + \displaystyle\int_0^L h_1(x,t,u)\frac{\partial u}{\partial t}\,dx \end{cases}$$

this equality is called the <u>energy equation</u> and the expression

$(\dagger)_1$ $$E(t) := \frac{1}{2}\int_0^L \rho(x)\left(\frac{\partial u}{\partial t}\right)^2 dx + \frac{1}{2}\int_0^L \tau\left(\frac{\partial u}{\partial x}\right)^2 dx$$

is called the <u>total energy</u> of the vibrating rope, with

$(\dagger)_2$ $$K(t) := \frac{1}{2}\int_0^L \rho(x)\left(\frac{\partial u}{\partial t}\right)^2 dx, \quad \underline{\text{kinetic energy}}$$

$(\dagger)_3$ $$V(t) := \frac{1}{2}\int_0^L \tau\left(\frac{\partial u}{\partial x}\right)^2 dx, \quad \underline{\text{potential energy}}$$

Note that in the case of the rope without external forces ($h_1 \equiv 0$) and freely standing extremities ($\frac{\partial u}{\partial x} = 0$ at $x = 0, L$) the energy equation becomes

$$\frac{d}{dt}E(t) = 0 \quad , \forall t$$

which means that the total energy is a conserved quantity, i.e., the wave equation (with those boundary conditions) is a conservative system and

$(\ddagger)_1$ $$E(t) = E(0), \quad \forall t$$

where $E(0)$, the initial energy, is computed from the initial condition in the place of $\frac{\partial u}{\partial t}$ and $\frac{\partial u}{\partial x}$ in $(\dagger)_1$:

$(\ddagger)_2$ $$E(0) = \frac{1}{2}\int_0^L \rho(x)\,\psi(x)^2 dx + \frac{1}{2}\int_0^L \tau\,(\varphi'(x))^2 dx$$

this is also the case if the extremities of the rope are kept fixed because then $\frac{\partial u}{\partial t}(0,t) = 0 = \frac{\partial u}{\partial t}(L,t)$, since $u(0,t) = u_0$ (a constant) and the same for the value of u at L.

Recalling that the Fourier method resulted in the expression of $u(x,t)$ as a series, $u(x,t) = \sum_{k=1}^{\infty} u_k(x,t)$, of functions $u_k(x,t)$ which, in the case of the boundary value problem corresponding to $(*)_1$, are

$$(*)_1 \qquad u_k(x,t) = \alpha_k \sin\left(\frac{k\pi}{L}x\right)\cos\left(\frac{k\pi c}{L}t\right) + \beta_k \sin\left(\frac{k\pi}{L}x\right)\cos\left(\frac{k\pi c}{L}t\right)$$

which can be written as

$$(*_2) \qquad u_k(x,t) = \gamma_k \sin\left(\frac{k\pi c}{L}t + \theta_k\right)\sin\left(\frac{k\pi}{L}x\right)$$

where

$$\gamma_k = \sqrt{\alpha_k^2 + \beta_k^2}$$

and the <u>phase</u> θ_k is defined by

$$\theta_k = \operatorname{arctg}\frac{\alpha_k}{\beta_k}.$$

the function $(*)_1$ or $(*)_2$ is called the <u>k^{th} harmonic</u> of the solution $u(x,t)$ and the smallest k for which $\alpha_k, \beta_k \neq 0$ (and so $u_k(x,t) \not\equiv 0$) is called

the fundamental harmonic.

It is clear from $(*)_2$ that, for each fixed $\hat{n}$, the point of the rope has a periodic movement with amplitude $r_k \sin\frac{k\pi\hat{n}}{L}$ and period $T_k = \frac{2L}{kc}$; the corresponding frequency is $\omega_k = T_k^{-1}$, which are called the amplitude and the frequency of the k^{th} harmonic.

From $(*)_2$ we can compute the kinetic and the potential energy of the k^{th} harmonic, obtaining the k^{th} harmonic total energy

$$E_k = \frac{1}{2}\int_0^L \rho(x)\, r_k^2\, \frac{k^2\pi^2c^2}{L^2}(\cos^2\beta_k)\,\mathrm{sen}^2\left(\frac{k\pi x}{L}\right)dx +$$

$$+ \frac{1}{2}\int_0^L \tau\, r_k^2\, \frac{k^2\pi^2}{L^2}(\sin^2\beta_k)\cos^2\left(\frac{k\pi x}{L}\right)dx$$

where $\beta_k := \frac{k\pi c}{L}t + \theta_k$ is independent of x; after some computations we get, if ρ is constant,

$$E_k = \frac{k^2\pi^2}{4L}\rho c^2 r_k^2 = M\pi^2 r_k^2 \omega_k^2$$

) using $c^2 = \tau/\rho$, $M = L\rho$ and the definition of ω_k

Because $u(x,t) = \sum_{k=1}^{\infty} u_k(x,t)$, and remembering the orthogonality relations of the sine and cosine functions (session 11), we can conclude, after using $(\sharp)_2$ and the conservation of energy $(\sharp)_1$ with ρ and τ constants independent of x and t, respectively, that, for all t,

$$E(t) = E(0) = \frac{1}{2}\sum_{k=1}^{\infty} \rho \frac{k^2\pi^2c^2}{L^2}\beta_k^2 \frac{L}{2} + \frac{1}{2}\sum_{k=1}^{\infty} \tau \frac{k^2\pi^2}{L^2}\alpha_k^2 \frac{L}{2}$$

$$= \sum_{k=1}^{\infty} \frac{k^2\pi^2}{4L}\rho c^2 \gamma_k^2$$

$$= \sum_{k=1}^{\infty} E_k .$$

So the energy of the vibrating string is the sum of the energy of all its harmonics.

To end this brief study of the linear wave equation we observe from $u(x,t)$ given in $(*)_2$ a characteristic that will be explored more generaly in the next session, and is the following.

Recalling the trigonometric identities

$$\cos(a+b) = \cos a \cos b - \sin a \sin b$$

$$\sin(a+b) = \sin a \cos b + \cos a \sin b$$

we can write

$$\sin a \cos b = \frac{1}{2}\left(\sin(a+b) + \sin(a-b)\right)$$

$$\sin a \sin b = \frac{1}{2}\left(\cos(a-b) - \cos(a+b)\right)$$

which, when used in the expression for the formal solution $(*)_2$ gives, after rearranging the terms

$$u(x,t) = \frac{1}{2}\sum_{k=1}^{\infty}\left(\alpha_k \sin\frac{k\pi(x+ct)}{L} - \beta_k \cos\frac{k\pi(x+ct)}{L}\right) +$$

$$+ \frac{1}{2}\sum_{k=1}^{\infty}\left(\alpha_k \sin\frac{k\pi(x-ct)}{L} + \beta_k \cos\frac{k\pi(x-ct)}{L}\right)$$

So, the solution of $(*)_1$ can be written as the sum of two functions

$(*)_5$ $$u(x,t) = F(x+ct) + G(x-ct)$$

with

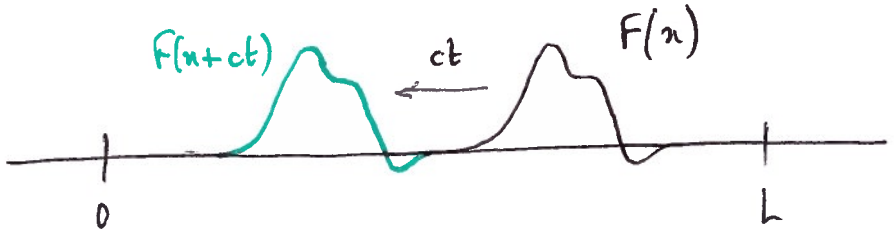
$$F(\xi) = \frac{1}{2}\sum_{k=1}^{\infty}\left(\alpha_k \sin\frac{k\pi\xi}{L} - \beta_k \cos\frac{k\pi\xi}{L}\right)$$

and

$$G(\xi) = \frac{1}{2}\sum_{k=1}^{\infty}\left(\alpha_k \sin\frac{k\pi\xi}{L} + \beta_k \cos\frac{k\pi\xi}{L}\right)$$

If we recall the study developed in session 8 we

note that $F(x+ct)$ is the translation of the graph of $F(x)$ by ct units to the left: the constant c is, indeed a velocity with physical meaning, as stated in session 13. Likewise the graph of $G(x-ct)$ is the translation of the graph of G by ct units to the right. Schematically,



and similarly for G.

This is a very interesting result: the solution of $(*)_1$ is the sum of a "wave" traveling to left with another travelling to the right, constructed from the initial condition, both travelling with the same absolute speed c, dependent only on the physical properties (tension τ and density ρ) of the vibrating material.

The importance of this result is that it is independent of the boundary conditions: every solution of $\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$ exhibits this behaviour, as the following theorem, proved next, shows:

Theorem

Let $u = u(x,t)$ be a solution of the wave equation $\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$, with c a constant.

Then, there exists functions $F, G : \mathbb{R} \to \mathbb{R}$ such that $u(x,t) = F(x+ct) + G(x-ct)$.

Proof:

the proof is very simple: let $\xi := x+ct$, $\eta := x-ct$, and define N by

$$N(\xi, \eta) = N(x+ct, x-ct) := u(x,t).$$

then
$$\frac{\partial u}{\partial x} = \frac{\partial N}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial N}{\partial \eta}\frac{\partial \eta}{\partial x} = \frac{\partial N}{\partial \xi} + \frac{\partial N}{\partial \eta}$$

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial \xi}\left(\frac{\partial N}{\partial \xi} + \frac{\partial N}{\partial \eta}\right)\frac{\partial \xi}{\partial x} + \frac{\partial}{\partial \eta}\left(\frac{\partial N}{\partial \xi} + \frac{\partial N}{\partial \eta}\right)\frac{\partial \eta}{\partial x}$$

$$= \frac{\partial^2 N}{\partial \xi^2} + 2\frac{\partial^2 N}{\partial \xi \partial \eta} + \frac{\partial^2 N}{\partial \eta^2}$$

and the same type of computations for $\frac{\partial u}{\partial t}$, $\frac{\partial^2 u}{\partial t^2}$ gives

$$\frac{\partial u}{\partial t} = \frac{\partial N}{\partial \xi}c - \frac{\partial N}{\partial \eta}c$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 N}{\partial \xi^2}c^2 - 2\frac{\partial^2 N}{\partial \xi \partial \eta}c^2 + \frac{\partial^2 N}{\partial \eta^2}c^2$$

and so

we get the wave equation in the new variables ξ, η, v as follows

$$\frac{\partial^2 v}{\partial \xi \partial \eta} = 0 \quad , \forall_{\xi,\eta} \, ,$$

and this can be easily integrated, first in relation to ξ giving

$$\frac{\partial v}{\partial \eta} = G_1(\eta)$$

and then in relation to η to give

$$v = \int G_1(\eta) d\eta + f(\xi) .$$

As the integral in the righ-hand side is a function of η only we denoted by $G(\eta)$ (to simplify the notation) and we get

$$v(\xi, \eta) = F(\xi) + G(\eta)$$

and reverting the change of variables we obtain

$$u(x,t) = f(x+ct) + G(x-ct)$$

as wanted. ■

GMIT

Differential Equations and Applications

Session 15: Linear wave equation in 1-dim unbounded intervals: d'Alembert formula and propagation of waves

In the last session we observed that the solutions to the wave equation in a bounded interval obtained via fourier series can be written in the form

$$(*) \qquad u(x,t) = F(x+ct) + G(x-ct)$$

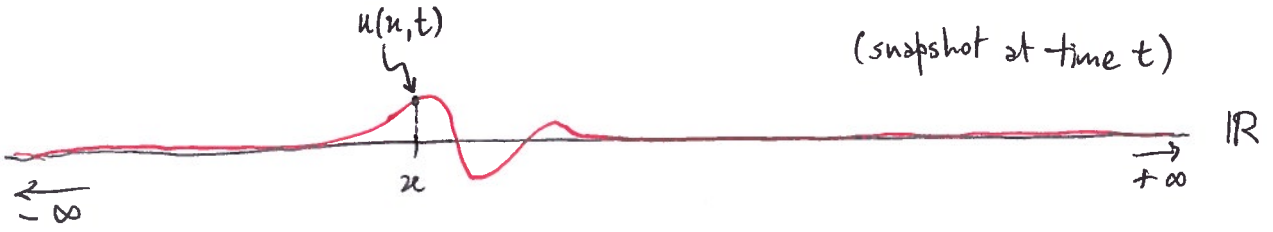
with F and G obtained in terms of fourier series.

We even conclude that (*) is still true if the wave equation is not valid only on a bounded interval: all solutions of

$$(**)_1 \qquad \frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \quad , \quad (x,t) \in \mathbb{R}^2$$

are of this type. The first problem that we need to solve in this last case is how are F and G in (*) related to the additional conditions we must impose on $(**)_1$.

Let us consider the wave equation $(**)_1$ as a model of a very, very, long rope whose length is so long that we assume it to be infinite



It seems natural to model the vibrations of such a rope by the wave equation $(**)_1$ and to fix the initial position and the initial velocity of the rope if we want to have a single solution to the equation.

So, we consider the <u>initial value problem</u>

$$(**)_2 \begin{cases} \dfrac{\partial^2 u}{\partial t^2} = c^2 \dfrac{\partial^2 u}{\partial x^2} & (x,t) \in \mathbb{R}^2 \\ u(x,0) = f(x) \\ \dfrac{\partial u}{\partial t}(x,0) = g(x) \end{cases}$$

where f and g are the initial configuration of the rope, and g the initial velocity.

We know that solutions of $(**)_1$ are of the type $(*)$, so let us use the initial conditions to find F and G:

$$f(x) = u(x,0) = F(x) + G(x)$$

$$g(x) = \frac{\partial u}{\partial t}(x,0) = cF'(x) - cG'(x)$$

Now, integrating the equation for g we get

$$\int_0^x g(s)\,ds = cF(x) - cG(x) + \alpha$$

where $\alpha = cG(0) - cF(0)$ is a constant.

So, we now have

$$F(x) + G(x) = f(x)$$

$$F(x) - G(x) = \frac{1}{c}\int_0^x g(s)\,ds - \frac{\alpha}{c}$$

Thus, summing these two we now have

$(x)_1$ $$F(x) = \frac{1}{2}f(x) + \frac{1}{2c}\int_0^x g(s)\,ds - \frac{\alpha}{2c}$$

and subtracting the above equations we get

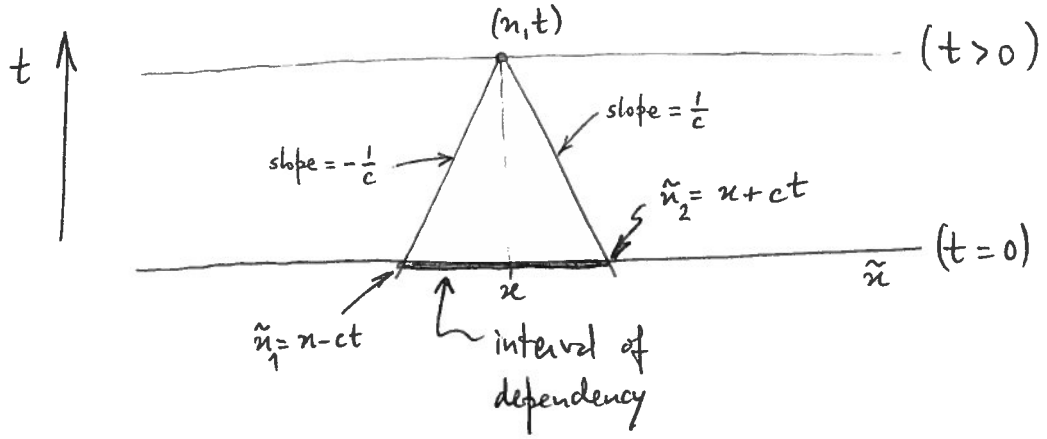
$(x)_2$ $$G(x) = \frac{1}{2}f(x) - \frac{1}{2c}\int_0^x g(s)\,ds + \frac{\alpha}{2c}$$

Hence, substituting $(x)_1$ and $(x)_2$ into $(*)$ we get the following expression for the solution of the initial value problem $(**)_2$:

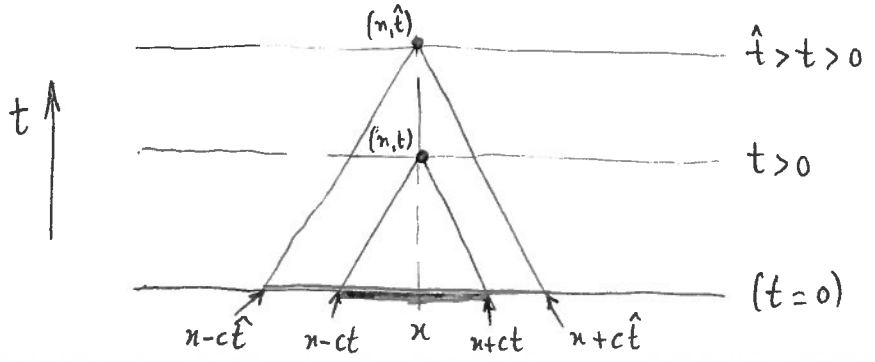
$(\dagger)$ $$\boxed{u(x,t) = \frac{1}{2}\left(f(x+ct) + f(x-ct)\right) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)\,ds}$$

This is D'Alembert's formula, and its analysis provides important information about properties of solutions to the wave equation, as we will see now.

formula (†) tell us that the value of $u(x,t)$ at a point (x,t) depends only on the initial position f on the points $x+ct$ and $x-ct$, and on the initial velocity g in all the points of the interval $[x-ct, x+ct]$. So, the value of u at (x,t) depends only on the initial conditions on $[x-ct, x+ct]$. This is called the <u>interval of dependency</u> of the point (x,t)

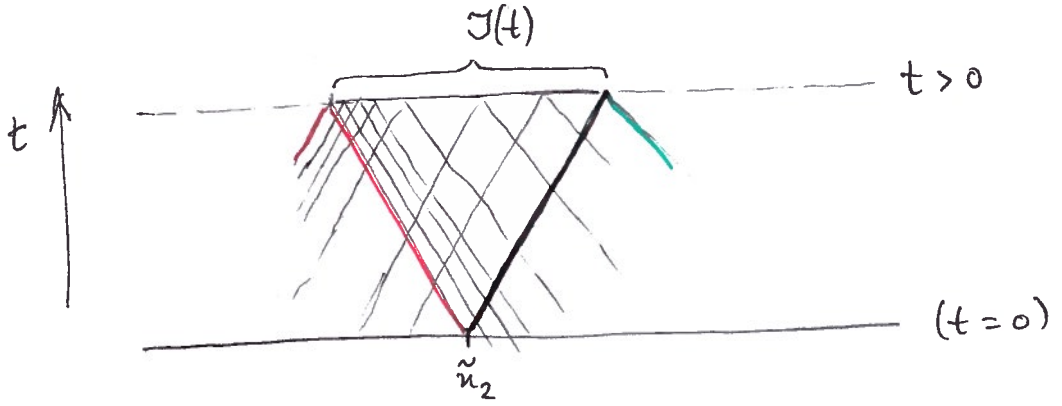


this means that whatever happens at $t=0$ to the right of $\tilde{x}_2 = x+ct$ and to the left of $\tilde{x}_1 = x-ct$ have absolutely no effect in the value of u at position x and time t; it can, of course, influence what happens at x at a later time $\hat{t}$:



We can reverse this study and instead of fixing a point (x,t) and asking what was the part of the initial data that influenced it, ask now what are the regions of the rope that are influenced by it at a later time.

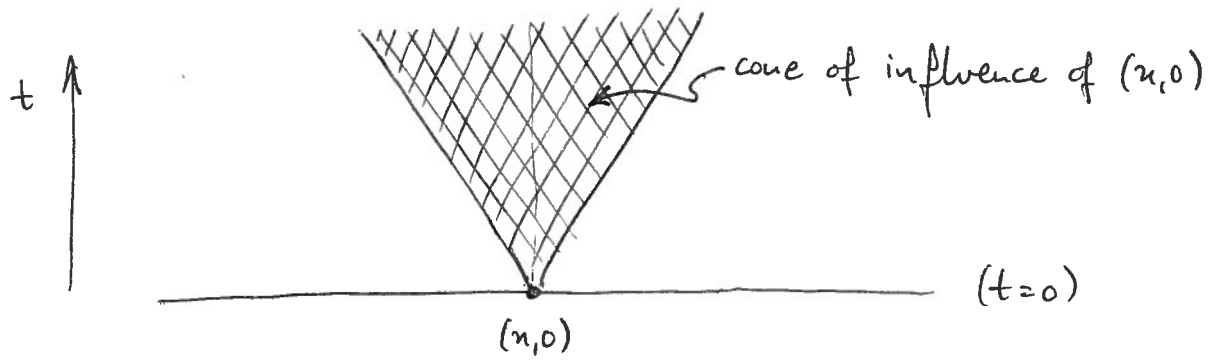
It is easy to understand what happens if we repeat the first plot considering only the point $\tilde{x}_2$ and adding all possible plots for which this point becomes inside the interval of dependency:



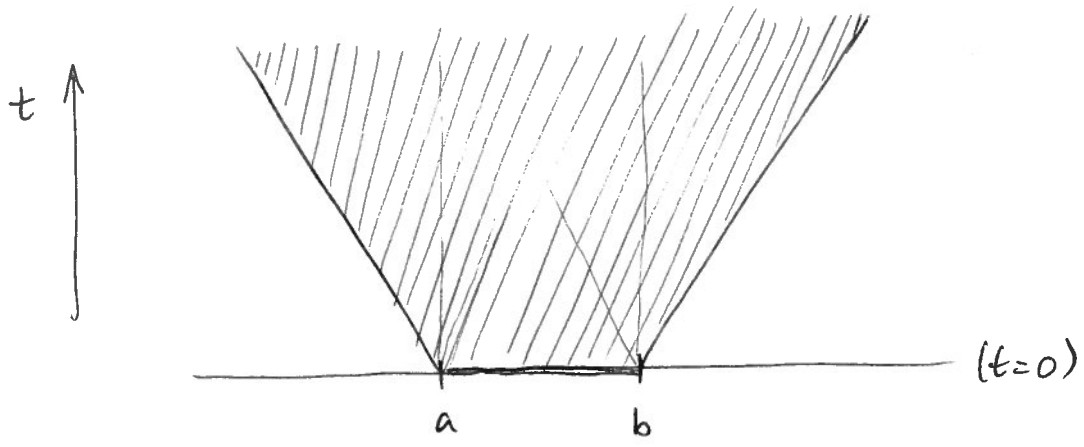
So, for all points $x \in J(t)$ the point $\tilde{x}_2$ is in their intervals of dependency, so what happens to the initial data f and g at the point $\tilde{x}_2$ influences the points in $J(t)$, at time t, and only those points are influenced by f and g at $\tilde{x}_2$ at that time.

If we consider smaller times than t the region $J(t)$ will be correspondingly smaller, and the reverse for larger times.

Collecting what happens in all times $t>0$ we have the so called region of influence, or cone of influence, of the point $(x,0)$ represented next:

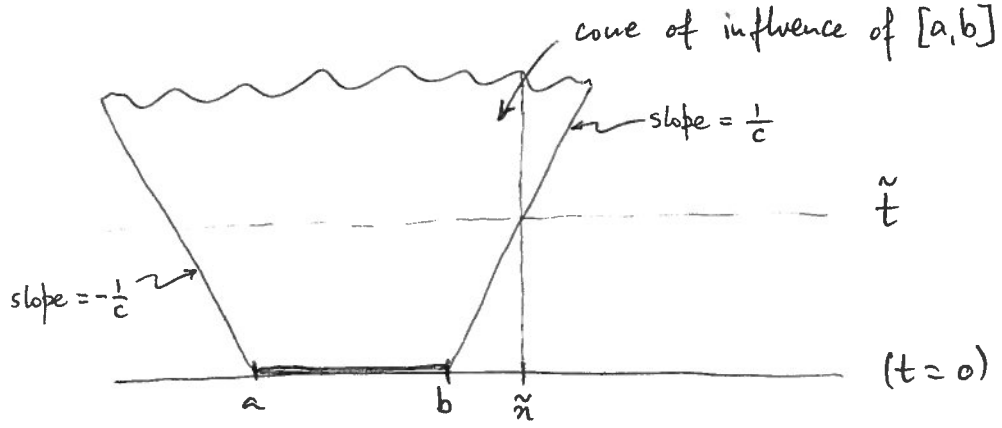


It should be obvious that the cone of influence of the initial data on the interval $[a,b]$ is the union of the cones of influence of the points $(x,0)$ with $x\in[a,b]$, like in the next figure



An alternative way to interpret these results is to observe that they show that solutions of the wave equation given by D'Alembert formula correspond to the propagation of

the information contained in the initial data to the whole rope with a velocity c and $-c$. In fact, suppose the initial conditions are concentrated in an interval $[a,b]$, i.e., $f = g = 0$ in $\mathbb{R}\setminus[a,b]$, and consider a point of the rope at location $\tilde{x} > b$.
From what we saw before we have



and so the point $\tilde{x}$ enters the cone of influence of $[a,b]$ only at a time $\tilde{t}$ given by $\tilde{x} = b + c\tilde{t} \iff \tilde{t} = \frac{\tilde{x}-b}{c}$, and only after that time can the initial data in $[a,b]$ have some influence in the value of $u(\tilde{x},t)$.

this means that solutions to the wave equation propagate signals at a velocity c (and $-c$, as we would have concluded if we had put $\tilde{x} < a$). Of course that we already knew this from $(*)$ which was proved in the last session.

Let us consider one interesting case of $(\dagger)$ correspondent to

the situation where we perturb the rope in a bounded region but do not impress initial velocity, i.e.: $g(x) = 0$ everywhere.

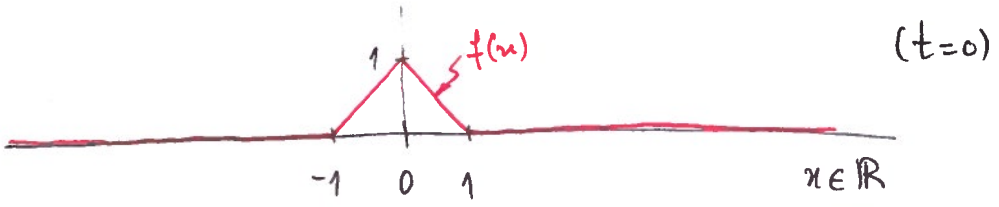
Consider, then, $f(x) = 1 - |x|$ for $x \in [-1, 1]$, and zero outside, and $g(x) \equiv 0$. Then the solution given by D'Alembert formula for the initial value problem

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} = \dfrac{\partial^2 u}{\partial x^2} & , \ (x,t) \in \mathbb{R}^2 \\ u(x,0) = f(x) & , \ x \in \mathbb{R} \\ \dfrac{\partial u}{\partial t}(x,0) = 0 & , \ x \in \mathbb{R} \end{cases}$$

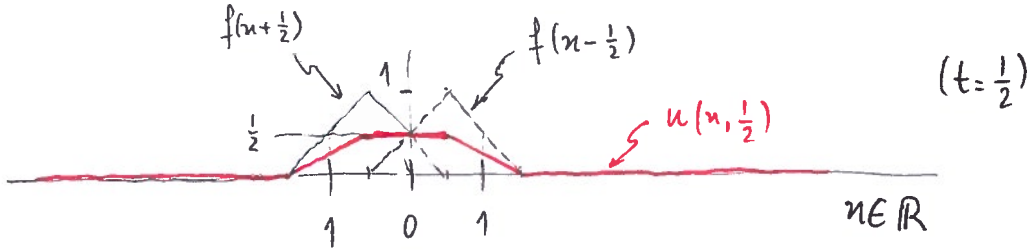
is

$$(\diamond) \qquad u(x,t) = \frac{1}{2}\Big(f(x+t) + f(x-t)\Big)$$

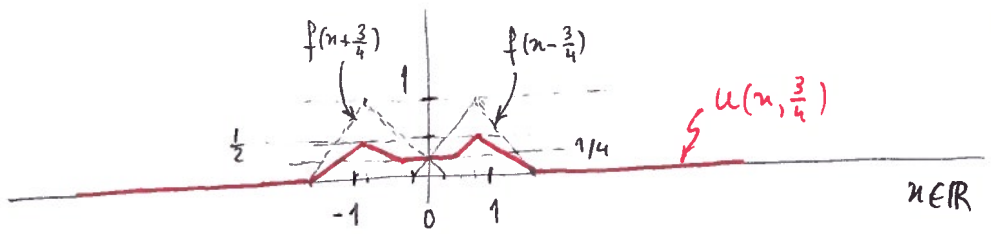
Let's see, graphically, how $u(x,t)$ evolves with time.
At $t = 0$ we have $u(x,0) = \frac{1}{2}(f(x) + f(x)) = f(x)$ [obviously!]



Now consider the shape of u, given by $(\diamond)$, at time $t = \frac{1}{2}$

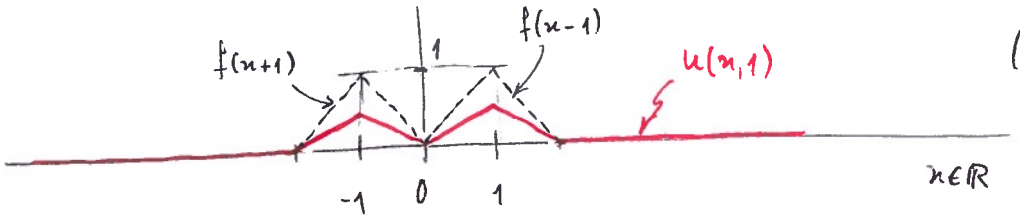


Proceed a little more with time, to $t = \frac{3}{4}$



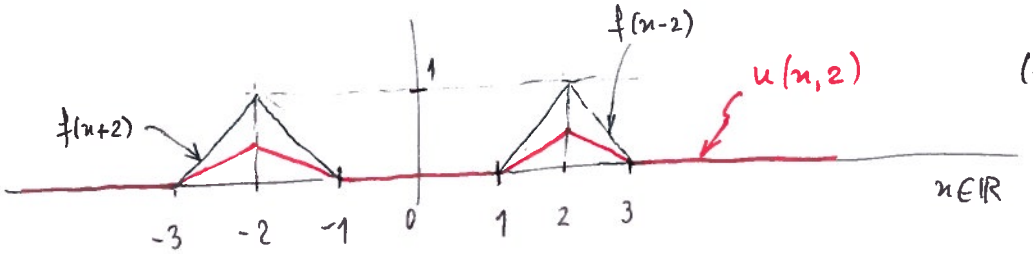
$(t = \frac{3}{4})$

At a still later time, $t = 1$,



$(t = 1)$

and still later, $t = 2$,



$(t = 2)$

This example clearly shows the propagation of the initial disturbance f to the position of the rope being propagated with constant velocity c and $-c$, as we had concluded before it must happen.

It also shows a very interesting phenomenon which is the following: once the initial perturbation passes through a certain point x of the rope, that point remains at rest.

this may seem natural (and it is...) but it is not what happens if we do the same study in dimension 2 : an initial perturbation of the surface of water in a pond, at a small region passes points away from that region after some time but the water does not get back to rest but it keeps oscillating : try throwing a stone to a pond and see what happens !

In dimension 3 we have again the situation of dimension 1 ; if the perturbation passes a location that point remains at rest ; and this is the reason we can communicate with sound (speak !) without hearing over and over the same thing , like a multiple echo.

Let us now consider the case of a semi-infinite rope, i.e., a rope that can be modelled by points in $\mathbb{R}_0^+ = [0, \infty)$. In this case we need to know what is happening at the point $x=0$ of the rope. It can be fixed for all time, it can move as when someone shakes the rope

in any case an additional boundary condition must be provided at $x=0$. the initial-boundary value problem is now

$$(\Delta) \quad \begin{cases} \dfrac{\partial^2 u}{\partial t^2} = c^2 \dfrac{\partial^2 u}{\partial x^2} \;, & (x,t) \in \mathbb{R}^{2+} \\ u(0,t) = h(t) \;, & t > 0 \\ u(x,0) = f(x), \; \dfrac{\partial u}{\partial t}(x,0) = g(x) \;, & x > 0 \end{cases}$$

Proceeding as before, we know that solutions of the wave equation are of the type

$$u(x,t) = F(x+ct) + G(x-ct)$$

where, using the initial conditions as before, we now have,

$$F(x) = \frac{1}{2} f(x) + \frac{1}{2c} \int_0^x g(s)\,ds + k \;, \quad x > 0$$

$$G(x) = \frac{1}{2} f(x) - \frac{1}{2c} \int_0^x g(s)\,ds - k \;, \quad x > 0$$

However, note that now these expressions are valid only for $x > 0$. this means that we do not know what is $G(x-ct)$ if t is so large that $x - ct < 0$.

The only hope to get that kind of information is to use the only information in (Δ) that was not yet used : the boundary condition $u(0,t) = h(t)$.

$$h(t) = u(0,t) = f(ct) + G(-ct), \quad \forall t > 0$$

thus, denoting ct by y, we have, for all $y > 0$,

$$\begin{aligned} G(-y) &= h\left(\frac{y}{c}\right) - f(y) \\ &= h\left(\frac{y}{c}\right) - \frac{1}{2} f(y) - \frac{1}{2c} \int_0^y g(s)\,ds - k. \end{aligned}$$

Now we can complete the answer : if $x - ct > 0$ the solution $u(x,t)$ is given by

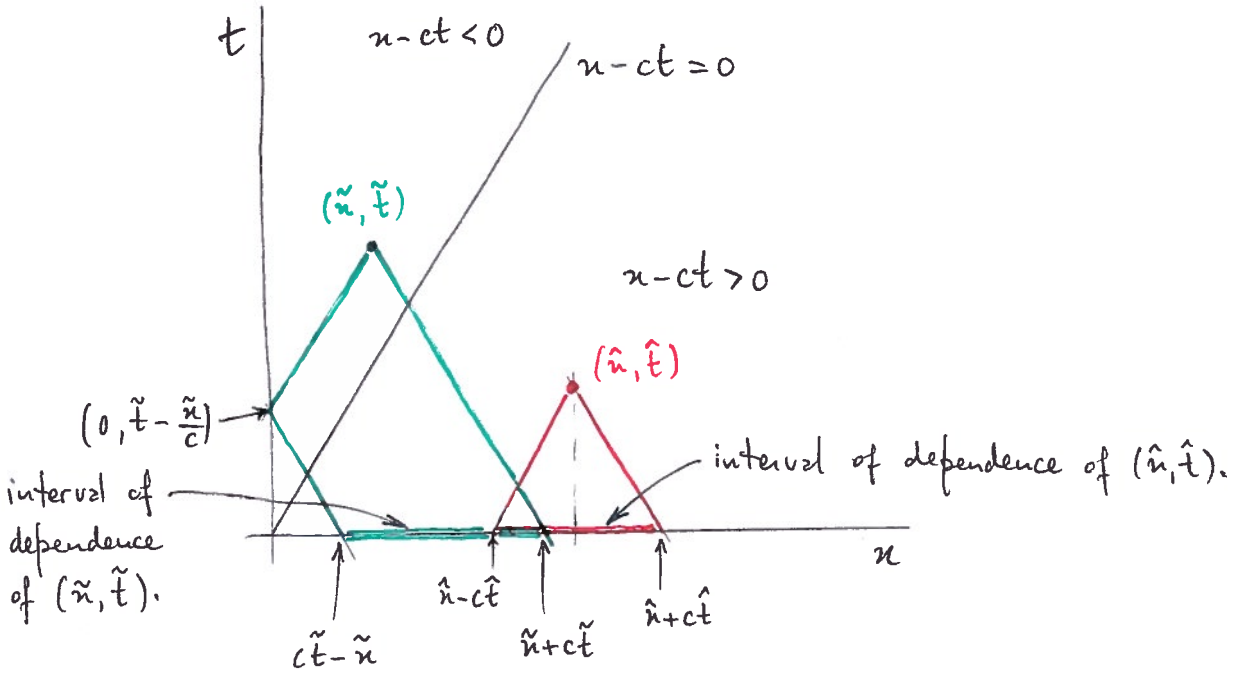
$$(\tilde{\Delta})_1 \qquad u(x,t) = \frac{f(x+ct) + f(x-ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} g(s)\,ds .$$

However, if $x - ct < 0$ then $u(x,t)$ is

$$(\tilde{\Delta})_2 \qquad u(x,t) = \frac{f(x+ct) - f(ct-x)}{2} + \frac{1}{2c} \int_{ct-x}^{x+ct} g(s)\,ds + h\left(t - \frac{x}{c}\right)$$

It is interesting to analyse this solution in terms of the concepts of influence and dependence introduced

earlier in the infinite rope context. This is easy to do by reference to the following plot:



Consider a point of the rope at position $\hat{x}$. If the time $\hat{t}$ is so small that $\hat{x} - c\hat{t} > 0$, i.e., if $\hat{t} < \frac{\hat{x}}{c}$, then at that point the fact that the rope ends at $x=0$ is not felt: for that position and at those times, it is as if the rope were infinite.

However, if, for some point $\tilde{x}$ of the rope we consider $\tilde{t}$ sufficiently high so that $\tilde{x} - c\tilde{t} < 0$, i.e., $\tilde{t} > \frac{\tilde{x}}{c}$, then $(\tilde{x}, \tilde{t})$ is influenced by what happens at the origin of the rope $x=0$ at time $t = \tilde{t} - \frac{\tilde{x}}{c}$, and at this point the rope has a behaviour which is naturally

influenced by h and by what happens at $t=0$ at the point $x = c\tilde{t} - \tilde{x}$.

Before we finish the study of the wave equation in an infinite or semi-infinite line we observe that, if we require the solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

to be a "classical" solution, i.e., if we require the function f to be twice differentiable in the variables x and t, symbolically $u \in C^2_x \cap C^2_t$, then we need to use initial conditions $f \in C^2$ and $g \in C^1$, and, in the case of semi-infinite rope, $h \in C^2$.

These are conditions that are too restrictive and are not fullfield in many cases of interest in applications.

This motivates the need to consider a notion of solution that requires less differentiability of u.

How can this be? the idea arises from the following computation: suppose we take a function $\varphi \in C^\infty_0(\mathbb{R}^2)$, i.e., a function that is C^∞ and has compact support, which means it is equal to zero outside some closed and bounded set of $\mathbb{R}^2$. Multiply the wave

equation by $\varphi(x,t)$ and integrate over $\mathbb{R}^2$. We obtain

$$(*) \qquad \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x,t)\frac{\partial^2 u}{\partial t^2}(x,t)\,dt\,dx = c^2 \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x,t)\frac{\partial^2 u}{\partial x^2}(x,t)\,dt\,dx$$

Now, if we suppose $u \in C_t^2(\mathbb{R}^2)$, we can integrate by parts in t the integral on the left-hand side of $(*)$ twice and, because φ has compact support, we get

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x,t)\frac{\partial^2 u}{\partial t^2}(x,t)\,dt\,dx = \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{\partial^2 \varphi}{\partial t^2}(x,t)\,u(x,t)\,dt\,dx$$

Now integrating by parts twice in x the integral in the right-hand side of $(*)$ we get

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x,t)\frac{\partial^2 u}{\partial x^2}(x,t)\,dt\,dx = \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{\partial^2 \varphi}{\partial x^2}(x,t)\,u(x,t)\,dt\,dx$$

Substituting all this into $(*)$ we get

$$(**) \qquad \int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial^2 \varphi}{\partial t^2} - c^2\frac{\partial^2 \varphi}{\partial x^2}\right) u(x,t)\,dt\,dx = 0.$$

Now, this expression only includes $u(x,t)$, not its derivatives, which have been "moved" to the function $\varphi(x,t)$.

this means that (*) makes sense even if u is only continuous (actually, it makes sense even for functions u that are much less regular than continuous.

This is the idea behind the notion of "weak solution" of the wave equation: we say that a function $u: \mathbb{R}^2 \to \mathbb{R}$ is a <u>weak</u>, or <u>generalized</u>, <u>solution</u> of the wave equation

$$\frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \quad , (x,t) \in \mathbb{R}^2$$

if, for all $\varphi \in C_0^\infty(\mathbb{R})$, it holds the equality

$$\iint_{\mathbb{R}^2} u(x,t) \left(\frac{\partial^2 \varphi}{\partial t^2} - c^2 \frac{\partial^2 \varphi}{\partial x^2} \right) dt\, dx = 0,$$

and u^2, $\left(\frac{\partial u}{\partial t}\right)^2$, and $\left(\frac{\partial u}{\partial x}\right)^2$ are integrable in every compact subset of $\mathbb{R}^2$.

In fact, in order that this new definition of solution to be useful, we need a better definition of integral than that used in usual calculus courses (the Riemann integral) and a more general differential calculus, obtained by a similar generalization of the notion of derivative. These generalized notions are the ones used in current research in PDEs, but require much more time to introduce than what we have in this course.

GMIT

Differential Equations and Applications

Session 16: Nonlinear waves (Burgers' equation): shocks and the need for a weaker notion of solution.

In the last session we saw the need to a more general notion of solution for the wave equation due to problems arising from the lack of regularity of the initial data and the fact that the wave equation does not regularize its solutions as time progresses.

In fact, this problem is much more relevant when the equations are non-linear (as most of truly realistic mathematical models are) where, in many cases, solutions cease to be differentiable, or even continuous, at some time and location, even if the initial data is completely smooth (say C^∞).

In this final session we will consider one such case, the Burgers' equation which is not only the easiest equation exhibiting this kind of behaviour but it is also very useful in many applications.

We recall the concept of conservation law in a one dimensional setting that we treated previously in session 9.

Consider a physical quantity distributed in a one dimensional medium such that, at position x and instant t, its concentration is denoted by $u(x,t)$. Then, the integral

$$\int_{x_1}^{x_2} u(x,t)\,dx$$

represents the total amount of that physical quantity. If there are no sources or sinks the change in this total amount is determined by the flux through the border of the interval; if we call the flux positive if it is from left to right, and negative otherwise, then

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)\,dx = -q(u(x_2,t)) + q(u(x_1,t))$$

where $q = q(u)$ is the flux.

As we already saw in session 9 we can use the fundamental theorem of calculus (if q and u are sufficiently smooth) to write the right-hand side

as $-\int_{x_1}^{x_2} \frac{\partial q}{\partial x}(u(x,t))\,dx$ and so write

$$\int_{x_1}^{x_2} \left(\frac{\partial u}{\partial t}(x,t) + \frac{\partial q}{\partial x}(u(x,t)) \right) dx = 0$$

and, due to the arbitrariness of the interval $[x_1, x_2]$, (and the presumed regularity of the integrand functions) this implies u must satisfy the conservation law

$$(*) \qquad \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} q(u) = 0 .$$

As we saw in session 9, we need to have a <u>constitutive relation</u> between q and u in order to be able to study this equation. There, we were studying the process of diffusion, for which q was given by Fick's, or Fourier's, law.

Let us now consider a different constitutive relation: suppose we take q a linear function of u,

$$(**) \qquad q(u) = v u .$$

this means that the flux q is determined only by the value of u, and corresponds, physically, to the existence of a convection behaviour in the quantity

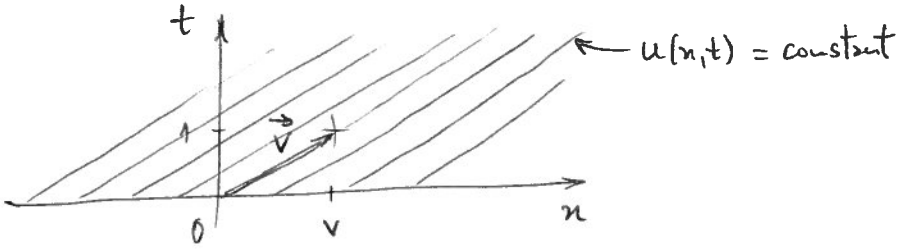
being studied. Let us, then, consider (*) with (**). the PDE is

$$(+)_1 \qquad \frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0 \qquad x \in \mathbb{R},\ t \in \mathbb{R}$$

Observe that $(+)_1$ can be written as

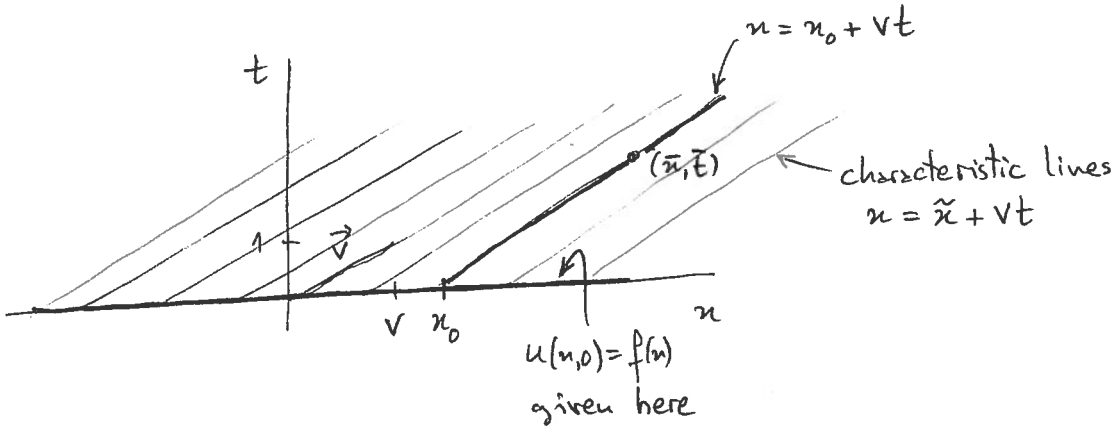
$$(+)_2 \qquad (\nabla u) \cdot \vec{v} = 0$$

where $\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial t} \right)$ is the gradient of u, and $\vec{v} = (v, 1)$. Equation $(+)_2$ has an easy geometric interpretation: since $\nabla u(x,y)$ is a vector orthogonal to the level sets of $u(x,y)$, equality $(+)_2$ states that the level sets of $u(x,y)$ are parallel to $\vec{v}$, i.e., along straight



lines parallel to $\vec{v}$ the function $u(x,t)$ is constant. The lines parallel to $\vec{v}$, meaning: the lines of equation $(x,t) = (\tilde{x}, 0) + \vec{v}t \iff x = \tilde{x} + vt$, are called the characteristics of $(+)_1$, and what we have just concluded is that along characteristic lines

the value of $u(x,t)$ is constant! Hence, if we know it at one point of each of these lines we know it everywhere! Complementing equation $(+)_1$ with an initial condition $u(x,0) = f(x)$, or any other condition given along a line not parallel to the characteristics, allow us to compute the solution to $(+)_1$ very easily:



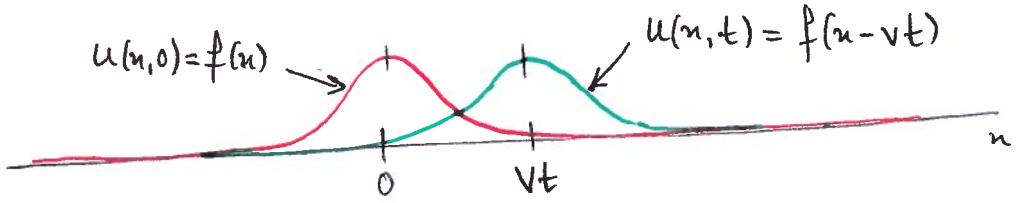
At $(\bar{x},\bar{t})$ the value of u is $u(\bar{x},\bar{t}) = u(x_0,0) = f(x_0)$, where x_0 is the abscissa of the point where the characteristic intersects the x-axis: from $\bar{x} = x_0 + v\bar{t}$ we get

$$x_0 = \bar{x} - v\bar{t}.$$

So $u(\bar{x},\bar{t}) = f(\bar{x} - v\bar{t})$, and since $(\bar{x},\bar{t})$ is arbitrary we can conclude that the solution of $(+)_1$ with initial condition $u(x,0) = f(x)$ is

$(+)_3$ $$u(x,t) = f(x - vt), \quad \forall x,t \in \mathbb{R}$$

This is, of course, the reason equation $(+)_1$ is called the drift equation, or the transport equation: the initial condition f is transported by translation given by $x \mapsto x - Vt$



Let us consider what looks like just a slightly more complicated equation, but really is a much more interesting, and difficult, equation: take the flux q to be the nonlinear function

$$q(u) = \frac{1}{2}u^2. \qquad (***)$$

Thus, the conservation law $(*)$ now becomes

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \quad , \quad x \in \mathbb{R}, t > 0. \qquad (*)_1$$

This is called the inviscid Burgers' equation, an equation that first appeared in fluid mechanics and it is, together with some modifications, extremely important in mathematics

Observe that $(x)_1$ looks very much like the transport equation $(t)_1$ but now the transport velocity V changes because it is now equal to $u(x,t)$ and so the "transport" should be faster if u is bigger and slower if u is smaller. this is the intuition, but is it true?

Proceeding as in the transport equation $(t)_1$ we look for a line $x(t)$ in the (x,t)-plane such that the solution $u = u(x,t)$ is constant along that line. this will be a characteristic line for $(x)_1$, just like the straight line $x(t) = x_0 + Vt$ was the characteristic line for $(t)_1$.

So we want to find a line $x(t)$ such that $u(x(t),t)$ is constant, independent of t. this means that

$$\frac{d}{dt} u(x(t),t) = 0$$

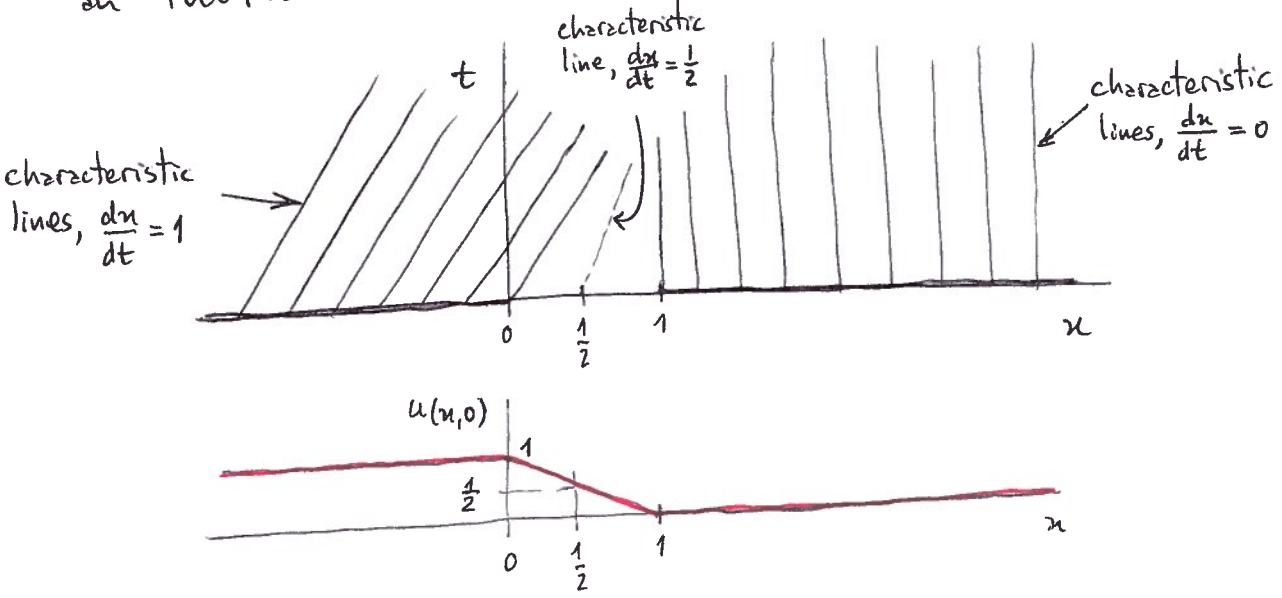
which, by the chain rule, gives

$$(x)_2 \qquad \frac{\partial u}{\partial t} + \frac{dx}{dt}\frac{\partial u}{\partial x} = 0$$

But now compare $(x)_2$ with $(x)_1$. If we choose $x(t)$ as a solution of

$$\frac{dx}{dt} = u(x,t)$$

with $u(x,t)$ a solution of $(x)_1$, then $(x)_2$ is satisfied which means that $u(x,t)$ is constant along this line. So, if u is constant along that line and $\frac{dx}{dt} = u(x,t)$, then $\frac{dx}{dt}$ is constant: the characteristic lines have constant slope: they are straight lines just like in the case of the transport equation but the slopes now are not all equal (to v) but vary with u, and so now depend on its value at, say, an initial time $t = 0$. Suppose the following:



This means that the initial value problem

$$(x)_3 \quad \begin{cases} \dfrac{\partial u}{\partial t} + u \dfrac{\partial u}{\partial x} = 0 & x \in \mathbb{R},\ t > 0 \\ u(x,0) = f(x) \end{cases}$$

with $f(x) = \begin{cases} 1 & \text{if } x \leq 0 \\ 1-x & \text{if } x \in (0,1) \\ 0 & \text{if } x \geq 1 \end{cases}$

has solution given by

$(x)_4$ $$u(x,t) = f(x - t\,u(x,t)).$$

Note the similarity with $(+)_3$. However, this is not yet the end of the story: the equality $(x)_4$ is an equation for $u(x,t)$ and we need to guarantee that $(x)_4$ has a solution. This can be done using the implicit function theorem if f is differentiable.

In fact, writing $(x)_4$ in the form

$$F(u,x,t) = 0$$

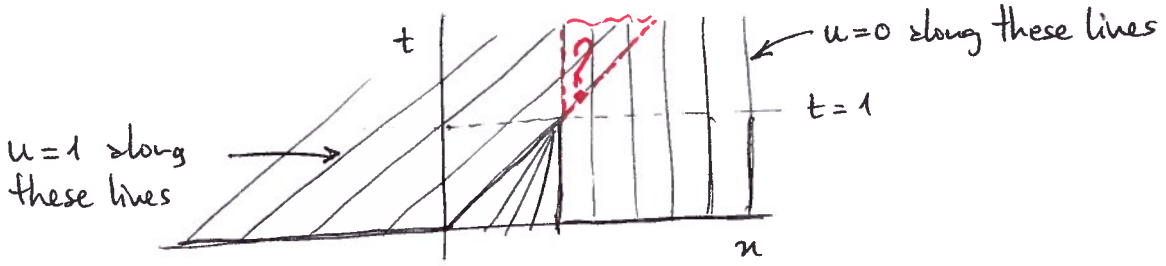
with $F(u,x,t) = u - f(x - tu)$, the implicit function theorem guarantees the existence of a function $u = u(x,t)$ for (x,t) close to $(x,0)$ if $\frac{\partial F}{\partial u} \neq 0$. But

$$\frac{\partial F}{\partial u} = 1 + f' \cdot t$$

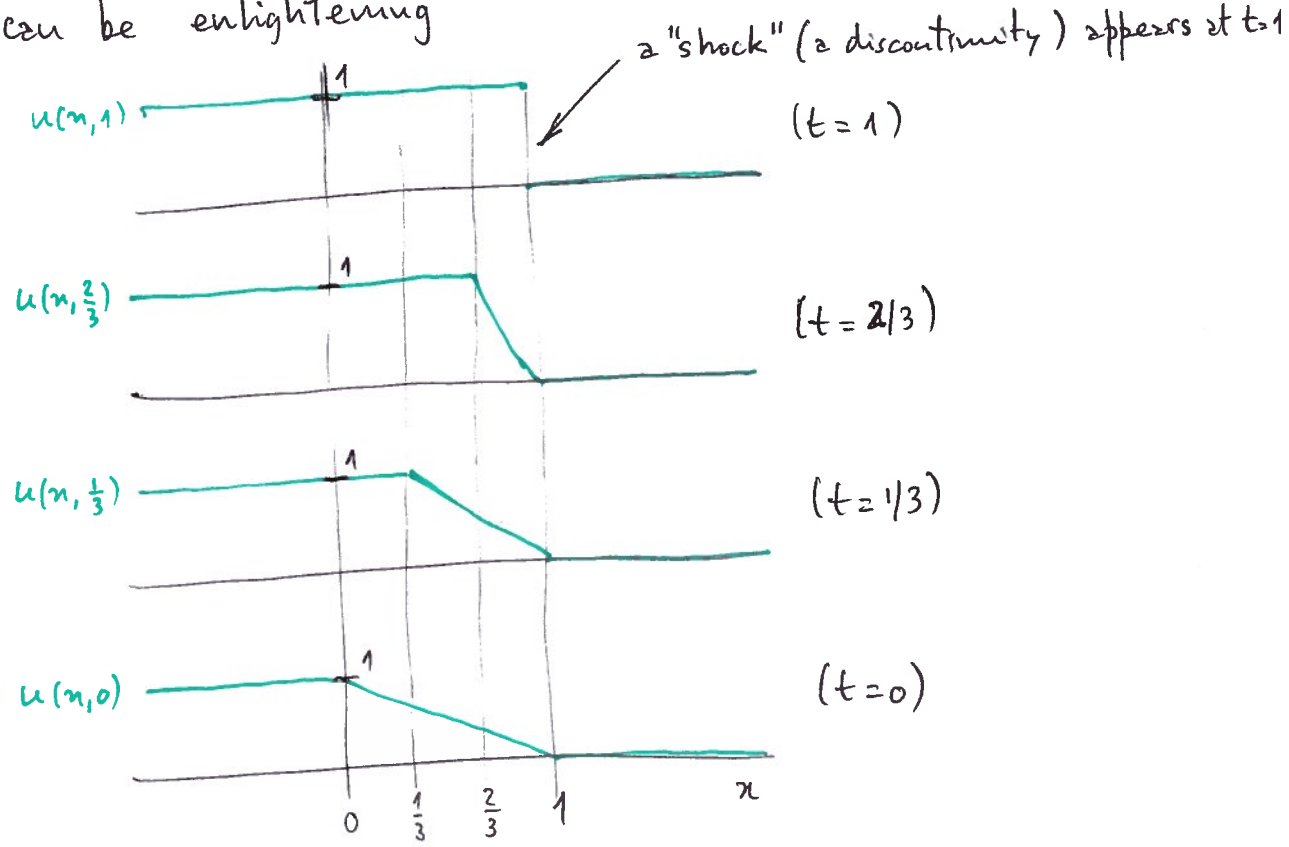
is always different from zero (at least) for small enough $t > 0$.

Can the existence be extended for larger values of t? Inspecting the last plot we see imediatly

that we must have problems at $t=1$ (for that example):



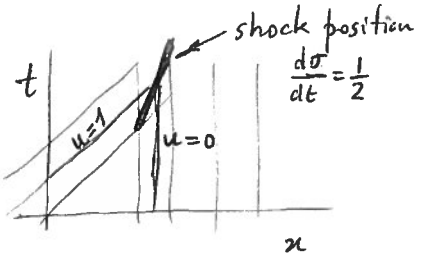
In the region in red our construction cannot be valid because it is telling that $u(x,t)$ is multivalued at each point! Looking to what is happening at several values of t can be enlightening



Note that after $t=1$ the equation "wants" to push $u=1$ to the right with velocity $u=1$, but the part of the solution with $u=0$ do not move (because its velocity of translation is $u=0$); so we have a problem: a

collision of wills, if we can speak like this, i.e., a shock! After $t=1$ we need to have a more general notion of solution, allowing for the existence of discontinuities (shocks) and we have to establish a theory of shock dynamics, i.e., to decide how is the shock to move for $t>1$. this requires some additional assumptions to be made after the shock. For example, if we assume that the conservation law is to be satisfied after the onset of the shock then the shock propagates with a velocity that is equal to the mean of the values of u on both sides of the shock $x=\sigma(t)$:

$$\frac{d\sigma}{dt} = \frac{u^-(t) + u^+(t)}{2}.$$

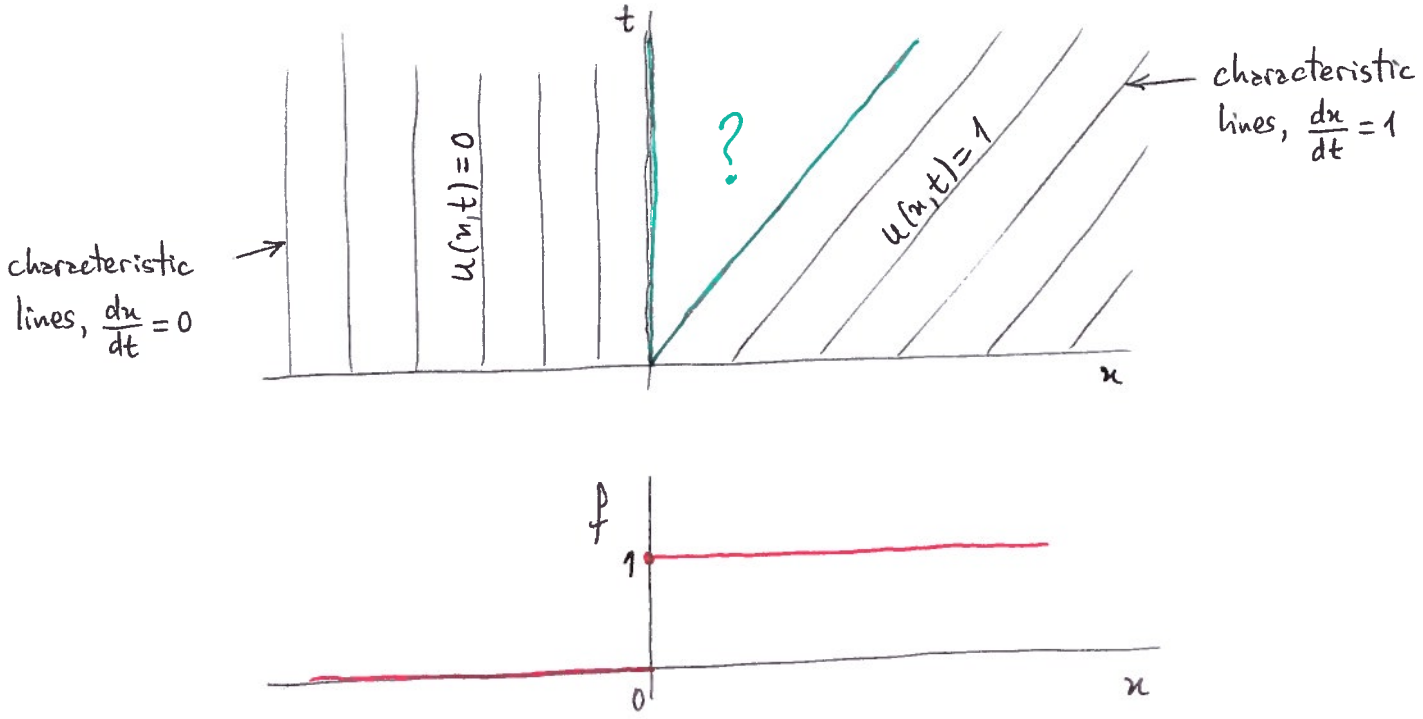


this is called the Rankine-Hugoniot condition, determining the position of the shock after its formation.

Another consequence of the application of the implicit function theorem to $(X)_4$ is that if the initial condition f is an increasing function of x, then $f'(x)>0$ and so $\frac{\partial F}{\partial u} > 1 > 0$ for all $t>0$ implying that $u(x,t)$ is defined for all $(x,t)\in \mathbb{R}\times\mathbb{R}^+$.

Now we have a different kind of problem. To see what it is let's go back to the geometric picture with the characteristic lines, along which the solution $u(x,t)$ is constant, and consider now the initial condition

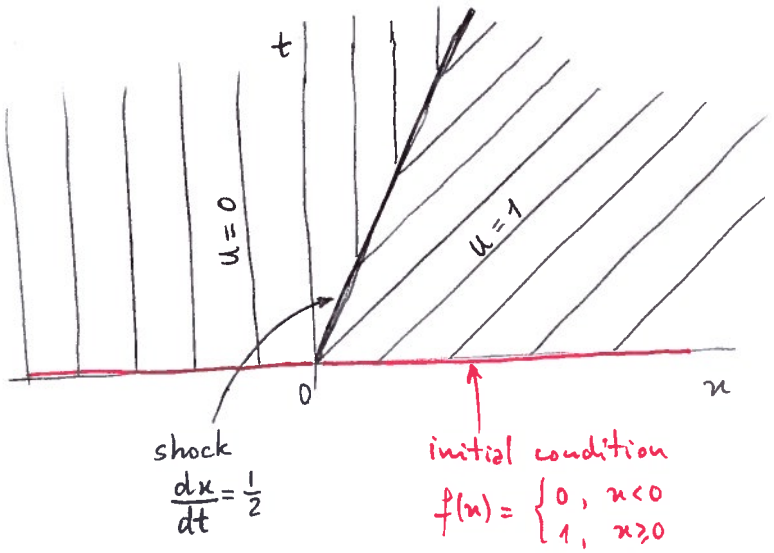
$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$



In the region in green the partial differential equation under consideration (Burgers' equation $(X)_3$) does not inform us of what is going on!

So, we are free to <u>define</u> $u(x,t)$ in the region in green in such a way that satisfies $(X)_3$ and "glues" well with the $u(x,t)$ in the rest of $\mathbb{R} \times \mathbb{R}^+$. Let's see:

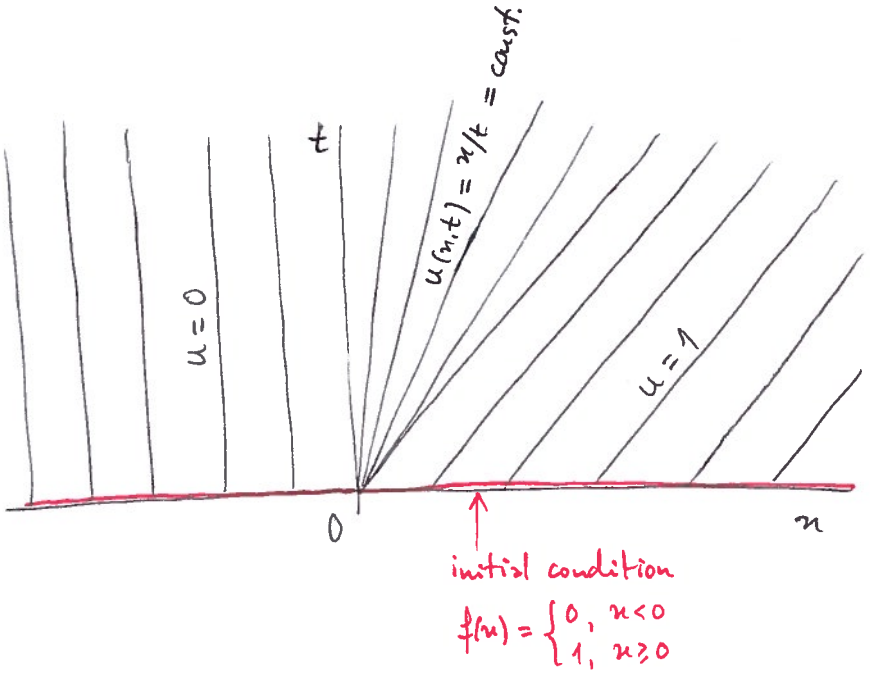
One possibility is to construct a shock in the green region separating $u=0$ from $u=1$. To keep the conservation law valid we need the shock to satisfy the Rankine-Hugoniot condition and so its slope should be $\frac{d\sigma}{dt} = \frac{1}{2}$. In this case we would have



and the solution to Burgers' equation with this initial condition will be

$$(\nabla) \qquad u(x,t) = \begin{cases} 0, & \text{if } x < \frac{t}{2} \\ 1, & \text{if } x > \frac{t}{2} \end{cases}$$

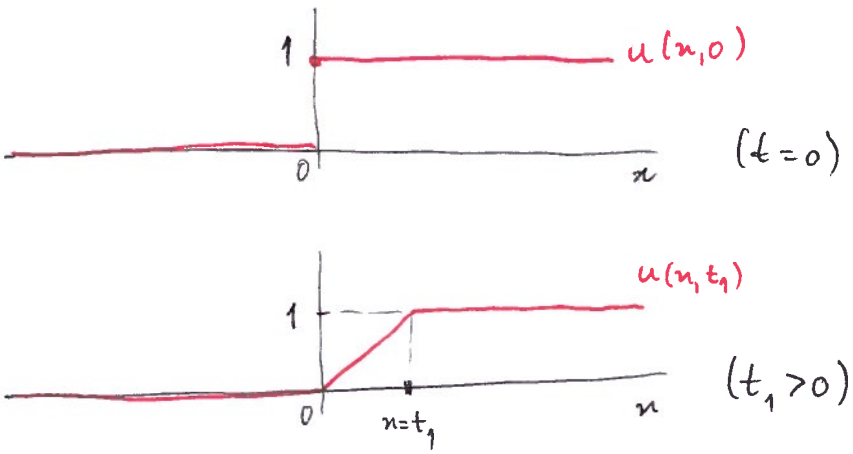
But (∇) is not the only solution to define a solution in the region painted green. Another possibility that looks reasonable is to fill the region in green with characteristic straight lines with slopes changing continuously from $\frac{d\sigma}{dt} = 0$ to $\frac{d\sigma}{dt} = 1$, as in what follows:

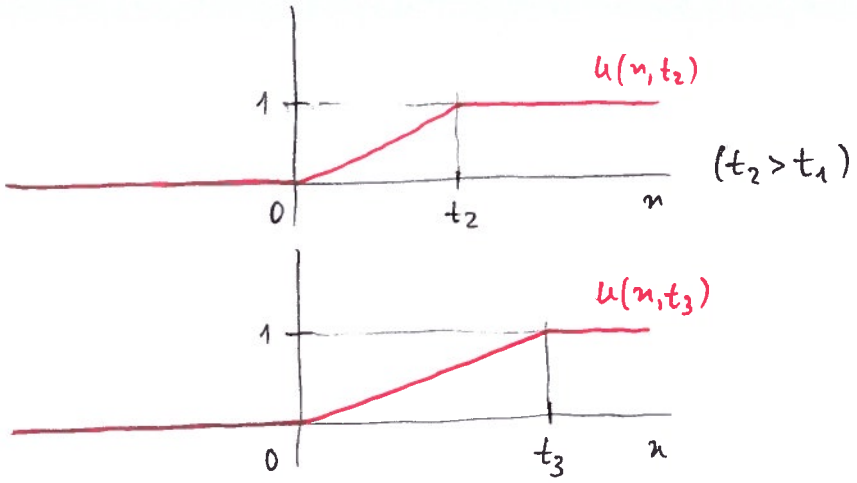


and the solution of Burgers' equation with this initial condition is

$$u(x,t) = \begin{cases} 0 & \text{if } x<0 \\ x/t & \text{if } 0 \leq x < t \\ 1 & \text{if } x \geq t. \end{cases}$$

This type of solution is called a <u>rarefaction wave</u>, a name coming from fluid mechanics and that we can understand by ploting at the plots of the solutions in several increasing instants of time





So, the region to the right of $x=0$, which initially had a value of $u=1$ is becoming, progressively, with lower and lower values of u: the quantity u is being rarefied.

The fact that there are (at least) two solutions poses an important problem for applying the differential equation to modelling of scientific or technical situations: which of the solutions describes the phenomenon? which shall we use? This is a very important question, the answer of which cannot be found in the differential equation itself (since both solutions satisfy the differential equation!)

The decision as to which solution should be considered valid is suggested from the physical setting for which the PDE is considered a model. In particular, in the case of Burgers' equation, it comes from an analogy with gas dynamics, in which the entropy of

the system increases across a shock curve. There are a number of entropy criteria that translate this idea to rigorous mathematical terms. These criteria allow us to decide which shocks are "valid" and which should be "rejected", and they are automatically satisfied by rarefaction waves, which are continuous functions.

One important criterion is Lax entropy condition stating that if the flux function $q(u)$ is either convex or concave (which is the case of $q(u) = \frac{1}{2}u^2$), then the solution u satisfies Lax entropy condition if across a shock $x = \sigma(t)$ we have

$$q'(u^+) < \frac{d\sigma}{dt} < q'(u^-).$$

We can easily observe that the solution (∇) does not satisfy Lax entropy condition because $q'(u) = u$ and so, as to the right of the shock we have $u^+ = 1$ and to the left $u^- = 0$ and $\frac{d\sigma}{dt} = \frac{1}{2}$, what is true for that shock is

$$q'(u^+) = u^+ = 1 > \frac{d\sigma}{dt} > 0 = u^- = q'(u^-)$$

which is exactly the opposite of Lax's condition. So, solution (∇) should be rejected as a valid solution for being unphysical according to this entropy

criterium.

the rigorous deduction of these criteria of selecting the "right" shock, or the "right" solution (Rankine-Hugoniot, Lax, Liu, etc.) would take another module to present; together with many other important results in the theory of non-linear transport equations it must be left to another occasion.