Otmar Scherzer

COMAT 4

Lecture Notes – Preliminary Draft

Summer 2015

Computational Science Center
University of Vienna
A-1090 Vienna, Austria

Chapter 1

Examples of ODEs and PDEs

1.1 Simple Examples of Ordinary Differential Equations and Separation of Variables

1.1.1 Movement of a Falling Body

We describe the movement of a vertically falling body. Then its position at time t is described by its height $h(t)$.

Newton's second law of motion implies that the acceleration of the body, that is, the change of its speed, is proportional to the forces acting on the body. That is,

$$F = ma\,,$$

where F denotes the forces, m denotes the mass of the body, and a is its acceleration.

Now, the acceleration is the change of the speed, which is itself the change of the position of the body. Therefore

$$F = m\ddot{h}(t)\ . \tag{1.1}$$

We still have to model the acting forces.

The main force F is gravity, which, for small heights h, equals approximately mg, where $g \approx 9.81\text{m/s}^2$ is the gravitational acceleration at the earth's surface. The gravitation is acting downwards such that from (1.1) it follows that

$$m\ddot{h}(t) = -mg\ .$$

If either the body is very light or it is falling fast, it is necessary to take into account air friction, which slows down the fall of the body. One possibility is to model air friction as a force proportional to the square of the body's velocity. Because friction always works against the current movement, the sign of the corresponding force will be opposite to the sign of $\dot{h}$. Thus we obtain the refined model

$$m\ddot{h}(t) = -c\,\mathrm{sgn}(\dot{h}(t))\,\dot{h}(t)^2 - mg\,,$$

where c is some material constant describing the drag of the body.

In order to obtain a complete description of the movement of the body, we will need in addition a description of the state of the body at some initial time t_0, where we begin our considerations. More precisely, we will need its initial position h_0 and its initial velocity v_0. Then, assuming this model is correct, the movement of the body is completely described by the *differential equation*

$$\begin{aligned} m\ddot{h}(t) &= -c\,\mathrm{sgn}(\dot{h}(t))\,\dot{h}(t)^2 - mg\,, \\ h(t_0) &= h_0\,, \\ \dot{h}(t_0) &= v_0\,. \end{aligned}$$

1.1.2 Separation of Variables at the Example of Population Dynamics

Now consider a simple model that describes the evolution of a population over some period of time. That is, we know the population p_0 at some given time t_0, and we want to obtain an estimate $p(t)$ of the population at some future time $t > t_0$.

As a basic model, we assume that the *rate of change* of the population is given by some function $N(t, p)$ that depends only on the time and the size of the population. Then the function p that describes the population solves the differential equation

$$\dot{p}(t) = N\big(t, p(t)\big)\,, \qquad p(t_0) = p_0\,. \tag{1.2}$$

One very simple model assumes that the number of births and deaths within a certain amount of time is proportional to the size of the population that is,

$$N\big(t, p(t)\big) = (R - S)p(t)\,,$$

with R, S are the birth and death rates, respectively. Thus (1.2) becomes

$$\dot{p}(t) = (R - S)\, p(t)\ .$$

Using the initial state $p(t_0) = p_0$, we obtain with this model the population dynamics

$$p(t) = p_0 e^{(R-S)(t-t_0)}\ .$$

That is, depending on the sign of $R - S$, either the population increases or decreases exponentially.

Now we try to introduce the effects of overpopulation into the model by assuming that the death rate depends on the size of the population. That is, instead of assuming a constant death rate $S > 0$, we assume that S is a function of p. The simplest model is to assume the death rate being proportional to p, setting

$$S(p) = \sigma\, p$$

for some constant $\sigma > 0$. Then we obtain the equation (the *logistic differential equation*)

$$\dot{p}(t) = \big(R - \sigma p(t)\big)\, p(t)\ . \tag{1.3}$$

In the following, we will compute the analytic solution of this equation. We note first that the derivative of p is positive if $R > \sigma p$ (and the population p is positive, which we tacitly assume), while it is negative if $R < \sigma p$. In other words, the population increases as long as $p < R/\sigma$, while it decreases for $p > R/\sigma$. In particular, this implies that the long term behavior of the population will be stagnating at the value $p = R/\sigma$.

In order to solve the logistic differential equation, we define

$$\rho := R/\sigma\ ,$$

and rewrite the equation as

$$\frac{1}{(\rho - p(t))\, p(t)} \frac{dp(t)}{dt} = \sigma\ .$$

Integrating both sides of this equation with respect to t, we obtain

$$\int_0^{\hat{t}} \frac{1}{(\rho - p(t))\, p(t)} \frac{dp(t)}{dt}\, dt = \int_0^{\hat{t}} \sigma\, dt\ .$$

Now, we make a change of variables $t \to p := p(t)$, such that formally

$$\frac{dp(t)}{dt} dt = dp \, .$$

Using this identity we obtain the equation

$$\int_{p(0)}^{p(\hat{t})} \frac{1}{(\rho - p)p} \, dp = \sigma \hat{t} \, , \quad \forall \hat{t} \geq 0 \, . \tag{1.4}$$

Now note that

$$\begin{aligned}
&\int_{p(0)}^{p(\hat{t})} \frac{1}{(\rho - p)p} \, dp \\
=&\frac{1}{\rho} \int_{p(0)}^{p(\hat{t})} \frac{1}{\rho - p} + \frac{1}{p} \, dp \\
=&\frac{1}{\rho} \left(\left(- \ln \left| \rho - p(\hat{t}) \right| + \ln \left| p(\hat{t}) \right| \right) - \underbrace{\left(- \ln |\rho - p(0)| + \ln |p(0)| \right)}_{=:C} \right) \\
=&\frac{1}{\rho} \ln \left| \frac{p(\hat{t})}{\rho - p(\hat{t}))} \right| - C \, .
\end{aligned}$$

This, together with (1.4) shows that the function p satisfies:

$$\frac{1}{\rho} \ln \left| \frac{p(t)}{\rho - p(t)} \right| = \sigma t + C \, , \qquad \forall t \geq 0 \, .$$

Multiplying the equation with ρ and taking the exponential, it follows that

$$\left| \frac{p(t)}{\rho - p(t)} \right| = e^{\rho \sigma t + \rho C} = e^{\rho \sigma t} \, e^{\rho C} \, ,$$

which is equivalent to

$$\left| \frac{\rho - p(t)}{p(t)} \right| = e^{-\rho \sigma t} \, e^{-\rho C} \, , \tag{1.5}$$

Now we define a new constant

$$D := \pm e^{-\rho C} \, ,$$

where $D > 0$ if $\rho - p(t) > 0$ and else otherwise.

Then this last equation reads as

$$\frac{\rho}{p(t)} - 1 = De^{-\rho\sigma t}\,,$$

which in turn implies that

$$p(t) = \frac{\rho}{1 + De^{-\rho\sigma t}}\,.$$

This is the general form of a solution of the differential equation (1.3). The specific solution satisfying $p(t_0) = t_0$ is determined by determining D from

$$p_0 = p(t_0) = \frac{\rho}{1 + De^{-\rho\sigma t_0}}\,.$$

The method, which we applied in the last example is called *separation of variables*:

Definition 1.1. *An ODE that can be brought into the form*

$$f(y(t))\dot{y}(t) = g(t)\,, \tag{1.6}$$

where the function $f\colon \mathbb{R} \to \mathbb{R}$ only depends on y and not on t, and the function $g\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ only depends on t and not on y, is called ordinary differential equation (of first order) with separable variables*.*

The general strategy for solving such differential equations is to substitute $t \to y := y(t)$. Since

$$\dot{y} = \frac{dy}{dt}\,,$$

we can *formally* multiply (1.6) with dt and obtain the *formal* equation

$$f(y)\,dy = g(t)\,dt\,.$$

Now we can apply indefinite integrals to both sides and obtain the equation

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

where $C \in \mathbb{R}$ is some constant that appears due to the indefinite integration. Note, that the first integration is with respect to y and the right hand side reveals an integration with respect to t.

If it is possible to compute the integrals of f and g analytically, we obtain an equation the solution necessarily has to satisfy. If, in addition, it is possible to solve this equation for y, we indeed obtain an analytic (general) solution of the differential equation.

Example 1.2. *Consider the ODE*

$$(T^2 - t^2)\,\dot{y} + ty = 0\,,$$

where $T > 0$ is some given constant. This equation has separable variables, but in the form above they are not yet separated. In order to bring the equation in the form (1.6)*, we rewrite the equation as*

$$\frac{\dot{y}}{y} = -\frac{t}{T^2 - t^2}\,,$$

which is possible for $y \neq 0$ and $t \neq \pm T$. We rewrite this formally as

$$\frac{dy}{y} = -\frac{t}{T^2 - t^2}dt\,.$$

Now, integration of both sides of the equation leads to

$$\ln|y| = \frac{1}{2}\ln\left|T^2 - t^2\right| + C\,.$$

Taking the exponential of the equation, we obtain

$$|y| = e^C\sqrt{|T^2 - t^2|}\,.$$

Replacing the constant $e^C > 0$ by the constant $D \in \mathbb{R}$ also encoding the sign of y, we get

$$y(t) = D\sqrt{|T^2 - t^2|}\,. \tag{1.7}$$

The constant $D \in \mathbb{R}$ still has to be determined using the initial condition $y(t_0) = y_0$. Inserting this condition into the general solution, we see that

$$y_0 = y(t_0) = D\sqrt{|T^2 - t_0^2|}\,,$$

and therefore

$$D = \frac{y_0}{\sqrt{|T^2 - t_0^2|}}\,. \tag{1.8}$$

Note that we have assumed during the computation of the solution of the ODE that $y_0 \neq 0$ *and* $t \neq \pm T$. *It can be easily seen, however, that the derivation above also covers the situation where* $y_0 = 0$ *and* $t_0 \neq \pm T$. *There, the constant function* $y = 0$ *is the unique solution of the ODE, at least until the time reaches one of the values* $\pm T$.

The case $t_0 = \pm T$*, however, is different. Then, if* $y_0 = 0$*, for* every *constant* $D \in \mathbb{R}$ *the function* (1.7) *satisfies the ODE and therefore is a solution. If, however,* $y_0 \neq 0$*, then the ODE has no solution at all– the ODE and the initial conditions are inconsistent.*

Finally, note that all the solutions are valid only locally; that is, there exists at least a time interval $[t_0, t_0 + \varepsilon)$ *for some* $\varepsilon > 0$ *on which the solution exists and can be written as* (1.7) *with* D *given by* (1.8). *For general ODEs, this is all that can be said about the solution. In this special case, one can specify the length of the interval on which the solution looks like* (1.7)*: If* $t_0 > T$*, then the formula* (1.7) *is valid on* $[t_0, +\infty)$. *If, however,* $-T < t_0 < T$*, then the solution is*

$$y(t) = \begin{cases} D_1\sqrt{|T^2 - t^2|} & \text{if } t \in [t_0, T]\,, \\ D_2\sqrt{|T^2 - t^2|} & \text{if } t \in [T, +\infty)\,, \end{cases} \qquad \text{with } \begin{cases} D_1 = y_0/\sqrt{T^2 - t_0^2}\,, \\ D_2 \in \mathbb{R} \text{ arbitrary.} \end{cases}$$

In particular, the solution is only unique up to time T. *Similarly, if* $t_0 < -T$*, then*

$$y(t) = \begin{cases} D_1\sqrt{|T^2 - t^2|} & \text{if } t \in [t_0, -T]\,, \\ D_2\sqrt{|T^2 - t^2|} & \text{if } t \in [-T, T]\,, \\ D_3\sqrt{|T^2 - t^2|} & \text{if } t \in [T, +\infty)\,, \end{cases} \qquad \text{with } \begin{cases} D_1 = y_0/\sqrt{T^2 - t_0^2}\,, \\ D_2 \in \mathbb{R} \text{ arbitrary,} \\ D_3 \in \mathbb{R} \text{ arbitrary.} \end{cases}$$

1.1.3 Homogeneous ODEs

Definition 1.3. *An ODE of the form*

$$\dot{y} = f\left(\frac{y}{t}\right), \tag{1.9}$$

with $f\colon \mathbb{R} \to \mathbb{R}$*, is called of* homogeneous type*.*

If we are given an ODE of homogeneous type, we can solve it by starting with the substitution

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

For the right hand side of (1.9) we are left with the term $f(z)$. For the left hand side of (1.9) we use the product rule and obtain

$$\dot{y} = \frac{dy}{dt} = \frac{d(tz)}{dt} = z + t\frac{dz}{dt} = z + t\dot{z} \ .$$

Thus we have for the variable z the differential equation

$$z + t\dot{z} = f(z) \ .$$

Now it is easy to see that this ODE is of separable type: We can bring it in the form

$$\frac{\dot{z}}{f(z) - z} = \frac{1}{t} \ .$$

This ODE can now be solved by integration as in Section **??**, and we obtain a solution $z(t)$. At the end, we obtain the solution y by $y(t) = tz(t)$.

Example 1.4. *Consider the ODE*

$$\dot{y} = \left(\frac{y}{t}\right)^2 \ .$$

It is easy to see that this ODE is homogeneous with $f(y/t) = (y/t)^2$. Using the substitution $y = tz$ we obtain

$$z + t\dot{z} = z^2$$

and therefore

$$\frac{\dot{z}}{z^2 - z} = \frac{1}{t} \ .$$

Now, we follow Subsection **??** *and reformulate this equation to*

$$\frac{1}{z^2 - z} dz = \frac{1}{t} dt \ .$$

Integrating this equation, we obtain the indefinite integral equation

$$\int \frac{1}{z^2 - z} dz = \int \frac{1}{t} dt + C$$

or by calculating the integrals

$$\ln \left| \frac{z - 1}{z} \right| = \ln |t| + C \ .$$

Now, assuming that $t > 0$ and $z \geq 1$, which depends on the initial condition, we get

$$\frac{z-1}{z} = Dt$$

for some constant $D = \exp(C) \in \mathbb{R}^+$ depending on the initial value. Solving for z we obtain

$$z(t) = \frac{1}{1 - Dt}$$

(note that z is greater than 1) and, after substitution of $y = tz$

$$y(t) = \frac{t}{1 - Dt} \; .$$

1.2 Linear ODEs

Definition 1.5. *An ODE that can be written as*

$$\dot{y} + f(t)y = g(t)$$

for some functions $f \colon \mathbb{R} \to \mathbb{R}$ and $g \colon \mathbb{R} \to \mathbb{R}$ is called linear ODE of first order.

Here, first order means that the highest derivative of the unknown function y that appears in the equation is the first one. Linear means that all the expressions are linear in the unknown y and its derivatives.

As in the case of linear algebraic equations, the linearity of an equation has some implications on the structure of its solutions. To that end we consider the *homogeneous* equation[1]

$$\dot{y} + f(t)y = 0 \; .$$

If we are given two solutions y_1 and y_2 of this equation (with possibly different initial conditions), then

$$\dot{y}_1 + f(t)y_1 = 0 \, ,$$
$$\dot{y}_2 + f(t)y_2 = 0 \; .$$

Consequently also

$$\frac{d}{dt}(y_1 + y_2) + f(t)(y_1 + y_2) = \dot{y}_1 + f(t)y_1 + \dot{y}_2 + f(t)y_2 = 0 \, ,$$

[1] Homogeneous means that the right hand side of the equation is zero, that is, $g = 0$

which shows that also $y_1 + y_2$ is a solution of the ODE. More general, if y_1 and y_2 solve the ODE and c_1, $c_2 \in \mathbb{R}$, then the linear combination $c_1 y_1 + c_2 y_2$ is also a solution.

In order to solve the (inhomogeneous) equation

$$\dot{y} + f(t)y = g(t) \tag{1.10}$$

we first observe that (1.10) is equivalent to

$$h(t)\dot{y} + h(t)f(t)y = h(t)g(t)\,, \tag{1.11}$$

at least, if $h\colon \mathbb{R} \to \mathbb{R}$ is a function that is different from zero.

Now the idea is to choose the function h in such a way that the left hand side of (1.11) is itself a derivative. More precisely, we try to find $h\colon \mathbb{R} \to \mathbb{R}$ such that

$$h(t)\dot{y} + h(t)f(t)y = \frac{d}{dt}(hy) = \dot{h}y + h\dot{y}\ . \tag{1.12}$$

If (1.12) holds, then the equation (1.11) reads as follows,

$$\frac{d}{dt}(hy) = h(t)g(t)\,,$$

which after integration becomes:

$$h(t)y(t) = \int^t g(s)h(s)\,ds + C\ . \tag{1.13}$$

For this reason, a function h satisfying (1.12) is called an *integrating factor* for the ODE (1.10).

Thus, (1.12) is satisfied, if h satisfies

$$h(t)f(t)y = \dot{h}(t)y\ .$$

Dividing this equation by y, we see that h has to satisfy the ODE

$$\dot{h} = f(t)h\ .$$

This ODE can be solved by separation of the variables, and we obtain the integrating factor

$$h(t) = D\exp\Big(\int^t f(s)\,ds\Big)\ .$$

Inserting this integrating factor in (1.13), we obtain

$$y(t) = \frac{\int^t \Big[g(s) D \exp\Big(\int^s f(r)\,dr\Big)\Big] ds + C}{D \exp\Big(\int^t f(s)\,ds\Big)}\,,$$

or, setting $\tilde{C} := C/D$,

$$y(t) = \frac{\int^t \Big[g(s) \exp\Big(\int^s f(r)\,dr\Big)\Big] ds + \tilde{C}}{\exp\Big(\int^t f(s)\,ds\Big)}\,.$$

1.3 Simple Examples of Partial Differential Equations

Definition 1.6. *A partial differential equation (PDE) is an equation for more than two different derivatives of a function $u(x_1, x_2, \ldots, x_n)$ on a domain $\Omega \subseteq \mathbb{R}^n$.*

Example 1.7. *For instance*

$$y\frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial y} = x^2 yu \text{ or } yu_{xx} + u_y = x^2 yu\,, \tag{1.14}$$

which actually means

$$y\frac{\partial^2 u}{\partial x^2}(x,y) + \frac{\partial u}{\partial y}(x,y) = x^2 yu(x,y) \text{ for } (x,y) \in \Omega\,. \tag{1.15}$$

This is a PDE for a function $u(x,y)$ in two variables.

We use the notation that

$$u_x = \frac{\partial u}{\partial x}, u_y = \frac{\partial u}{\partial y},$$

and

$$u_{xx} = \frac{\partial^2 u}{\partial x^2} = \frac{\partial \frac{\partial u}{\partial x}}{\partial x}\,, u_{xy} = \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial \frac{\partial u}{\partial x}}{\partial y} = \frac{\partial \frac{\partial u}{\partial y}}{\partial x} = u_{yx}\,.$$

- The variables x, y are called *independent* variables.

- u is called *dependent* variable.

The *order* of the differential equation is the order of the highest derivative of the dependent variables in the differential equation. The PDE (1.15) is a differential of second order. The differential equation

$$xu_x u_{xxy} + u_x^4 = 0$$

is of third order.

Most PDEs with relevance in practice are of first or second order.

Example 1.8. *1. The electrostatic potential* $u(x, y, z)$ *which is determined by a charge density* $\rho(x, y, z)$ *satisfies the* Poisson equation

$$\Delta u = u_{xx} + u_{yy} + u_{zz} = 4\pi\rho \ .$$

Δ *denotes the* Laplace *operator (in space dimension three).*

2. The wave equation *is the PDE*

$$\frac{1}{c^2} u_{tt} = \Delta u \ .$$

In air, $u(x, y, z, t)$ *denotes the density of air at a location* (x, y, z) *at time* t*.* c *denotes the sound speed.*

3. Heat *or* diffusion *equation:*

$$u_t = \alpha \Delta u \, ,$$

with some $\alpha > 0$*.*

4. The velocity $\vec{v}(\vec{x}, t) = (v_1, v_2, v_3)(\vec{x}, t)$ *and the pressure* $p(\vec{x}, t)$ *of an incompressible fluid as a function of space* $\vec{x} = (x, y, z)$ *and time* t *satisfies the* Navier-Stokes-equations

$$\vec{v}_t + (\vec{v} \cdot \nabla)\vec{v} = \nu \Delta \vec{v} + \nabla p \, , \nabla \cdot \vec{v} = 0 \, ,$$

for some constant ν*. In the above equation*

$$\nabla \cdot \vec{v} := (v_1)_x + (v_2)_y + (v_3)_z$$

denotes the divergence*.*

$$(\vec{v} \cdot \nabla)\vec{v} = \begin{pmatrix} v_1(v_1)_x + v_2(v_1)_y + v_3(v_1)_z \\ v_1(v_2)_x + v_2(v_2)_y + v_3(v_2)_z \\ v_1(v_3)_x + v_2(v_3)_y + v_3(v_3)_z \end{pmatrix} .$$

This is a system of four equations in four unknowns.

Definition 1.9. *A PDE is called* linear *if u and its derivative only appear* linearly*. More precisely a linear PDE has the form*

$$Lu = b\,,$$

where L is a differential operator and b is a given function.

Example 1.10. *Equation* (1.14) *is linear with*

$$L = y\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial y} - x^2 y\,, \qquad b = 0\,.$$

The Poisson equation is linear with $L = \Delta$ and $b = 4\pi\rho$. The wave equation and the heat equation are linear, respectively. The Navier-Stokes equation is nonlinear. Another, frequently used nonlinear PDE is the Burger's equation

$$u_x + uu_y = 0\,.$$

1.3.1 Conservation Principles

Differential equations are frequently derived from conservation of physical quantities like mass, energy, temperature and so on. To illustrate this we consider the temperature distribution in a homogeneous, non insulating slab of length L. We denote now by $u(x,t)$ the temperature in a point $x \in [0, L]$ at time $t \geq 0$.

We are modeling the following principles:

1. Conservation of energy: The timely variation of thermal energy in every interval $[a, b] \subseteq [0, L]$ is equal to the heat flux across a and b.

2. The energy density (energy per length's unit) is $\rho c u$. Thereby ρ denotes the density, c is the specific heat. Both ρ and c are assumed to be constant here.

3. Fourier's law: The heat flux is proportional to the gradient of the temperature (heat is fluctuating from warm to cold), and the proportionality constant $k > 0$ is called heat conductivity. That is, we have

$$\frac{d}{dt}\int_a^b \rho c u(x,t)\,dx = -k\frac{\partial u}{\partial x}(a,t) + k\frac{\partial u}{\partial x}(b,t)\,. \tag{1.16}$$

The left hand side is the variation of the total energy in the slab. The right hand side is the energy, which migrates in and out of the slab per time unit. Equation (1.16) should hold for all intervals $[a,b]$. Thus by the fundamental theorem of integration we get from (1.16)

$$\int_a^b \left(\rho c\frac{\partial u}{\partial t}(x,t)\,dx - k\frac{\partial^2 u}{\partial x^2}(x,t)\right)\,dx = 0\,.$$

Because this holds for arbitrary intervals $[a, a+\varepsilon]$ we see that

$$\begin{aligned} 0 &= \frac{1}{\varepsilon}\int_a^{a+\varepsilon} \left(\rho c\frac{\partial u}{\partial t}(x,t)\,dx - k\frac{\partial^2 u}{\partial x^2}(x,t)\right)\,dx \\ &\sim \left(\rho c\frac{\partial u}{\partial t}(a,t) - k\frac{\partial^2 u}{\partial x^2}(a,t)\right) \end{aligned}$$

This should hold for all a, which gives that

$$0 = \frac{\partial u}{\partial t}(a,t) - \underbrace{\frac{k}{\rho c}}_{:=\alpha}\frac{\partial^2 u}{\partial x^2}(a,t)\,.$$

This is the heat equation in $\mathbb{R}\times\mathbb{R}_0^+$.

In $\mathbb{R}^n$ the derivation from Fourier's law is similar and results in

$$0 = \frac{\partial u}{\partial t}(a,t) - \underbrace{\frac{k}{\rho c}}_{:=\alpha}\Delta u(a,t)\,,$$

for all $a \in \mathbb{R}^3$. Here

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}$$

denotes the Laplace operator.

Chapter 2

Classification of Linear Partial Differential Equations

For the sake of simplicity of presentation we restrict attention to linear partial differential equations of second order in two variables. Such an equation for a function $u = u(x, y)$ reads as follows:

$$Au_{xx} + 2B_{xy} + Cu_{yy} + Du_x + Eu_y + Fu + G = 0 \ . \qquad (2.1)$$

Here $A = A(x, y), \dots, G = G(x, y)$ are again functions.

Definition 2.1. *A PDE* (2.1) *is called*

- elliptic *if* $AC - B^2 > 0$ *,*
- parabolic *if* $AC - B^2 = 0$ *, and*
- hyperbolic *if* $AC - B^2 < 0$ *.*

Example 2.2.

The wave equation

$$\frac{1}{c^2}u_{xx} - u_{yy} = 0$$

(note we changed the notation from t to y) is of the from (2.1) *with*

$$A = c^{-2}, B = 0, C = -1 \ .$$

Because

$$AC - B^2 = -c^{-2} < 0 \, ,$$

the equation is hyperbolic.

The Laplace equation

$$u_{xx} + u_{yy} = 0$$

is of the from (2.1) *with*

$$A = 1, B = 0, C = 1 \; .$$

Because

$$AC - B^2 = 1 \geq 0 \, ,$$

the equation is elliptic.

For the heat equation

$$u_x - u_{yy} = 0$$

(note we changed the notation from t to x and x to y) is of the from (2.1) *with*

$$A = 0, B = 0, C = -1 \; .$$

Because

$$AC - B^2 = 0 \, ,$$

the equation is parabolic.

Remark 2.3. 1. *For the classification on the* main symbol

$$Au_{xx} + 2Bu_{xy} + Cu_{yy}$$

is relevant. These are the terms of the differential equation of highest order (in our case this is 2).

2. $AC - B^2$ *is the determinant of the symmetric matrix*

$$M = \begin{pmatrix} A & B \\ B & C \end{pmatrix} \; .$$

Denoting by $m_{11} = A$, $m_{12} = m_{21} = B$, $m_{22} = C$ and $x_1 = x$, $x_2 = y$ the main symbol reads as follows

$$\sum_{i,j=1}^{2} m_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} \; .$$

3. *If the coefficients* A, B, C *are not constant, but functions which depend on* x *and* y *in a non-trivial manner, then the type of the partial differential equation can be different at various points* (x, y).

   For instance the differential equation

   $$xu_{xx} + u_{yy} = 0$$

   is elliptic for $x > 0$, *parabolic for* $x = 0$, *and hyperbolic for* $x < 0$.

4. *The terminology elliptic, parabolic and hyperbolic is motivated from conic sections. A curve* $(X, Y(X))$, *which satisfies the equation*

   $$AX^2+2BXY+CY^2+DX+EY+F = 0 \qquad \text{(with constant coefficients)}$$

   is either an ellipsis, parabola, or an hyperbola, depending on the sign of $AC - B^2$. *For instance*

   (a) $A = C = 1$, $B = D = E = 0$, $F = -1$ *leads to* $X^2 + Y^2 = 1$, *a circle.*

   (b) *More general* $A = \frac{1}{a^2}$, $C = \frac{1}{b^2}$, $B = D = E = 0$, $F = -1$ *leads to* $\frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1$, *an ellipsis.*

   (c) $A = F = 1$, $C = -1$ *and* $B = D = E = 0$ *lead to,* $y^2 = x^2 + 1$, *which is an hyperbola.*

   (d) $A = 1$, $B = C = D = F = 0$, $E = -1$ *leads to* $Y = X^2$, *which is a parabola.*

5. *A different terminology is used in linear algebra, where* quadratic forms *are investigated:*

   $$Q(X, Y) = AX^2 + 2BXY + CY^2 .$$

   We assume that one of the coefficients A, B, C *is not identical* 0.

   - Q *is called positive (negative) definite, if* $Q(X, Y) > 0$ ($Q(X, Y) < 0$) *for all* $(X, Y) \neq (0, 0)$.
   - Q *is called positive (negative) semi-definite, if* $Q(X, Y) \geq 0$ ($Q(X, Y) \leq 0$) *for all* $(X, Y) \neq (0, 0)$.
   - *If* Q *is not semi-definite, then it is called* indefinite.

Theorem 2.4.
- *Q is positive or negative (semi-) definite if and only if $AC - B^2 > (\geq)0$;*
- *Q is indefinite iff $AC - B^2 < 0$;*
- *If $AC - B^2 = 0$, then Q is semi-definite.*

Proof. We first note that for $A \neq 0$

$$\begin{aligned} Q(X,Y) &= AX^2 + 2BXY + CY^2 \\ &= A\left(X^2 + 2\frac{B}{A}XY + \frac{B^2}{A^2}Y^2\right) + \left(C - \frac{B^2}{A}\right)Y^2 \\ &= A\left(X + \frac{B}{A}Y\right)^2 + \left(C - \frac{B^2}{A}\right)Y^2 \\ &= \frac{(AX+BY)^2}{A} + \frac{AC - B^2}{A}Y^2 \,. \end{aligned} \tag{2.2}$$

(a) The first item:

- For the $\Rightarrow$ direction let Q be positive of negative (semi-)definite.
  - If Q is positive (semi-)definite, then
  $$A = Q(1,0) > (\geq)0 \text{ and } C = Q(0,1) > (\geq)0 \,. \tag{2.3}$$
  - If Q is negative (semi-) definite, then
  $$A = Q(1,0) < (\leq)0 \text{ and } C = Q(0,1) < (\leq)0 \,. \tag{2.4}$$

  Thus $AC > (\geq)0$ in both cases.
  We differ again between two cases:

  i. If $B = 0$, then (2.2) and (2.3), (2.4) imply that
  $$AC - B^2 = AC > (\geq)0 \,.$$
  ii. Let $B \neq 0$. If Q is positive semi-definite and $B \neq 0$, then $A \neq 0$: Note that otherwise, for $A = 0$, the positive semi-definiteness guarantees that
  $$0 \leq Q\left(-\frac{|C|+1}{2B}, 1\right) = -|C| - 1 + C < 0 \,.$$
  which gives a contradiction.

Thus $A > 0$ if Q is positive definite or positive semi-definite, respectively.
If Q is positive (semi-)definite, we have

$$0 < (\leq) Q\left(1, -\frac{A}{B}\right) = (AC - B^2)\frac{A}{B^2} \ .$$

Because $A > 0$ it then follows that $AC - B^2 > 0$.

iii. If Q is negative (semi-) definite, the argumentation is analogous to show that $AC - B^2 > (\geq) 0$.

- For the opposite direction, let $AC - B^2 > (\geq)0$, then $AC > (\geq)0$. Then from (2.2) it follows that
  –

  $$\mathrm{sgn}(A)Q(X,Y) \geq (AC - B^2)Y^2 > (\geq)0 \,, \qquad \forall X \in \mathbb{R} \text{ and } Y \neq 0 \ .$$

  Thus $Q(X,Y)$ is either positive or negative (semi-) definite, depending on the sign of A.
  – If $Y = 0$, then

  $$\mathrm{sgn}(A)Q(X,Y) = |A|\, X^2 > 0 \,, \qquad \forall X \neq 0 \ .$$

(b) The negation of Q is semi-definite is that Q is indefinite. Q is semi-definite is equivalent to $AC - B^2 \geq 0$. Thus Q is indefinite if $AC - B^2 < 0$.

(c) The third item: Let $AC - B^2 = 0$, then

- If $B = 0$ also $A = 0$ or $C = 0$. Without loss of generality we assume that $C = 0$. Then

  $$Q(X,Y) = AX^2 \geq 0 \text{ if } A \geq 0 \text{ or } Q(X,Y) = AX^2 \leq 0 \text{ if } A \leq 0 \ .$$

  That means that Q is positive, negative definite, depending on the sign of A, respectively.
- If $B \neq 0$ also $A \neq 0$ and $C \neq 0$. Moreover, A and C have the same sign. Then

  $$\begin{aligned} Q(X,Y) &= AX^2 + 2BXY + CY^2 \\ &= \mathrm{sgn}(A)\left(|A|\, X^2 + 2\sqrt{AC}XY + |C|\, Y^2\right) \\ &= \mathrm{sgn}(A)(|A|\, X + |C|\, Y)^2 \ . \end{aligned}$$

That means that Q is positive, negative definite, depending on the sign of A, respectively.

□

By a change of coordinates every quadratic form can be transformed into

- $X^2 + Y^2$ *if* $AC - B^2 > 0$,
- $X^2 - Y^2$ *if* $AC - B^2 < 0$,
- X^2 *if* $AC - B^2 = 0$.

For instance,

$$\begin{aligned} Q(X,Y) &= X^2 + 4XY + Y^2 \\ &= (X+2Y)^2 - 4Y^2 + Y^2 \\ &= (X+2Y)^2 - 3Y^2 \\ &= (X')^2 - (Y')^2 , \end{aligned}$$

where

$$X' = X + 2Y \text{ and } Y' = \sqrt{3}Y \ .$$

Remark 2.5. *An important property of the classification, Definition 2.1, is that it is invariant under coordinate transformations: A coordinate transformation does not change the type of the differential equation.*

We consider the change of coordinates:

$$\begin{aligned} L_1(x,y) &= ax + by \, , \\ L_2(x,y) &= cx + dy \ . \end{aligned}$$

This shows that

$$\begin{aligned} \frac{\partial(u \circ L)}{\partial x}(x,y) &= a\partial_1 u(L(x,y)) + c\partial_2 u(L(x,y)) \, , \\ \frac{\partial(u \circ L)}{\partial y}(x,y) &= b\partial_1 u(L(x,y)) + d\partial_2 u(L(x,y)) \ . \end{aligned}$$

The notation is ugly, however, it should make aware that on the right hand side we differentiate with respect to the first component and not x variable.

Thus

$$\frac{\partial^2(u\circ L)}{\partial^2 x}(x,y) = a^2\partial_1^2 u(L(x,y)) + 2ac\partial_{12}^2 u(L(x,y)) + c^2\partial_2^2 u(L(x,y))\,,$$
$$\frac{\partial^2(u\circ L)}{\partial x\partial y}(x,y) = ab\partial_1^2 u(L(x,y)) + (ad+bc)\partial_{12}^2 u(L(x,y)) + cd\partial_2^2 u(L(x,y))\,,$$
$$\frac{\partial^2(u\circ L)}{\partial^2 y}(x,y) = b^2\partial_1^2 u(L(x,y)) + 2bd\partial_{12}^2 u(L(x,y)) + d^2\partial_2^2 u(L(x,y))\;.$$

This can be written now into compact matrix form:

$$\underbrace{\begin{pmatrix} \frac{\partial^2(u\circ L)}{\partial^2 x}(x,y) & \frac{\partial^2(u\circ L)}{\partial x\partial y}(x,y) \\ \frac{\partial^2(u\circ L)}{\partial x\partial y}(x,y) & \frac{\partial^2(u\circ L)}{\partial^2 y}(x,y) \end{pmatrix}}_{=:C}$$
$$= A\underbrace{\begin{pmatrix} \partial_1^2 u(L(x,y)) & \partial_{12}^2 u(L(x,y)) \\ \partial_{12}^2 u(L(x,y)) & \partial_2^2 u(L(x,y)) \end{pmatrix}}_{=:U''} A^T$$

with

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

Thus we have

$$(\det(A))^2\det(U'') = \det(C)\,,$$

and thus the determinants of C and A have equal signs. This shows that the type does not change by linear transformations.

2.1 Characteristics

The following considerations make evident the role of the different types of differential equations.

Exemplary, first, we consider an ordinary differential equation:

$$u''(x) = F(x, u(x), u'(x))\;.$$

Suppose that $u(x_0)$ and $u'(x_0)$ are known in a point x_0, then (compare the Euler method)

$$\begin{aligned} u(x_0+\Delta x) &\approx u(x_0) + u'(x_0)\Delta x\,, \\ u'(x_0+\Delta x) &\approx u'(x_0) + u''(x_0)\Delta x \\ &= u'(x_0) + F(x_0, u(x_0), u'(x_0))\Delta x\;. \end{aligned}$$

The equations get exact for $\Delta x \to 0$. That is, the solution of the differential equation can be determined approximately from the values at x_0. In the following we apply this idea to partial differential equations: We use the linearization:

$$u(x_0 + \Delta x, y_0 + \Delta y) \approx u(x_0, y_0) + u_x(x_0, y_0)\Delta x + u_y(x_0, y_0)\Delta y \ .$$

But we need also approximations for higher order derivatives which can be derived by approximation of u_x:

$$\begin{aligned} u_x(x_0 + \Delta x, y_0 + \Delta y) &\approx u_x(x_0, y_0) + u_{xx}(x_0, y_0)\Delta x + u_{xy}(x_0, y_0)\Delta y \ , \\ u_y(x_0 + \Delta x, y_0 + \Delta y) &\approx u_y(x_0, y_0) + u_{yx}(x_0, y_0)\Delta x + u_{yy}(x_0, y_0)\Delta y \ . \end{aligned}$$

This means that if you know the function value u at (x_0, y_0) and derivatives of up to second order at (x_0, y_0), then one knows also u and it's first derivatives in a neighborhood, that is at $(x_0 + \Delta x, y_0 + \Delta y)$.

This idea is generalized now to the PDE (2.1)

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} = -Du_x - Eu_y - Fu - G \ ,$$

We now assume that we know u_x, u_y and u on a curve (not just on (x_0, y_0)), with tangential directions $\vec{v}$ for every point on the curve. Then we also know the tangential derivatives of u_x, u_y,... in tangential direction $\vec{v}$ on the curve:

$$D_{\vec{v}} u_x = \lim_{t \to 0} \frac{u_x(x_0 + t\vec{v}) - u_x(x_0)}{t} \text{ and } D_{\vec{v}} u_y = \lim_{t \to 0} \frac{u_y(x_0 + t\vec{v}) - u_y(x_0)}{t} \ .$$

The tangential derivative can be expressed as

$$D_{\vec{v}} u_x = \nabla u_x \cdot \vec{v} \text{ and } D_{\vec{v}} u_y = \nabla u_y \cdot \vec{v} \ .$$

And therefore

$$\begin{aligned} v_1 u_{xx} + v_2 u_{xy} &= D_{\vec{v}} u_x \ , \\ v_1 u_{xy} + v_2 u_{yy} &= D_{\vec{v}} u_y \ . \end{aligned}$$

In summary, we know that for a given curve with tangential direction $\vec{v}$ and given u_x, u_y and u on a piece of the curve, the second order derivatives can be determined from the system

$$\begin{aligned} v_1 u_{xx} + v_2 u_{xy} \qquad\qquad &= D_{\vec{v}} u_x \ , \\ v_1 u_{xy} + \quad v_2 u_{yy} &= D_{\vec{v}} u_y \ , \\ Au_{xx} + 2Bu_{xy} + \quad Cu_{yy} &= -Du_x - Eu_y - Fu - G \ . \end{aligned} \tag{2.5}$$

This linear system has a unique solution if

$$0 \neq \det \begin{pmatrix} A & 2B & C \\ v_1 & v_2 & 0 \\ 0 & v_1 & v_2 \end{pmatrix} = Av_2^2 - 2Bv_1v_2 + Cv_1^2 =: Q(v_2, -v_1) \ . \qquad (2.6)$$

This means that we can solve the PDE (locally around a point on the curve) exactly.

Now, we reformulate the function Q:

$$\begin{aligned} Q(v_2, -v_1) &= Av_2^2 - 2Bv_1v_2 + Cv_1^2 \\ &= A\left(v_2^2 - 2\frac{B}{A}v_1v_2 + \frac{C}{A}v_1^2\right) \\ &= A\left(\underbrace{(v_2 - \frac{B}{A}v_1)}_{=:\tilde{v}_2}{}^2 - \left(\frac{B^2}{A^2} - \frac{C}{A}\right)v_1^2\right) \ . \end{aligned}$$

Depending on the sign of $AC - B^2$ we have different scenarios:

Elliptic case: In this case $AC > B^2$, which in particular implies that $A \neq 0$. Then $Q(v_2, -v_1) = 0$ implies that $\tilde{v}_2 = v_1 = 0$, and consequently $v_1 = v_2 = 0$.

Hyperbolic case: In the case $AC < B^2$, we can have the case $A = 0$ in which every $\vec{v}$ is a solution of $Q(v_2, -v_1) = 0$. If $A \neq 0$, then $\vec{v}$ is a solution of $Q(v_2, -v_1) = 0$ iff

$$\tilde{v}_2 = \pm\sqrt{\frac{B^2}{A^2} - \frac{C}{A}}v_1 \ .$$

Thus there exists a one-dimensional solution space, and thus the system (2.5) has nontrivial solutions too. This means that the solution of the PDE is not uniquely determined by the values u, u_x, u_y on the curve with tangent vector $\vec{v}$. Curves with such a property are called *characteristics*.

Because $\vec{n} = (v_2, -v_1)$ is the normal vector to the tangent we get the following definition of characteristics:

Definition 2.6. *Let $AC - B^2 < 0$. A curve in the xy-plane is called* characteristics *of the hyperbolic PDE* (2.1) *if the normal vector $\vec{n} = (n_1, n_2)$ satisfies in every point the equation*

$$An_1^2 + 2Bn_1n_2 + Cn_2^2 = 0 \ .$$

Example 2.7. *The characteristics of the equation*

$$u_{xx} - u_{yy} = 0 \,, \qquad (A = 1, C = -1, B = 0)$$

are the curves, where the normal vector satisfies

$$n_1^2 - n_2^2 = 0 \ .$$

That is $n_1 = \pm n_2$. That are the lines $x \pm y = const$. In this case the solution space $Q(v_2, -v_1)$ is a one dimensional linear space.

If we identify y with time t, this is the standard wave equation. The characteristics $x \pm t = const$ are the lines, where the waves propagate: If we consider waves, which move to the right, then this waves are given by

$$u(x,t) = F(x - t) \ .$$

Note, that on the characteristics the value is constant, and from this value the solution cannot be determined in normal direction. The ones which move to the left are

$$u(x,t) = G(x + t) \ .$$

An in general the waves are of the form

$$u(x,t) = F(x - t) + G(x + t) \ .$$

This example makes clear the roles of characteristics.

Chapter 3

Boundary Value Problems

For motivating purposes we study first boundary value problems for ordinary differential equations at hand of a simple test example:

$$\begin{aligned} L[u] = -u'' + bu' + cu = f \text{ in } (0,1)\,, \\ u(0) = u(1) = 0\,. \end{aligned} \tag{3.1}$$

b, c, and f can be functions on $(0,1)$.

It can be shown that this differential equation has a unique solution provided that

$$c(x) \geq 0\,, \qquad \forall x \in (0,1)\,.$$

This will always be assumed in the following.

Here, for the numerical solution, we consider *finite difference methods* (FDM). Later, alternatively, we will investigate *finite element methods* (FEM).

For the simplicity of presentation we consider an equidistant grid

$$\Delta_h = \{x_i = ih : i = 1, \ldots, n-1, h = 1/n\} \subseteq (0,1)\,. \tag{3.2}$$

We denote by

$$\vec{u} = (u(x_1), \ldots, u(x_{n-1}))^t \in \mathbb{R}^{n-1} \tag{3.3}$$

the vector of the exact solution u of (3.1) on the grid Δ_h (3.2). In addition, we assume Dirichlet boundary conditions

$$0 = u(x_0) = u(x_n) = 0\,.$$

For the numerical solution we look for an approximating vector

$$\vec{u}_h = (u_1, \ldots, u_{n-1})^t \in \mathbb{R}^{n-1}\,. \tag{3.4}$$

For this purpose we discretize L from (3.1) by approximating the derivatives of u at the positions $x = x_i$ via *difference quotients*. Thereby we have several alternatives:

- *One-sided forward-difference operator:*

$$D_h^+[u](x) = \frac{u(x+h) - u(x)}{h} \sim u'(x) \ .$$

- *One-sided backward-difference operator:*

$$D_h^-[u](x) = \frac{u(x) - u(x-h)}{h} \sim u'(x) \ .$$

- *Central difference quotient:*

$$D_h[u](x) = \frac{u(x+h) - u(x-h)}{2h} \sim u'(x) \ . \tag{3.5}$$

Moreover, the second derivative can be approximated by a central difference quotient

$$D_h^2[u](x) = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} \sim u''(x) \ . \tag{3.6}$$

Example 3.1. *We study a simple situation of* (3.1) *with* $b, c \equiv 0$*, that is* $-u'' = f$*. We approximate* u'' *by* $D_h^2[u]$ *at the nodal points* Δ_h*. Taking into account the Dirichlet boundary conditions* $u(x_0) = u(x_n) = 0$ *we get the discretized equation:*

$$\underbrace{\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{n-1}) \end{bmatrix}}_{=:\vec{f}} = - \begin{bmatrix} u''(x_1) \\ u''(x_2) \\ \vdots \\ u''(x_{n-1}) \end{bmatrix} \sim h^{-2} \underbrace{\begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix}}_{=:L_h} \underbrace{\begin{bmatrix} u(x_1) \\ u(x_2) \\ \vdots \\ u(x_{n-1}) \end{bmatrix}}_{\vec{u}} .$$

Because $\vec{u}$ *should be approximated by* $\vec{u}_h$*, we will use the following linear equation to determine* $\vec{u}_h$*:*

$$L_h \vec{u}_h = \vec{f} \ . \tag{3.7}$$

The Eigenvalues of L_h *are* $4h^{-2}\sin^2(kh\pi/2)4$, $k = 1, \dots, n-1$. *The function* $sinc(x) := \frac{\sin(x)}{x}$ *is monotonically decreasing in* $[0, \pi/2]$ *such that*

$$sinc(x) \geq sinc\left(\frac{\pi}{2}\right) = \frac{2}{\pi}, \quad \forall x \in [0, \pi/2],$$

which implies that:

$$\left\|L_h^{-1}\right\|_2 = \frac{1}{\lambda_{\min}(L_h)} = \max_{1 \leq k \leq n-1} \frac{h^2}{4\sin^2(kh\pi/2)} \leq \frac{1}{4}.$$

Consequently,

$$\begin{aligned} \|\vec{u} - \vec{u}_h\|_2 &= \left\|L_h^{-1}(L_h\vec{u} - \vec{f})\right\|_2 \\ &\leq \left\|L_h^{-1}\right\|_2 \left\|L_h\vec{u} - \vec{f}\right\|_2 \\ &\leq \frac{1}{4}\left\|L_h\vec{u} - \vec{f}\right\|_2 . \end{aligned} \tag{3.8}$$

If $\left\|L_h\vec{u} - \vec{f}\right\|_2$ *converges to* 0 *for* $h \to 0$, *then* L_h *is called* consistent. *If there exists an estimate of the form* (3.8), *then* consistency *implies* stability.

In the following we determine error estimates for difference quotients:

Lemma 3.2. *Let* $u \in C^2[0,1]$ *and* $x \in [h, 1-h]$. *Then, for one-sided difference quotients we have the estimate*

$$\left|D_h^{\pm}[u](x) - u'(x)\right| \leq \frac{1}{2}\|u''\|_\infty h .$$

For a central difference quotient and $u \in C^3[0,1]$ *we even have:*

$$|D_h[u](x) - u'(x)| \leq \frac{1}{6}\|u'''\|_\infty h^2 .$$

For D_h^2 *we have: Let* $u \in C^4[0,1]$ *and* $x \in [h, 1-h]$, *then:*

$$\left|D_h^2[u](x) - u''(x)\right| \leq \frac{1}{12}\|u''''\|_\infty h^2 . \tag{3.9}$$

Proof. We prove exemplary the assertion for the central difference quotient. Let $u \in C^3[0,1]$, then it follows from Taylor expansion around $x \in (0,1)$:

$$\begin{aligned} u(x+h) &= u(x) + hu'(x) + \frac{1}{2}h^2u''(x) + \frac{1}{6}h^3u'''(\zeta_+), \\ u(x-h) &= u(x) - hu'(x) + \frac{1}{2}h^2u''(x) - \frac{1}{6}h^3u'''(\zeta_-), \end{aligned}$$

for some $\zeta_\pm$ satisfying $x-h<\zeta_-<x<\zeta_+<x+h$. Therefore

$$u(x+h)-u(x-h)=2hu'(x)+\frac{1}{6}h^3(u'''(\zeta_+)+u'''(\zeta_-))\,,$$

and thus

$$\left|\frac{u(x+h)-u(x-h)}{2h}-u'(x)\right|\le\frac{1}{6}h^2\sup\{|u'''(\zeta)|:\zeta\in[0,1]\}\ ,$$

which gives the assertion. □

Example 3.3. *Applied to the differential equation* (3.1) *we find that, provided the solution of the differential equation is 4× continuously differentiable, that*

$$\left\|\underbrace{L_h}_{=-D_h^2}\vec{u}-\vec{f}\right\|_\infty\le\frac{1}{12}\left\|u''''\right\|_\infty h^2=\frac{1}{12}\left\|f''\right\|_\infty h^2\ .$$

In the following we discretize the operator L defined in (3.1)*. We use the discretization $D_h^2[u]$ for approximating u''. Moreover, the first derivative is approximated by either one of the difference quotients $D_h^+[u]$, $D_h^-[u]$, $D_h[u]$. Using different difference quotients gives different diagonal matrices:*

$$L_h=h^{-2}\begin{bmatrix}d_1 & s_1 & & 0\\ r_2 & d_2 & \ddots & \\ & \ddots & \ddots & s_{n-2}\\ 0 & & r_{n-1} & d_{n-1}\end{bmatrix}\in\mathbb{R}^{(n-1)\times(n-1)}\,, \tag{3.10}$$

where for

- D_h^+*:*

$$\begin{aligned}d_i&=2-hb(x_i)+h^2c(x_i)\,,\\ r_i&=-1\,,\\ s_i&=-1+hb(x_i)\,,\end{aligned} \tag{3.11}$$

- D_h^-*:*

$$\begin{aligned}d_i&=2+hb(x_i)+h^2c(x_i)\,,\\ r_i&=-1-hb(x_i)\,,\\ s_i&=-1\,,\end{aligned} \tag{3.12}$$

- D_h:

$$\begin{aligned} d_i &= 2 + h^2 c(x_i)\,, \\ r_i &= -1 - hb(x_i)/2\,, \\ s_i &= -1 + hb(x_i)/2\,. \end{aligned} \tag{3.13}$$

The approximate solution is determined as the solution of the linear system (3.7)*.*

Definition 3.4. *A difference method has* order of consistence q *if*

$$\left\| L_h \vec{u} - \vec{f} \right\|_\infty = \max |(L_h \vec{u})_i - f_i| \le C h^q\,.$$

Note, that in this definition $\vec{u}$ is the vector of the solution of the infinite dimensional problem at the nodal points.

Theorem 3.5. *Let the solution of the boundary value problem* (3.1) *be* $4\times$ *continuously differentiable (which is for instance the case if b, c, f are* $2\times$ *continuously differentiable). Then the difference method* (3.7) *has the order of consistency q:*

- $q = 2$*, if the central difference quotient D_h is used for approximating u';*
- $q = 1$*, if forward or backward difference quotients $D_h^{\pm}$ are used for approximating u'.*

Chapter 4

Partial Differential Equations

In this chapter we are considering *finite element methods* (FEM) for the solution of elliptic differential equations. We restrict attention to space dimension two for simplicity. The domain Ω, on which we solve the differential equation, has piecewise linear boundary Γ, is bounded and connected.

4.1 Finite Element Methods

The basis of finite element methods are *weak* solutions. We just give a short sketch of the basics of this theory:

Definition 4.1. $H^1(\Omega)$ *denotes the space of square integrable functions with square integrable derivatives and the inner product*

$$\langle u, v \rangle_{H^1(\Omega)} = \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega uv \, dx \, .$$

The associated norm is denoted by $\|\cdot\|_{H^1(\Omega)}$ *and the semi-norm is denoted by*

$$|u|_{H^1(\Omega)} = \int_\Omega |\nabla u|^2 \, dx \, .$$

Functions $u \in H^1(\Omega)$ are not necessarily continuous in Ω, but it is still possible to define boundary values. For us it is not really important how they can be defined rigorously, we just do as if they point evaluations can be done. Of particular importance are the set of zero-Dirichlet data:

$$H_0^1(\Omega) = \left\{ u \in H^1(\Omega) : u|_\Gamma = 0 \right\} \, ,$$

which is a closed linear subspace of $H^1(\Omega)$. For functions in $H_0^1(\Omega)$ the Poincare-Friedrich inequality is valid:

$$\gamma_\Omega \|u\|_{H^1(\Omega)} \leq |u|_{H^1(\Omega)} , \qquad \forall u \in H_0^1(\Omega) . \tag{4.1}$$

After this clarification of notation we are investigating now elliptic differential equations first:

$$L[u] := -\nabla \cdot (\sigma \nabla u) + cu = f \text{ in } \Omega \tag{4.2}$$

with Dirichlet boundary conditions

$$u = 0 \text{ on } \Gamma . \tag{4.3}$$

Aside from some smoothness conditions (which we do not discuss in detail) essential conditions are the following:

$$0 < \sigma_0 \leq \sigma(x) \leq \sigma_\infty \text{ and } 0 \leq c(x) \leq c_\infty .$$

This conditions are essential for ellipticity. A classical solution (referring to standard theory) is one, where the second derivative is continuous.

The basics of weak solutions are partial integration:

$$\begin{aligned}\int_\Omega fv\,dx \underbrace{=}_{(4.2)} &- \int_\Omega \nabla \cdot (\sigma \nabla u) v\,dx + \int_\Omega cuv\,dx \\ &= \int_\Omega \sigma \nabla u \nabla v\,dx + \int_\Omega cuv\,dx - \int_\Gamma v\sigma \frac{\partial u}{\partial n}\,ds .\end{aligned}$$

Definition 4.2. *A weak solution of the* homogenous Dirichlet-problem*, that is of* (4.2) *and* (4.3)*, is a solution of*

$$\int_\Omega fv\,dx = \int_\Omega \sigma \nabla u \nabla v\,dx + \int_\Omega cuv\,dx , \qquad \forall v \in H_0^1(\Omega) . \tag{4.4}$$

Remark 4.3. *The weak solution is unique.*

The *inhomogenous Dirichlet problem* consists in solving (4.2) together with boundary conditions:

$$u = g \text{ on } \Gamma . \tag{4.5}$$

We extend the function g from Γ to Ω and denote such an extension by u_0. With u_0 we reduce (4.2), (4.5) to a homogenous Dirichlet problem. In fact $w := u - u_0$ solves the homogenous Dirichlet problem

$$L[w] = f - L[u_0] , \qquad w|_\Gamma = 0 .$$

The (in-)homogenous *Dirichlet problem* has a unique solution:

Theorem 4.4. *Let σ, c and f be bounded functions satisfying*

$$0 \le c(x) \le c_\infty \text{ and } 0 < \sigma_0 \le \sigma(x) \le c_\infty \; .$$

Then, the Dirichlet problem has a unique weak solution.

The *Neumann problem* consists in the solution of (4.2) with boundary conditions:

$$\sigma \frac{\partial u}{\partial n} = g \text{ on } \Gamma \; . \tag{4.6}$$

The weak form of the equation is again derived by partial integration:

$$\begin{aligned}
\int_\Omega fv\,dx \underbrace{=}_{(4.2)} & \int_\Omega \nabla\cdot(\sigma\nabla u)v\,dx + \int_\Omega cuv\,dx \\
&= \int_\Omega \sigma\nabla u\nabla v\,dx + \int_\Omega cuv\,dx - \int_\Gamma v\sigma\frac{\partial u}{\partial n}\,ds \\
&\underbrace{=}_{(4.6)} \int_\Omega \sigma\nabla u\nabla v\,dx + \int_\Omega cuv\,dx - \int_\Gamma vg\,ds \; .
\end{aligned}$$

Existenca and uniqueness are as follows:

Theorem 4.5. *Let σ, c and f be bounded functions satisfying*

$$0 < c_0 \le c(x) \le c_\infty \text{ and } 0 < \sigma_0 \le \sigma(x) \le c_\infty \; .$$

Then, the Neumann problem has a unique weak solution.

For $c = 0$ we also have weak solutions, provided that

$$\int_\Omega f\,dx = -\int_\Gamma g\,ds \; . \tag{4.7}$$

In this case, however, the solution is not unique, and all solutions differ by a constant. Typically we select the solution, which satisfies $\int_\Omega u\,dx = 0$.

Now, we consider a general strategy for solving elliptic differential equations: Let

$$\begin{aligned}
a(u,v) &:= \int_\Omega \sigma\nabla u\cdot\nabla v\,dx + \int_\Omega cuv\,dx \, , \\
l(v) &:= \int_\Omega fv\,dx \; .
\end{aligned} \tag{4.8}$$

a is called *bilinear* form because it is linear in every component on $V = H^1(\Omega)$. Moreover, l is a linear operator on V. With this notation we have a compact formulation of the differential equation (4.2), (4.3):

$$a(u,v) = l(v)\,, \qquad \forall v \in H_0^1(\Omega)\,. \tag{4.9}$$

Note, that the space $H_0^1(\Omega)$ is designed such that the solution satisfies homogenous Dirichlet conditions.

Definition 4.6. *A bilinear form $a : V \times V \to \mathbb{R}$ on V is called*

- *symmetric, if $a(u,v) = a(v,u)$ for all $u, v \in V$,*
- *continuous, if there exists a number $a_\infty \in \mathbb{R}_+$ such that*

$$|a(u,v)| \le a_\infty \|u\|_V \|v\|_V\,, \qquad \forall u,v \in V\,,$$

- *W-elliptic, if there exists a constant $\alpha_0 > 0$ such that*

$$a(w,w) \ge a_0 \|w\|_V^2\,, \qquad \forall w \in W\,.$$

Proposition 4.7. *Let σ, c and f be bounded functions satisfying*

$$0 \le c(x) \le c_\infty \text{ and } 0 < \sigma_0 \le \sigma(x) \le \sigma_\infty\,.$$

Then a from (4.8) is symmetric and bounded on $V = H^1(\Omega)$ with

$$a_\infty = \max\{\sigma_\infty, c_\infty\}$$

and $H_0^1(\Omega)$-elliptic with constants

$$a_0 = \gamma_\Omega^2 \sigma_0$$

(γ_0 is the Poincare-Friedrich constant (4.1)). Moreover,

$$a(v,v) \ge 0\,, \qquad \forall v \in H^1(\Omega)\,.$$

Proof. • For alle $u, v \in H_0^1(\Omega)$ we have

$$\begin{aligned}
&|a(u,v)| \\
&\underbrace{\leq}_{\Delta-\text{Ineq.}} \left|\int_\Omega \sigma \nabla u \nabla v \, dx\right| + \left|\int_\Omega cuv \, dx\right| \\
&\leq a_\infty \left(\int_\Omega |\nabla u \cdot \nabla v| \, dx + \int_\Omega |uv| \, dx\right) \\
&\underbrace{\leq}_{\text{Cauchy-Schwarz for functions}} a_\infty \left(\|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}\right) \\
&\underbrace{\leq}_{\text{Cauchy-Schwarz for numbers}} a_\infty \left(\sqrt{\|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2} \cdot \sqrt{\|v\|_{L^2(\Omega)}^2 + \|\nabla v\|_{L^2(\Omega)}^2}\right) .
\end{aligned}$$

• For $v \in H^1(\Omega)$ we have:

$$\begin{aligned}
|a(v,v)| &= \int_\Omega \sigma |\nabla v|^2 \, dx + \int_\Omega cv^2 \, dx \\
&\geq \sigma_0 \int_\Omega \sigma |\nabla v|^2 \, dx \\
&\geq 0 .
\end{aligned} \tag{4.10}$$

• From (4.10) it follows for $v \in H_0^1(\Omega)$.

$$|a(v,v)| \geq \sigma_0 |v|_{H_0^1(\Omega)}^2 \underbrace{\geq}_{(4.1)} \sigma_0 \gamma_\Omega^2 \|v\|_{H^1(\Omega)}^2 .$$

□

To determine an approximate solution we use a *Galerkin*-approach: We select a finite dimensional subspace $V_h \subseteq H_0^1(\Omega)$ and determine $u_h \in V_h$ satisfying

$$a(u_h, v_h) = l(v_h) , \qquad \forall v_h \in V_h . \tag{4.11}$$

If $\{\phi_1, \ldots, \phi_n\}$ is a basis of V_h, then the ansatz

$$u_h = \sum_{i=1}^n u_i \phi_i$$

results in the linear equation

$$\begin{aligned} A\vec{u}_h = \vec{b} \text{ with } \vec{u}_h = (u_1, \dots, u_n)^T, \\ A = [a(\phi_i, \phi_j)]_{ij} \in \mathbb{R}^{n\times n}\,, \vec{b} = [l(\phi_j)]_j \in \mathbb{R}^n\,. \end{aligned} \tag{4.12}$$

The matrix A is called *stiffness matrix.*

Example 4.8. *We solve the following Dirichlet-problem by a Galerkin approach:*

$$-u'' = f \ \textit{in}\ (0,1) \ \textit{with}\ u(0) = u(1) = 0\,.$$

Let V_h be the space of linear splines on a equidistant grid:

$$\Delta_h = \{x_i = ih : 0 \le i \le n, h = 1/n\}$$

satisfying homogenous boundary data. The space of linear splines consists of linear combinations of hat-functions Λ_i, $i = 1, \dots, n-1$, which are equal to 1 at the nodal points i/n. That gives the stiffness matrix:

$$a(\Lambda_i, \Lambda_j) = \int_0^1 \Lambda_i'(x)\Lambda_j'(x)\,dx = \begin{cases} 2/h & i = j \\ -1/h & |i-j| = 1 \\ 0 & \textit{else} \end{cases}$$

In this case the Galerkin method requires to solve the equation:

$$\frac{1}{h}\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \vdots & \\ & \vdots & \vdots & -1 \\ & & -1 & 2 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \end{pmatrix},$$

where

$$b_i = \int_0^1 f(x)\Lambda_i(x)\,dx\,, \qquad i = 1, \dots, n-1\,.$$

Remark 4.9. *In order to solve the inhomogenous Dirichlet problem, we determine a function $\omega : \Omega \to \mathbb{R}$, which extends the boundary data onto Ω: Then we are looking for solution $u^\dagger$ of*

$$a(u_0 + \omega, v) = l(v)\,, \qquad \forall v \in H_0^1(\Omega)\,, \tag{4.13}$$

the solution of the inhomogenous problem is $u^\dagger + \omega$.

For the Neumann problem the right hand side reads as follows:

$$l(v) = \int_\Omega f v \, dx \; + \int_\Gamma g v \, ds$$

and the linear system reads as follows:

$$a(u, v) = l(v) \, , \qquad \forall v \in H^1(\Omega) \, .$$

For realizing a Galerkin method we need an appropriate *ansatz space* $V_h \subseteq H^1(\Omega)$. Typically finite Element methods are based on *triangulations* of the domain Ω.

Definition 4.10. *A set of open triangles $\Gamma = \{T_1, \ldots, T_m\}$ is called* regular triangulation *of Ω, if*

1. $T_i \cap T_j = \emptyset \quad \forall i \neq j$,
2. $\bigcup_{i=1}^m \overline{T_i} = \overline{\Omega}$,
3. *for $i \neq j$ we have either*
   (a) $\overline{T_i} \cap \overline{T_j} = \emptyset$,
   (b) $\overline{T_i} \cap \overline{T_j}$ *is a joint corner of T_i and T_j, or*
   (c) *a common edge.*

The corners of the triangle are called corners.

On a triangulation we define linear ansatz functions (in analogy to linear splines).

Theorem 4.11. *Let Γ be a regular triangulation of a polygonal domainΩ with nodal points x_i, $i = 1, \ldots, n$. Then there exist continuous functions $\Lambda_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, n$, satisfying:*

1. $\Lambda_i(x_j) = \delta_{ij} \, , \quad i, j = 1, \ldots, n \, ,$
2. $\Lambda_i(x) = \beta_{ik} + \alpha_{ik} \cdot x$ *for* $x \in T_k$ *with* $\alpha_{ik} \in \mathbb{R}^2$, $\beta_{ik} \in \mathbb{R}$.

$V^\Gamma = span\{\Lambda_1, \ldots, \Lambda_n\}$ consists of piecewise linear functions with respect to Γ.

The gradient of an element V^Γ is piecewise constant and we have $V^\Gamma \subseteq H^1(\Omega)$.

Definition 4.12. *The tupel* (Γ, V^Γ) *is called* finite elements.

The analog of Lagrange interpolation for finite elements reads as follows:

Theorem 4.13. *Let* Γ *be a regular triangulation of* $\Omega \subseteq \mathbb{R}^2$ *with nodal points* $\{x_i : i = 1, \ldots, n\}$. *Let be given* $\{y_i : i = 1, \ldots, n\}$. *Then* $\psi = \sum_{i=1}^n y_i \Lambda_i \in V^\Gamma$ *and*

$$\psi(x_i) = y_i\,, \quad i = 1, \ldots, n\,.$$

4.2 Stiffness Matrix

A finite element method is reduced to the solution of the linear matrix equation

$$A\vec{u}_h = b\,, \tag{4.14}$$

with *stiffness matrix* A, where

$$a_{ij} = a(\phi_i, \phi_j) = \int_\Omega \sigma \nabla\phi_i \cdot \nabla\phi_j + c\phi_i\phi_j \, dx\ . \tag{4.15}$$

We empghasize that the matrix is sparse.

For every triangle $T_k \in \Gamma$ the matrix

$$S_k = \left[\int_{T_k} \sigma \nabla\phi_i \cdot \nabla\phi_j + c\phi_i\phi_j \, dx\right]_{ij} \in \mathbb{R}^{n\times n} \tag{4.16}$$

consists of all integrals over the triangle T_k. These matrices S_k are called *element stiffness matrices*. Because of

$$\begin{aligned} a(\phi_i, \phi_j) &= \int_\Omega \sigma \nabla\phi_i \cdot \nabla\phi_j + c\phi_i\phi_j \, dx \\ &= \sum_{k=1}^m \int_{T_k} \sigma \nabla\phi_i \cdot \nabla\phi_j + c\phi_i\phi_j \, dx \end{aligned}$$

we have

$$A = \sum_{k=1}^m S_k\ . \tag{4.17}$$

To determine the element stiffness matrices one uses the transformation

$$\Phi(s,t) = x_1 + s(x_2 - x_1) + t(x_3 - x_1)\,, \tag{4.18}$$

which maps the *reference triangle*

$$D = \left\{z = (s,t)^t : s > 0, t > 0, s + t < 1\right\} \tag{4.19}$$

onto the triangle $T \in \Gamma$ with corners $x_i = (\zeta_i, \eta_i)^t$, $i = 1, 2, 3$. Therefore, we have

$$\Phi'(s,t) = [x_2 - x_1 \quad x_3 - x_1] = \begin{bmatrix} \zeta_2 - \zeta_1 & \zeta_3 - \zeta_1 \\ \eta_2 - \eta_1 & \eta_3 - \eta_1 \end{bmatrix} .$$

These two vectors are linear independent in a triangle which is not degenerate. Thus

$$d = \det\Phi' = (\zeta_2 - \zeta_1)(\eta_3 - \eta_1) - (\zeta_3 - \zeta_1)(\eta_2 - \eta_1) \neq 0 .$$

Thus

$$\Phi'^{-1}(x) = \frac{1}{d}\begin{bmatrix} \eta_3 - \eta_1 & \zeta_1 - \zeta_3 \\ \eta_1 - \eta_2 & \zeta_2 - \zeta_1 \end{bmatrix}(x) .$$

Example 4.14. *We are calculating the element stiffness matrix $S = [s_{ij}]$ for $L[u] = -\Delta u$ and a triangle $T \in \Gamma$ with corners x_1, x_2 and x_3. We denote by Λ_i, $i = 1, 2, 3$ the hut functions, with nodal value 1 at x_i and 0 else, respectively. Therefore,*

$$\begin{aligned} s_{ij} &= \int_T \nabla_x\Lambda_i(x) \cdot \nabla_x\Lambda_j(x)\,dx \\ &= \int_D \nabla_x\Lambda_i(\Phi(z)) \cdot \nabla_x\Lambda_j(\Phi(z))\,|det\Phi'|\;dz \\ &= |d| \int_D \Phi'^{-t}\nabla_z\Lambda_i(\Phi(z))) \cdot (\Phi'^{-t}\nabla_z\Lambda_j(\Phi(z))\,dz . \end{aligned}$$

The function $\Lambda_i(\Phi(\cdot))$ is again an hat function over D with nodal value 1 at z_i and 0 else. Therefore,

$$G := \begin{bmatrix} \nabla_z(\Lambda_1(\Phi(\cdot))^t \\ \nabla_z(\Lambda_2(\Phi(\cdot))^t \\ \nabla_z(\Lambda_3(\Phi(\cdot))^t \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

Therefeore, the integrands s_{ij} are constant on D. Moreover, the area of D is 0.5. Consequently,

$$\begin{aligned} [s_{ij}]_{i,j} &= \frac{|d|}{2} G\Phi'^{-1}\Phi'^{-t}G^t \\ &= \frac{1}{2\,|d|}\begin{bmatrix} \eta_2 - \eta_3 & \zeta_3 - \zeta_2 \\ \eta_3 - \eta_1 & \zeta_1 - \zeta_3 \\ \eta_1 - \eta_2 & \zeta_2 - \zeta_1 \end{bmatrix}\begin{bmatrix} \eta_2 - \eta_3 & \eta_3 - \eta_1 & \eta_1 - \eta_2 \\ \zeta_3 - \zeta_2 & \zeta_1 - \zeta_3 & \zeta_2 - \zeta_1 \end{bmatrix} \in \mathbb{R}^{3\times 3} \end{aligned}$$

4.3 Parabolic Differential Equations

In this section we study the numerical solution of parabolic initial value problems exemplary for the model problem

$$\begin{aligned} u_t + L[u] &= f \text{ for } (x,t) \in \Omega \times \mathbb{R}^+ , \\ u(x,t) &= 0 \text{ for } (x,t) \in \Gamma \times \mathbb{R}^+ , \\ u(x,0) &= u^0(x) \text{ for } x \in \Omega . \end{aligned} \tag{4.20}$$

u^0 is the initial value and $L[u]$ is as in (4.2) defined. As usual we associate with L the bilinear form:

$$a(v,w) = \int_\Omega \sigma \nabla v \cdot \nabla w + cvw \, dx . \tag{4.21}$$

4.3.1 Methods of Lines

With the methods of lines we determine for every fixed $t > 0$ an approximation $u_h(t) \in V_h$, which satisfies the equation

$$\int_\Omega u_h'(t) w \, dx + a(u_h(t), w) = \int_\Omega f(t) w \, dx , \qquad \forall w \in V_h . \tag{4.22}$$

The solution u_h can be determined with a Galerkin approach. For this purpose we set

$$u_h(t) = \sum_{j=1}^n \eta_j(t) \Lambda_j (u_h \text{ is a function of space and time}). \tag{4.23}$$

Let $w = \Lambda_k$, $k = 1, \ldots, n$ then from (4.22) we get the system of differential equations:

$$\sum_{j=1}^n \eta_j'(t) \int_\Omega \Lambda_j \Lambda_k \, dx + \sum_{j=1}^n \eta_j(t) a(\Lambda_j, \Lambda_k) = \int_\Omega f(t) \Lambda_k \, dx \\ , \qquad \forall k = 1, \ldots, n . \tag{4.24}$$

Denoting the vectors

$$y(t) = [\eta_j(t)]_j \in \mathbb{R}^n \text{ and } b(t) = \left[\int_\Omega f(t) \Lambda_k \, dx \right]_k \in \mathbb{R}^n ,$$

we can reformulate the system (4.24) to

$$Gy' = b - Ay\,, \quad y(0) = [\eta_j(0)]_j \in \mathbb{R}^n\,, \tag{4.25}$$

where G denotes Gram's matrix of hat-functions in $L^2(\Omega)$ and $A = [a(\Lambda_i, \Lambda_k)]_{jk}$. An equivalent description is

$$y' = G^{-1}b - G^{-1}Ay\,, \quad y(0) = [\eta_j(0)]_j \in \mathbb{R}^n\,. \tag{4.26}$$

The methods is called *(vertical) methods of lines*, because it reduces the PDE to a system of ordinary differential equations with respect time.

4.3.2 Crank-Nicolson Method

For the sake of simplicity of presentation we consider an equidistant grid with grid size $\tau > 0$:

$$\Delta_\tau = \{t_i = i\tau : i \in \mathbb{N}_0\}\ .$$

We define functions with respect to space

$$u_i = \sum_{j=1}^{n} \eta_{ij}\Lambda_j\,,$$

which approximate $u_h(t_i)$. The values η_{ij} approximate also the function values of u at the nodes $\{x_j : j = 1, \ldots, n\}$ of the triangulation Γ:

$$\eta_{ij} \approx u_h(x_i, t_j) \approx u(x_j, t_i)\ .$$

To solve (4.25) in a stable way we use the mid-point integration rule, which results in,

$$\Big(I + \frac{\tau}{2}G^{-1}A\Big)\, y_{i+1} = \Big(I - \frac{\tau}{2}G^{-1}A\Big)\, y_i + \tau G^{-1} b_{i+1/2}\,, \qquad i = 0, 1, 2\ldots, \tag{4.27}$$

where

$$b_{i+1/2} = \left[\int_\Omega f_{i+1/2}\Lambda_j\, dx\right]_j \ \text{with}\ f_{i+1/2} = f(t_i + \tau/2)\ .$$

An equivalente formulation of this method is:

$$\Big(G + \frac{\tau}{2}A\Big)\, y_{i+1} = \Big(G - \frac{\tau}{2}A\Big)\, y_i + \tau b_{i+1/2}\,, \qquad i = 0, 1, 2\ldots, \tag{4.28}$$

Because both G and A are positive definite, so is $G + \frac{\tau}{2}A\,, \qquad \forall tau > 0$. This recursion can be written as a finite dimensional variational problem for $u_{i+1} \in V_h$:

$$\begin{aligned} &\int_\Omega u_{i+1} w\,dx + \frac{\tau}{2} a(u_{i+1}, w) \\ = &\int_\Omega u_i w\,dx - \frac{\tau}{2} a(u_i, w) + \tau \int_\Omega f_{i+1/2} w\,dx\;. \end{aligned} \tag{4.29}$$

This is the Crank-Nicolson method.

4.4 Hyperbolic Differential Equations

In this section we study the numerical solution of conservation equations.

The simplest conservation equation is the transport equation

$$u_t = -a u_x\,, \quad x \in \mathbb{R}\,, t \geq 0\,, \tag{4.30}$$

where we assume $a > 0$. The case $a < 0$ can be treated in an analogous manner by considering $\tilde{u}(t,x) = u(t,-x)$.

For a given initial value und

$$u(x,0) = u^0(x) \tag{4.31}$$

the solution of (4.30) can be determined analytically:

$$u(x,t) = u^0(x - at)\;.$$

Routinely, the numerical solution of hyperbolic conservation equations is implemented via difference methods. For the sake of simplicity we again assume an equidistant grid with stepsize h in space and τ in time, respectively. In this way we get a two-dimensional cartesian grid:

$$\{x_j = jh : j \in \mathbb{Z}\} \times \{t_i = i\tau : i \in \mathbb{N}_0\} \subseteq \mathbb{R} \times \mathbb{R}_0^+\;. \tag{4.32}$$

In each node we determine an approximate value $u_{ij} \approx u(x_j, t_i)$. In the following we use the convention $\vec{u}_i = [u_{ij}]_{j \in \mathbb{Z}}$.

For approximation of the time derivative we use the difference quotient;

$$u_t(x_j, t_i) \approx \frac{u_{i+1,j} - u_{ij}}{\tau}\;.$$

For discretization of the space derivative we use one of the quotients D_h, D_h^+, D_h^- defined in Chapter 3. This results in the following iteration:

$$\vec{u}_{i+1} = \vec{u}_i - \tau a L_h \vec{u}_i = f A_h \vec{u}_i \,, \quad A_h = I - \tau a L_h \,, \tag{4.33}$$

where

$$A_h = \left[\begin{array}{ccc|ccc} \ddots & \ddots & & & & \\ \ddots & \alpha_0 & \alpha_1 & & & \\ & \alpha_{-1} & \alpha_0 & \alpha_1 & & \\ \hline & & \alpha_{-1} & \alpha_0 & \alpha_1 & \\ & & & \alpha_{-1} & \alpha_0 & \ddots \\ & & & & \ddots & \ddots \end{array}\right] , \tag{4.34}$$

where the entries are dependent of L_h as follows:

$$\begin{aligned} D_h^- : &\quad \alpha_{-1} = \gamma a \,, \quad \alpha_0 = 1 - \gamma a \,, \quad \alpha_1 = 0 \,, \\ D_h^+ : &\quad \alpha_{-1} = 0 \,, \quad \alpha_0 = 1 + \gamma a \,, \quad \alpha_1 = -\gamma a \,, \\ D_h : &\quad \alpha_{-1} = \gamma a/2 \,, \quad \alpha_0 = 1 \,, \quad \alpha_1 = -\gamma a/2 \,. \end{aligned} \tag{4.35}$$

In accordance with the initial conditions we set:

$$\vec{u}_0 = [u_{0j}]_{j \in \mathbb{Z}} \text{ with } u_{0j} = u^0(x_j) \,.$$

In the following we investigate some convergence and stability properties. Thus let

$$\vec{\hat{u}}_i = [u(x_j, t_i)]_{j \in \mathbb{Z}} \,, \quad i \in \mathbb{N}_0 \,.$$

Definition 4.15. *A difference method $\vec{u}_{i+1} = A_h \vec{u}_i$ is called consistent $\|\cdot\|_\infty$ if*

$$\left\| \vec{\hat{u}}_{i+1} - A_h \vec{\hat{u}}_i \right\|_\infty = o(h) \,, \qquad \forall i \in \mathbb{N}_0 \,, \tau > 0 \,.$$

If there exists some $q \in \mathbb{N}$ such that

$$\left\| \vec{\hat{u}}_{i+1} - A_h \vec{\hat{u}}_i \right\|_\infty \le C h^q \tau \,, \tag{4.36}$$

then the method has consistency order q.

Theorem 4.16. *The difference method* (4.33) *has consistency order $q = 1$, if L_h is defined via D_h, D_h^+ or D_h^-, respectively.*

Definition 4.17. *The difference method $\vec{u}_{i+1} = A_h\vec{u}_i$ is called stable, if for all $T > 0$ there exists a constant $C_T > 0$ such that*

$$\left\|A_h^i\right\|_\infty \le C_T\,, \text{ for } i\tau = \underbrace{i\gamma h}_{\text{time discretization}} \le T\,. \tag{4.37}$$

Theorem 4.18. *Let us assume that* (4.33) *has consistency ordnung $q \in \mathbb{N}$. Moreover, let us assume that there exists $\sigma > 0$ such that*

$$\|A_h\|_\infty \le 1 + \sigma h\,. \tag{4.38}$$

Then, the difference method is stable, and with the constant C from (4.36) *we have:*

$$\left\|\vec{\hat{u}}_i - \vec{u}_i\right\|_\infty \le \frac{\gamma C}{\sigma} e^{\sigma T/\gamma} h^q\,, \qquad \forall t_i = i\tau \text{ mit } 0 \le t_i \le T\,. \tag{4.39}$$

In the following we analyze (4.33) with the three different possibilities of approximation of the derivative:

Theorem 4.19. *Exclusively the left-sided difference operator D_h^- makes* (4.33) *stable. Thereby, one has to make sure that τ and h satisfy the* Courant-Friedrichs-Levi (CFL) *–condition,*

$$a\gamma = a\frac{\tau}{h} \le 1\,. \tag{4.40}$$

In this case $\|A_h\|_\infty \le 1$ and

$$\inf_{x\in\mathbb{R}} u^0(x) \le u_{ij} \le \sup_{x\in\mathbb{R}} u^0(x)\,, \qquad \forall i \in \mathbb{N}_0\,, j \in \mathbb{Z}\,.$$

If $a > 0$, then we have to use D_h^+ to make (4.33). Algorithms, which in dependency of the coefficients of the difference quotients locally differently are called *upwind – schemes*. The name comes form the fact that the transport equation also describes wind. Therefore, in dependence of the wind orientation one chooses different difference quotients.

Bibliography