Appendix A

Basic Space Discretisation Methods

Partial differential equations (PDEs) appear in all fields in which mathematical models are applied to describe the time evolution of certain space-dependent quantities. Since these PDEs might be of complicated form, it is often difficult or even impossible to solve them analytically. Therefore, one has to apply numerical schemes to obtain an approximation to the exact solution. In the present Appendix we deal with *space discretisation* methods, i.e., with numerical procedures approximating the spatial differential operator appearing in the PDE.

Let us consider the following linear PDE in a general form with the differential operator L, and the unknown function $w : [0, \infty) \times \Omega \to \mathbb{R}$ for $t \in [0, \infty)$ and $x \in \Omega \subseteq \mathbb{R}^d$:

$$\begin{aligned} \partial_t^\alpha w(t,x) &= Lw(t,x), \quad t > 0, x \in \Omega \\ w(0,x) &= w_0(x), \quad x \in \Omega, \end{aligned} \tag{A.1}$$

subject to appropriate boundary conditions. Here Ω denotes an open set, and $\alpha \in \mathbb{N}$ is the order of the time derivative. We are merely interested in the cases $\alpha = 1$, when $\partial_t w$ appears on the left-hand side of equation (A.1), or $\alpha = 2$ with $\partial_{tt} w$. We note that for $\alpha = 0$, the case $f(x) = Lw(x)$ with a given function $f : \Omega \to \mathbb{R}$ is also possible.

Example A.1. a) *Heat equation*

$$\partial_t w(t,x) = \partial_{xx} w(t,x)$$

in one dimension with $\alpha = 1$, $x \in (0, \pi)$, and $Lw(t,x) := \partial_{xx} w(t,x)$. Together with the boundary condition $w(t,0) = w(t,\pi) = 0$, this problem was already investigated in Section 1.1.

b) *Transport equation*

$$\partial_t w(t,x) = \partial_x w(t,x)$$

in one dimension with $\alpha = 1$, $x \in (0,1)$, and $Lw(t,x) := \partial_x w(t,x)$. Together with the boundary condition $w(t,1) = 0$, this problem was already investigated in Section 1.2.

c) *Wave equation*

$$\partial_{tt} w(t,x) = \partial_{xx} w(t,x)$$

in one dimension with $\alpha = 2$, $x \in (0,1)$, and $Lw(t,x) := \partial_{xx} w(t,x)$ with $w(t,0) = \partial_t w(t,0) = 0$. We will come back to this example later on during the lectures.

The main idea behind the simplest discretisation type[1] of PDEs is the following. First we discretise the operator L on the right-hand side with respect to the space variable x. By this we obtain an ordinary differential equation which is then solved by using time discretisation methods.

In what follows we discuss two ways of approximating the operator L and briefly introduce the two main classes of space discretisation methods: finite differences and Galerkin methods.

[1] It is called the *method of lines*.

A.1 Finite difference methods

Here we discuss finite difference methods in the one-dimensional case for the heat and transport equation, that is, for $d = 1$ and $\Omega = (a, b)$. In order to discretise problems like (A.1) in 1D, we divide the interval (a, b) into N pieces of sub-intervals with length $\Delta x = \frac{b-a}{N}$. Then the points $x_j = a + j\Delta x$, $j = 0, ..., N$, are called **grid points**.

Now, we would like to approximate the exact solution at time level $t \geq 0$ and at the points x_j, i.e., $w(t, x_j) = w(t, a+j\Delta x)$ by the values $w_j(t)$ for $j = 0, ..., N$. To this end we use Taylor's formula with respect to the second variable:

$$\begin{aligned} w(t, x_{j+1}) &= w(t, x_j + \Delta x) \\ &= w(t, x_j) + \Delta x \cdot \partial_x w(t, x_j) + \tfrac{1}{2}(\Delta x)^2 \cdot \partial_{xx} w(t, x_j) + \cdots , \\ \text{or} \qquad w(t, x_{j-1}) &= w(t, x_j - \Delta x) \\ &= w(t, x_j) - \Delta x \cdot \partial_x w(t, x_j) + \tfrac{1}{2}(\Delta x)^2 \cdot \partial_{xx} w(t, x_j) + \cdots . \end{aligned}$$

The first- and second-order partial derivatives of w with respect to the space variable x, appearing in the expression of Lw, can now be written as:

$$\begin{aligned} \partial_x w(t, x_j) &= \frac{w(t, x_{j+1}) - w(t, x_j)}{\Delta x} + \mathcal{O}\big((\Delta x)^2\big), \quad \text{or} \\ \partial_x w(t, x_j) &= \frac{w(t, x_j) - w(t, x_{j-1})}{\Delta x} + \mathcal{O}\big((\Delta x)^2\big), \quad \text{and} \\ \partial_{xx} w(t, x_j) &= \frac{w(t, x_{j+1}) - 2w(t, x_j) + w(t, x_{j-1})}{(\Delta x)^2} + \mathcal{O}\big((\Delta x)^3\big), \end{aligned}$$

where $\big|\frac{\mathcal{O}((\Delta x)^p)}{(\Delta x)^p}\big| \leq$ const. for small values of Δx. Neglecting the higher-order terms motivates us to define the following approximation formulae to the spatial derivatives:

$$\begin{aligned} \partial_x w(t, x_j) &\approx \frac{w_{j+1}(t) - w_j(t)}{\Delta x}, \qquad (A.2) \\ \text{or} \qquad \partial_x w(t, x_j) &\approx \frac{w_j(t) - w_{j-1}(t)}{\Delta x}, \\ \text{and} \qquad \partial_{xx} w(t, x_j) &\approx \frac{w_{j+1}(t) - 2w_j(t) + w_{j-1}(t)}{(\Delta x)^2} \end{aligned}$$

for $j = 1, ..., N-1$. In general, one can derive the approximating formula for any derivative by using the Taylor series expansion of the function $w(t, x_j)$, and taking into account that $x_j = x_{j\pm k} \pm k\Delta x$ for any k with $x_{j\pm k} \in [a, b]$.

Example A.2. For Example A.1.a), the spatially discretised[2] problem takes the form:

$$\tfrac{\mathrm{d}}{\mathrm{d}t} w_j(t) = \frac{w_{j+1}(t) - 2w_j(t) + w_{j-1}(t)}{(\Delta x)^2}, \quad \text{for } j = 1, ..., N-1.$$

The cases $j = 0$ and $j = N$ are given by the boundary condition $w(0, t) = 0$ as $w_0(t) = w_N(t) = 0$ for all $t \geq 0$. The ordinary differential equations above can be formulated as a system of ordinary

[2] It is sometimes called *semi-discretisation*.

differential equations $\frac{\mathrm{d}}{\mathrm{d}t}W(t) = MW(t)$ with the matrix

$$M = \frac{1}{(\Delta x)^2} \cdot \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix} \in \mathbb{R}^{(N-1)\times(N-1)} \tag{A.3}$$

and the vector $W(t) = \big(w_1(t), ..., w_{N-1}(t)\big) \in \mathbb{R}^{N-1}$. We note that matrix M is of special form, it is a *tridiagonal matrix*, i.e., $M = \frac{1}{(\Delta x)^2}\mathrm{tridiag}(1, -2, 1)$ that has non-zero elements only in its main diagonal and sub-diagonals. Linear systems of this type are much easier to treat numerically.

Example A.3. Example A.1.b) can be spatially discretised as follows:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}w_j(t) = \frac{w_{j+1}(t) - w_j(t)}{\Delta x}, \quad \text{for } j = 1, ..., N$$

with $w_0(t) = 0$ for all $t \geq 0$. Introducing the vector $W(t) = \big(w_1(t), ..., w_N(t)\big) \in \mathbb{R}^N$ and the matrix $M = \frac{1}{\Delta x}\mathrm{tridiag}(-1, 1, 0) \in \mathbb{R}^{N\times N}$, we have the system of ordinary differential equations $\frac{\mathrm{d}}{\mathrm{d}t}W(t) = MW(t)$.

A.2 Galerkin methods

In contrast to finite difference methods, which approximate the exact solution at certain grid points, Galerkin methods use a linear combination of some basis functions. This time let us formulate the problem on the abstract Hilbert space H which is equipped with the inner product $\langle\cdot,\cdot\rangle$. Then for some operator $A : D(A) \subseteq H \to H$ and a given $f \in H$ we consider the following problem:

$$f = Au \quad \text{in } H. \tag{A.4}$$

For instance, let $Au = u''$ on $H = \mathrm{L}^2(0, \pi)$, see Section 1.1. Taking finite dimensional subspaces $H_m \subset H$, $\dim H_m = m$, with a corresponding basis $\{\varphi_1^m, ..., \varphi_m^m\} \subset H_m$, each element $u_m \in H_m$ can be written as the linear combination of the basis functions, i.e.,

$$u_m = \sum_{k=1}^{m} c_k \varphi_k^m$$

with coefficients $c_k \in \mathbb{C}$, $k = 1, ..., m$. The main idea behind the Galerkin methods is that the exact solution $u \in H$ is approximated by a sequence $(u_m) \subset H_m$ for $m \to \infty$. For notational simplicity we shall drop the superscript from φ_j^m and write φ_j only.

Take the inner product of both sides of problem (A.4) with φ_j for $j = 1, ..., m$:

$$\langle f, \varphi_j\rangle = \langle Au, \varphi_j\rangle, \quad j = 1, ..., m.$$

We define the Galerkin methods by replacing u by $u_m = \sum_{k=1}^m c_k\varphi_k$ in the equation above. Using the linearity of A, we obtain

$$\langle f, \varphi_j\rangle = \left\langle A\sum_{k=1}^{m} c_k\varphi_k, \varphi_j \right\rangle = \sum_{k=1}^{m} c_k\big\langle A\varphi_k, \varphi_j\big\rangle, \quad j = 1, ..., m.$$

The definitions of the vectors

$$\Phi := \begin{pmatrix} \langle f, \varphi_1 \rangle \\ \vdots \\ \langle f, \varphi_m \rangle \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

and the matrix

$$A_m = \begin{pmatrix} \langle A\varphi_1, \varphi_1 \rangle & \dots & \langle A\varphi_m, \varphi_1 \rangle \\ \langle A\varphi_1, \varphi_2 \rangle & \dots & \langle A\varphi_m, \varphi_2 \rangle \\ \vdots & & \vdots \\ \langle A\varphi_1, \varphi_m \rangle & \dots & \langle A\varphi_m, \varphi_m \rangle \end{pmatrix}$$

enable us to formulate the problem as a system of linear equations

$$A_m C = \Phi.$$

The idea and "procedure" of the Galerkin methods is the following.

1. Choose a subspace $H_m \subset D(A) \subset H$ with $\dim H_m = m$, and a basis $\{\varphi_1, ..., \varphi_m\} \subset H_m$.
2. Solve the system $A_m C = \Phi$. Its solution is $C = \big(c_1, ..., c_m\big)$.
3. The approximation u_m to u is then constructed as $u_m = \sum_{k=1}^m c_k \varphi_k$.

Under appropriate assumptions one can prove that $u_m \to u$ as $m \to \infty$ in the H-norm.

We remark that in some special cases the Galerkin method can be formulated as a variational problem called **Ritz method**, see Exercise 2.

The question arises how to choose the basis functions φ_j, $j = 1, ..., m$. There are two basic possibilities: (i) eigenfunctions of A, (ii) functions being zero outside a small interval where they piecewise coincide with low order polynomials. The corresponding methods are called **spectral method** and **finite element method**, respectively.

Example A.4. Consider the one-dimensional *Poisson equation* on $\mathrm{L}^2(0, \pi)$ with a given function f and homogeneous boundary condition:

$$\begin{aligned} \tfrac{\mathrm{d}^2}{\mathrm{d}x^2} w(x) &= f(x), \quad x \in (0, \pi) \\ w(0) &= w(\pi) = 0. \end{aligned} \tag{A.5}$$

Here we have $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ and $D(A) = \mathrm{H}^2(0, \pi) \cap \mathrm{H}_0^1(0, \pi)$. In order to solve this equation by using the Galerkin method, we choose a finite dimensional subspace $H_m \subset D(A)$ and the appropriate basis functions φ_j, $j = 1, ..., m$. As before, we define the approximation $w_m \in H_m$ as:

$$w_m(x) := \sum_{k=1}^m c_k \varphi_k(x)$$

with coefficients c_k, $k = 1, ..., m$, to be determined from the system

$$\begin{aligned} \sum_{k=1}^m c_k \langle A\varphi_k, \varphi_j \rangle &= \langle f, \varphi_j \rangle \\ \sum_{k=1}^m c_k \int_0^\pi (A\varphi_k)(x) \varphi_j(x) \mathrm{d}x &= \int_0^\pi f(x) \varphi_j(x) \mathrm{d}x \\ \sum_{k=1}^m c_k \int_0^\pi \tfrac{\mathrm{d}^2}{\mathrm{d}x^2} \varphi_k(x) \varphi_j(x) \mathrm{d}x &= \int_0^\pi f(x) \varphi_j(x) \mathrm{d}x. \end{aligned}$$

Integrating the left-hand side by parts, we obtain

$$\sum_{k=1}^{m} c_k \left(\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\varphi_j(x)\Big|_{x=0}^{x=\pi} - \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x) \right),$$

and hence,

$$\sum_{k=1}^{m} -c_k \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x)\mathrm{d}x = \int_0^\pi f(x)\varphi_j(x)\mathrm{d}x \quad \text{for } j = 1, ...m, \tag{A.6}$$

where we used the "boundary condition" of the basis functions (they vanish at 0 and at π).

In what follows we determine the approximate solution $w_m(x)$ of problem (A.5) by applying two different sets of basis functions which correspond to spectral and finite element methods, respectively.

a) *Spectral method* (cf. Example 3.3): Since $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ with $D(A) = \mathrm{H}^2(0,\pi) \cap \mathrm{H}_0^1(0,1)$, we can choose the finite dimensional subspace as $\mathrm{lin}\{\sin(jx) : j = 1, ..., m\}$ (cf. Exercise 1.1). Then equation (A.6) results in

$$\begin{aligned}\sum_{k=1}^{m} -c_k \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\sin(kx)\tfrac{\mathrm{d}}{\mathrm{d}x}\sin(jx)\mathrm{d}x &= \sum_{k=1}^{m} -c_k \int_0^\pi k\cos(kx) j\cos(jx)\mathrm{d}x \\ &= \sum_{k=1}^{m} -c_k jk \int_0^\pi \cos(kx)\cos(jx)\mathrm{d}x = \int_0^\pi f(x)\sin(jx)\mathrm{d}x.\end{aligned}$$

Due to the orthogonality of basis functions, the Kronecker delta δ_{jk} appears on the left-hand side:

$$\sum_{k=1}^{m} -c_k jk\delta_{jk}\frac{\pi}{2} = \int_0^\pi f(x)\sin(jx)\mathrm{d}x,$$

which leads to the values

$$c_j = -\frac{2}{j^2\pi}\int_0^\pi f(x)\sin(jx)\mathrm{d}x \quad \text{for all } j = 1, ..., m.$$

The approximation $w_m(x)$ to $w(x)$ is then

$$w_m(x) = -\frac{2}{\pi}\sum_{k=1}^{m}\frac{1}{k^2}\sin(kx)\int_0^\pi f(s)\sin(ks)\mathrm{d}s.$$

In this case, the matrix A_m has entries $(A_m)_{jk} = -jk\frac{\pi}{2}\delta_{jk}$ only in its main diagonal which contains then the square numbers $1, ..., m^2$ multiplied by $-\frac{\pi}{2}$.

b) *Finite element method*: Another possible choice for basis functions are the functions

$$\varphi_j(x) := \begin{cases} 0, & \text{for } x < (j-1)\Delta x \\ \dfrac{x}{\Delta x} - (j-1), & \text{for } (j-1)\Delta x \le x < j\Delta x \\ (j+1) - \dfrac{x}{\Delta x}, & \text{for } j\Delta x \le x < (j+1)\Delta x \\ 0, & \text{for } (j+1)\Delta x \le x, \end{cases}$$

which are sometimes called "hat functions", see Example 3.4. Here $\Delta x = \frac{\pi}{m}$ for some $m \in \mathbb{N}$. Their first derivative exists piecewise and can be calculated easily:

$$\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x) := \begin{cases} 0, & \text{for } x < (j-1)\Delta x \\ \dfrac{1}{\Delta x}, & \text{for } (j-1)\Delta x < x < j\Delta x \\ -\dfrac{1}{\Delta x}, & \text{for } j\Delta x < x < (j+1)\Delta x \\ 0, & \text{for } (j+1)\Delta x < x. \end{cases}$$

Equation (A.6) yields

$$\sum_{k=1}^{m} -c_k \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x)\mathrm{d}x = \Delta x \left(-c_{j-1}\frac{-1}{(\Delta x)^2} - c_j\frac{2}{(\Delta x)^2} - c_{j+1}\frac{-1}{(\Delta x)^2} \right)$$

$$= \frac{1}{\Delta x}(c_{j-1} - 2c_j + c_{j+1}) = \int_0^\pi f(x)\varphi_j(x)\mathrm{d}x \quad \text{for } j = 1, ..., m,$$

with the basis functions φ_j defined above. The matrix A_m has now the tridiagonal form $A_m = \frac{1}{\Delta x}\mathrm{tridiag}(1, -2, 1)$.

Note that the basis functions φ_j, $j = 1, ..., m$, do not belong to $\mathrm{H}^2(0, \pi)$, hence the matrix A_m has to be defined by using (A.6), i.e., the weak formulation is necessary here.

Exercises

1. Consider the heat equation in two dimensions for $(x, y) \in \Omega = (0, \pi) \times (0, \pi)$ and $t \geq 0$:

$$\partial_t w(t, x, y) = \partial_{xx} w(t, x, y) + \partial_{yy} w(t, x, y)$$

with the boundary condition

$$w(t, x, y) = 0 \qquad \text{on } \partial\Omega,$$

where $\partial\Omega$ denotes the boundary of Ω. Derive the form of the corresponding matrix obtained when applying finite differences to discretise the operator $L = \partial_{xx} + \partial_{yy}$ (cf. matrix M in (A.3)).

2. Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and $A : D(A) \subset H \to H$ be a linear densely defined operator possessing the following properties:

a) A is *symmetric* on $D(A)$, that is, $\langle Au, v\rangle = \langle u, Av\rangle$ for all $u, v \in D(A)$, and

b) A is *strongly elliptic*, that is, there exists a constant $c > 0$ such that $\langle Au, u\rangle \geq c\|u\|^2$ for all $u \in D(A)$.

For all $v \in D(A)$ and a given element $f \in H$ define the functional $F : D(A) \to \mathbb{R}$ by

$$F(v) := \langle Av, v\rangle - 2\langle f, v\rangle.$$

Show that if $Au = f$ for $u \in D(A)$ then the functional F is minimal, i.e. $F(u) < F(v)$ for all $v \in D(A)$, $v \neq u$.

3. Derive the form of matrix A_m in Examples A.4.a) and b).