

CHAPTER 1

PARTIAL DIFFERENTIAL EQUATIONS

Many natural processes can be sufficiently well described on the macroscopic level, without taking into account the individual behavior of molecules, atoms, electrons, or other particles. The averaged quantities such as the deformation, density, velocity, pressure, temperature, concentration, or electromagnetic field are governed by partial differential equations (PDEs). These equations serve as a language for the formulation of many engineering and scientific problems. To give a few examples, PDEs are employed to predict and control the static and dynamic properties of constructions, flow of blood in human veins, flow of air past cars and airplanes, weather, thermal inhibition of tumors, heating and melting of metals, cleaning of air and water in urban facilities, burning of gas in vehicle engines, magnetic resonance imaging and computer tomography in medicine, and elsewhere. Most PDEs used in practice only contain the first and second partial derivatives (we call them second-order PDEs).

Chapter 1 provides an overview of basic facts and techniques that are essential for both the qualitative analysis and numerical solution of PDEs. After introducing the classification and mentioning some general properties of second-order equations in Section 1.1, we focus on specific properties of elliptic, parabolic, and hyperbolic PDEs in Sections 1.2–1.4. Indeed, there are important PDEs which are not of second order. To mention at least some of them, in Section 1.5 we discuss first-order hyperbolic problems that are frequently used to model transport processes such as, e.g., inviscid fluid flow. Fourth-order problems rooted in the bending of elastic beams and plates are discussed later in Chapter 6.

1.1 SELECTED GENERAL PROPERTIES

Second-order PDEs (or PDE systems) encountered in physics usually are either elliptic, parabolic, or hyperbolic. Elliptic equations describe a special state of a physical system, which is characterized by the minimum of certain quantity (often energy). Parabolic problems in most cases describe the evolutionary process that leads to a steady state described by an elliptic equation. Hyperbolic equations describe the transport of some physical quantities or information, such as waves. Other types of second-order PDEs are said to be undetermined. In this introductory text we restrict ourselves to linear problems, since nonlinearities induce additional aspects whose understanding requires the knowledge of nonlinear functional analysis.

1.1.1 Classification and examples

Let \mathcal{O} be an open connected set in \mathbb{R}^n . A sufficiently general form of a linear second-order PDE in n independent variables $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ is

$$-\sum_{i,j=1}^n \frac{\partial}{\partial z_i} \left(a_{ij} \frac{\partial u}{\partial z_j} \right) + \sum_{i=1}^n \left(\frac{\partial}{\partial z_i} (b_i u) + c_i \frac{\partial u}{\partial z_i} \right) + a_0 u = f, \quad (1.1)$$

where $a_{ij} = a_{ij}(\mathbf{z})$, $b_i = b_i(\mathbf{z})$, $c_i = c_i(\mathbf{z})$, $a_0 = a_0(\mathbf{z})$ and $f = f(\mathbf{z})$. For all derivatives to exist in the classical sense, the solution and the coefficients have to satisfy the following regularity requirements: $u \in C^2(\mathcal{O})$, $a_{ij} \in C^1(\mathcal{O})$, $b_i \in C^1(\mathcal{O})$, $c_i \in C^1(\mathcal{O})$, $a_0 \in C(\mathcal{O})$, $f \in C(\mathcal{O})$. These regularity requirements will be reduced later when the PDE is formulated in the weak sense, and additional conditions will be imposed in order to ensure the existence and uniqueness of solution. If the functions a_{ij} , b_i , c_i , and a_0 are constants, the PDE is said to be with constant coefficients. Since the order of the partial derivatives can be switched for any twice continuously differentiable function u , it is possible to symmetrize the coefficients a_{ij} by defining

$$a_{ij}^{new} := (a_{ij}^{orig} + a_{ji}^{orig})/2$$

and adjusting the other coefficients accordingly so that the equation remains in the form (1.1). This is left to the reader as an exercise. Based on this observation, in the following we always will assume that the coefficient matrix $A(\mathbf{z}) = \{a_{ij}\}_{i,j=1}^n$ is symmetric.

Recall that a symmetric $n \times n$ matrix A is said to be positive definite if

$$\mathbf{v}^T A \mathbf{v} > 0 \quad \text{for all } 0 \neq \mathbf{v} \in \mathbb{R}^n$$

and positive semidefinite if

$$\mathbf{v}^T A \mathbf{v} \geq 0 \quad \text{for all } \mathbf{v} \in \mathbb{R}^n.$$

Analogously one defines negative definite and negative semidefinite matrices by turning the inequalities. Matrices which do not belong to any of these types are said to be indefinite.

Definition 1.1 (Elliptic, parabolic and hyperbolic equations) Consider a second-order PDE of the form (1.1) with a symmetric coefficient matrix $A(\mathbf{z}) = \{a_{ij}\}_{i,j=1}^n$.

1. The equation is said to be elliptic at $\mathbf{z} \in \mathcal{O}$ if $A(\mathbf{z})$ is positive definite.
2. The equation is said to be parabolic at $\mathbf{z} \in \mathcal{O}$ if $A(\mathbf{z})$ is positive semidefinite, but not positive definite, and the rank of $[A(\mathbf{z}), b(\mathbf{z}) + c(\mathbf{z})]$ is equal to n .

3. The equation is said to be hyperbolic at $z \in \mathcal{O}$ if $A(z)$ has one negative and $n - 1$ positive eigenvalues.

An equation is called elliptic, parabolic, or hyperbolic in the set \mathcal{O} if it is elliptic, parabolic, or hyperbolic everywhere in \mathcal{O} , respectively.

Remark 1.1 (Temporal variable t) In practice we distinguish between time-dependent and time-independent PDEs. If the equation is time-independent, we put $n = d$ and $z = \mathbf{x}$, where d is the spatial dimension and \mathbf{x} the spatial variable. This often is the case with elliptic equations. If the quantities in the equation depend on time, which often is the case with parabolic and hyperbolic equations, we put $n = d + 1$ and $z = (\mathbf{x}, t)$, where t is the temporal variable. In such case the set \mathcal{O} represents some space-time domain. If the spatial part of the space-time domain \mathcal{O} does not change in time, we talk about a space-time cylinder $\Omega \times (0, T)$, where $\Omega \subset \mathbb{R}^d$ and $(0, T)$ is the corresponding time interval.

Notice that, strictly speaking, the type of the PDE in Definition 1.1 is not invariant under multiplication by -1 . For example, the equation

$$-\Delta u = f \quad \left(\text{where } \Delta = \sum_{i=1}^3 \frac{\partial^2}{\partial x_i^2} \text{ in } \mathbb{R}^3 \right) \quad (1.2)$$

is elliptic everywhere in \mathbb{R}^3 since its coefficient matrix A is positive definite,

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

However, the type of the equation

$$\Delta u = -f$$

cannot be determined since its coefficient matrix

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

is negative definite. In such cases it is customary to multiply the equation by (-1) so that Definition 1.1 can be applied. Moreover, notice that Definition 1.1 only applies to second-order PDEs. Later in this text we will discuss two important cases outside of this classification: hyperbolic first-order systems in Section 1.5 and elliptic fourth-order problems in Chapter 6.

Remark 1.2 Sometimes, linear second-order PDEs are found in a slightly different form

$$-\sum_{i,j=1}^n \tilde{a}_{ij}(z) \frac{\partial^2 u}{\partial z_i \partial z_j} + \sum_{i=1}^n \tilde{b}_i(z) \frac{\partial u}{\partial z_i} + \tilde{a}_0(z)u = f(z), \quad (1.3)$$

usually with a symmetric coefficient matrix $\tilde{A}(z) = \{\tilde{a}_{ij}\}_{i,j=1}^n$. When transforming (1.3) into the form (1.1), it is easy to see that the matrices $\tilde{A}(z)$ and $A(z)$ are identical, and

thus either one can be used to determine the ellipticity, parabolicity, or hyperbolicity of the problem. Moreover, if the coefficients \tilde{a}_{ij} and \tilde{b}_i are sufficiently smooth, the two forms are equivalent.

Operator notation It is customary to write elliptic PDEs in a compact form

$$Lu = f,$$

where L defined by

$$Lu = - \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial u}{\partial x_j} \right) + \sum_{i=1}^n \left(\frac{\partial}{\partial x_i} (b_i u) + c_i \frac{\partial u}{\partial x_i} \right) + a_0 u \quad (1.4)$$

is a second-order elliptic differential operator. The part of L with the highest derivatives,

$$- \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial}{\partial x_j} \right), \quad (1.5)$$

is called the principal (leading) part of L . Most parabolic and hyperbolic equations are motivated in physics, and therefore one of the independent variables usually is the time t . The typical operator form of parabolic equations is

$$\frac{\partial u}{\partial t} + Lu = f, \quad (1.6)$$

where L is an elliptic differential operator. Typical second-order hyperbolic equation can be seen in the form

$$\frac{\partial^2 u}{\partial t^2} + Lu = f, \quad (1.7)$$

where again L is an elliptic differential operator. The following examples show simple elliptic, parabolic, and hyperbolic equations.

■ **EXAMPLE 1.1 (Elliptic PDE: Potential equation of electrostatics)**

Let the function $\rho \in C(\overline{\Omega})$ represent the electric charge density in some open bounded set $\Omega \subset \mathbb{R}^d$. If the permittivity ϵ is constant in Ω , the distribution of the electric potential φ in Ω is governed by the Poisson equation

$$-\epsilon \Delta \varphi = \rho. \quad (1.8)$$

Notice that (1.8) does not possess a unique solution, since for any solution φ the function $\varphi + C$, where C is an arbitrary constant, also is a solution. In order to yield a well-posed problem, every elliptic equation has to be endowed with suitable boundary conditions. This will be discussed in Section 1.2.

■ **EXAMPLE 1.2 (Parabolic PDE: Heat transfer equation)**

Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and $q \in C(\overline{\Omega})$ the volume density of heat sources in Ω . If the thermal conductivity k , material density ρ , and specific heat c are constant in Ω , the parabolic equation

$$\frac{\partial \theta}{\partial t} - \frac{k}{\rho c} \Delta \theta = \frac{q}{\rho c} \quad (1.9)$$

describes the evolution of the temperature $\theta(x, t)$ in Ω . The steady state of the temperature ($\partial \theta / \partial t = 0$) is described by the corresponding elliptic equation

$$-k \Delta \theta = q.$$

Similarly to the previous case, the solution θ is not determined by (1.9) uniquely. Parabolic equations have to be endowed with both boundary and initial conditions in order to yield a well-posed problem. This will be discussed in Section 1.3.

■ **EXAMPLE 1.3 (Hyperbolic PDE: Wave equation)**

Let $\Omega \subset \mathbb{R}^d$ be an open bounded set. The speed of sound a can be considered constant in Ω if the motion of the air is sufficiently slow. Then the hyperbolic equation

$$\frac{\partial^2 p}{\partial t^2} - a^2 \Delta p = 0 \quad (1.10)$$

describes the propagation of sound waves in Ω . Here the unknown function $p(x, t)$ represents the pressure, or its fluctuations around some arbitrary constant equilibrium pressure. Again the function p is not determined by (1.10) uniquely. Hyperbolic equations have to be endowed with both boundary and initial conditions in order to yield a well-posed problem. Definition of boundary conditions for hyperbolic problems is more difficult compared to the elliptic or parabolic case, since generally they depend on the choice of the initial data and on the solution itself. We will return to this issue in Example 1.4 and in more detail in Section 1.5.

1.1.2 Hadamard's well-posedness

The notion of well-posedness of boundary-value problems for partial differential equations was established around 1932 by Jacques Salomon Hadamard.

J.S. Hadamard was a French mathematician who contributed significantly to the analysis of Taylor series and analytic functions of the complex variable, prime number theory, study of matrices and determinants, boundary value problems for partial differential equations, probability theory, Markov chains, several areas of mathematical physics, and education of mathematics.

Definition 1.2 (Hadamard's well-posedness) *A problem is said to be well-posed if*

1. *it has a unique solution,*
2. *the solution depends continuously on the given data.*

Otherwise the problem is ill-posed.



Figure 1.1 Jacques Salomon Hadamard (1865–1963).

As the reader may expect, well-posed problems are more pleasant to deal with than the ill-posed ones. The requirement of existence and uniqueness of solution is obvious. The other condition in Definition 1.2 denies well-posedness to problems with unstable solutions. From the point of view of numerical solution of PDEs, the computational domain Ω , boundary and initial conditions, and other parameters are not represented exactly in the computer model. Additional source of error is the finite computer arithmetics. If a problem is well-posed, one has a chance to compute a reasonable approximation of the unique exact solution as long as the data to the problem are approximated reasonably. Such expectation may not be realistic at all if the problem is ill-posed.

The concept of well-posedness deserves to be discussed in more detail. First let us show in Example 1.4 that well-posedness may be violated by endowing a PDE with wrong boundary conditions.

■ **EXAMPLE 1.4 (Ill-posedness due to wrong boundary conditions)**

Consider an interval $\Omega = (-a, a)$, $a > 0$, and the (inviscid) Burgers' equation

$$\frac{\partial}{\partial t}u(x, t) + u(x, t)\frac{\partial}{\partial x}u(x, t) = 0. \quad (1.11)$$

This equation is endowed with the initial condition

$$u(x, 0) = u_0(x) = x, \quad x \in \Omega, \quad (1.12)$$

where u_0 is a function continuous in $(-a, a)$ such that $u_0(\pm a) = \pm a$, and the boundary conditions

$$u(\pm a, t) = \pm a, \quad t > 0. \quad (1.13)$$

The (inviscid) Burgers' equation is an important representant of the class of first-order hyperbolic problems that will be studied in more detail in Section 1.5. In particular, after reading Paragraph 1.5.5 the reader will know that every function $u(x, t)$ that satisfies both equation (1.11) and initial condition (1.12) is constant along the lines

$$x_{x_0}(t) = x_0(t+1), \quad x_0 \in \Omega, \quad (1.14)$$

depicted in Figure 1.2.

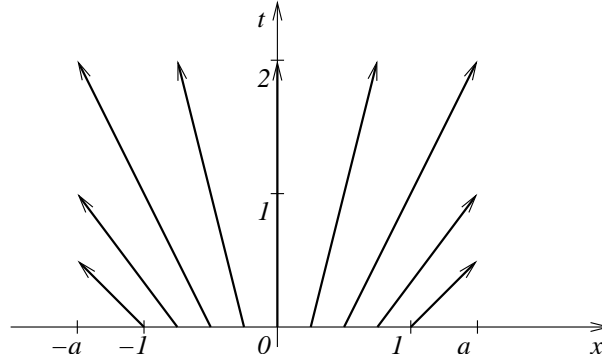


Figure 1.2 Isolines of the solution $u(x, t)$ of Burgers' equation.

It is easy to check the constantness of the solution u along the lines (1.14) by performing the derivative

$$\frac{d}{dt}u(x_{x_0}(t), t).$$

From this fact it follows that the solution to (1.11), (1.12) cannot be constant in time at the endpoints of Ω . Hence the problem (1.11), (1.12), (1.13) has no solution.

Some problems are ill-posed because of their very nature, despite their initial and boundary conditions are defined appropriately. This is illustrated in Example 1.5.

■ **EXAMPLE 1.5 (Ill-posed problem with unstable solution)**

Consider the one-dimensional version of the heat transfer equation (1.9) with normalized coefficients,

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0, \quad (1.15)$$

describing the temperature distribution within a thin slab $\Omega = (0, \pi)$ in the time interval $(0, T)$. We choose an initial temperature distribution $u(x, 0) = u_0(x)$ such that $u_0(0) = u_0(\pi) = 0$, fix the temperature at the endpoints to $u(0) = u(\pi) = 0$ and ask about the solution $u(x, t)$ of (1.15) for $t \in (0, T)$. The initial condition $u_0(x)$ can be expressed by means of the Fourier expansion

$$u_0(x) = \sum_{n=1}^{\infty} c_n \sin(nx). \quad (1.16)$$

Thus it is easy to verify that the exact solution $u(x, t)$ has the form

$$u(x, t) = \sum_{n=1}^{\infty} c_n e^{-n^2 t} \sin(nx) \quad (1.17)$$

and hence that

$$u(x, T) = \sum_{n=1}^{\infty} c_n e^{-n^2 T} \sin(nx) \quad (1.18)$$

is the solution corresponding to the time $t = T$. Notice that the coefficients $c_n e^{-n^2 t}$ converge to zero very fast as the time grows, and therefore after a sufficiently long time T the solution will be very close to zero in Ω . Hence, the heat transfer problem evidently is a well-posed in the sense of Hadamard.

Now let us reverse the time by defining a new temporal variable $s = T - t$. The backward heat transfer equation has the form

$$\frac{\partial \hat{u}}{\partial s} + \frac{\partial^2 \hat{u}}{\partial x^2} = 0.$$

We consider an initial condition $\hat{u}_0(x)$ corresponding to $s = 0$, i.e., to $t = T$. Again, $\hat{u}_0(x)$ can be expressed as

$$\hat{u}_0(x) = \sum_{n=1}^{\infty} d_n \sin(nx), \quad (1.19)$$

and the exact solution $\hat{u}(x, s)$ has the form

$$\hat{u}(x, s) = \sum_{n=1}^{\infty} d_n e^{n^2 s} \sin(nx).$$

Notice that now the coefficients $d_n e^{n^2 s}$ are amplified exponentially as the backward temporal variable s grows. This means that the solution of the backward heat transfer equation does not depend continuously on the initial data $\hat{u}_0(x)$, i.e., that the backward problem is ill-posed.

Suppose that we calculate some numerical approximation of the solution $u(x, T)$ for some sufficiently large time T and then use it as the initial condition $\hat{u}_0(x)$ for the backward problem. What we will observe when solving the backward problem is that the solution $\hat{u}(x, s)$ begins to oscillate immediately and the computation ends with a floating point overflow or similar error very soon. Because of the ill-posedness of the backward problem, chances are slim that one can get close to the original initial condition $u_0(x)$ at $s = T$.

Remark 1.3 (Inverse problems) *The ill-posed backward heat transfer equation from Example 1.5 was an inverse problem. There are various types of ill-posed inverse problems: For example, it is an inverse problem to identify suitable initial state and/or parameters for some physical process to obtain a desired final state. Usually, the better-posed the forward problem, the worse the posedness of the inverse problem.*

1.1.3 General existence and uniqueness results

Prior to discussing various aspects of the elliptic, parabolic, and hyperbolic PDEs in Sections 1.2–1.5, we find it useful to mention a few important abstract existence and uniqueness results for general operator equations. Since this paragraph uses some abstract functional analysis, readers who find its contents too difficult may skip it in the first reading and continue with Section 1.2.

In the following we consider a pair of Hilbert spaces V and W , and an equation of the form

$$Lu = f, \quad (1.20)$$

where $L : D(L) \subset V \rightarrow W$ is a linear operator and $f \in W$. The existence of solution to (1.20) for any right-hand side $f \in W$ is equivalent to the condition $R(L) = W$, while the uniqueness of solution is equivalent to the condition $N(L) = \{0\}$.

Theorem 1.1 (Hahn–Banach) *Let U be a subspace of a (real or complex) normed space V , and $f \in U'$ a linear form over U . Then there exists an extension $g \in V'$ of f such that $g(u) = f(u)$ for all $u \in U$, moreover satisfying $\|g\|_{V'} = \|f\|_{U'}$.*

Proof: The proof can be found in standard functional-analytic textbooks. See, e.g., [34, 65] and [100]. ■

Theorem 1.1 has important consequences: If $v_0 \in V$ and $f(v_0) = 0$ for all $f \in V'$, then $v_0 = 0$. Further, for any $v_0 \in V$ there exists $f \in V'$ such that $\|f\|_{V'} = 1$ and $f(v_0) = \|v_0\|_V$. The following result is used in the proof of the basic existence theorem: For any two disjoint subsets $A, B \subset V$, where A is compact and B convex, there exists $f \in V'$ and $\gamma \in \mathbb{R}$ such that $f(a) < \gamma < f(b)$ for all $a \in A$ and $b \in B$.

Theorem 1.2 (Basic existence result) *Let V, W be Hilbert spaces and $L : D(L) \subset V \rightarrow W$ a bounded linear operator. Then $R(L) = W$ if and only if both $R(L)$ is closed and $R(L)^\perp = \{0\}$.*

Proof: If $R(L) = W$, then obviously $R(L)$ is closed and $R(L)^\perp = \{0\}$. Conversely, assume that $R(L)$ is closed, $R(L)^\perp = \{0\}$ but $R(L) \neq W$. The linearity and boundedness of L implies that $R(L)$ is a closed subspace of W . Let $w \in W \setminus R(L)$. The set $\{w\}$ is compact and the closed set $R(L)$ obviously is convex. By the Hahn–Banach theorem there exists a $w^* \in W'$ such that $(w^*, w) > 0$ and $(w^*, Lv) = 0$ for all $v \in D(L)$. Therefore $0 \neq w^* \in R(L)^\perp$, which is a contradiction. ■

In order to see under what conditions $R(L)$ is closed, let us generalize the notion of continuity by introducing closed operators:

Definition 1.3 (Closed operator) *An operator $T : D(T) \subset V \rightarrow W$, where V and W are Banach spaces, is said to be closed if for any sequence $\{v_n\}_{n=1}^\infty \subset D(T)$, $v_n \rightarrow v$ and $T(v_n) \rightarrow w$ imply that $v \in D(T)$ and $w = Tv$.*

It is an easy exercise to show that every continuous operator is closed. However, there are closed operators which are not continuous:

■ EXAMPLE 1.6 (Closed operator which is not continuous)

Consider the interval $\Omega = (0, 1) \subset \mathbb{R}$, the Hilbert space $V = L^2(\Omega)$ and the Laplace operator $L : V \rightarrow V$, $Lu = -\Delta u = -u''$. This operator is not continuous, since,

e.g., $Lv \notin V$ for $v = x^{-1/3} \in V$. We know that the space $C_0^\infty(\Omega)$ is dense in $L^2(\Omega)$ (see Paragraph A.2.10). To show that L is closed in V , for an element $v \in V$ consider some sequence $\{v_n\}_{n=1}^\infty \subset C_0^\infty(\Omega)$ such that $v_n \rightarrow v$, and such that the sequence $\{-\Delta v_n\}_{n=1}^\infty$ converges to some $w \in V$. Passing to the limit $n \rightarrow \infty$ in the relation

$$\int_{\Omega} -\Delta v_n \varphi \, d\mathbf{x} = - \int_{\Omega} v_n \Delta \varphi \, d\mathbf{x} \quad \text{for all } \varphi \in C_0^\infty(\Omega),$$

we obtain

$$\int_{\Omega} w \varphi \, d\mathbf{x} = - \int_{\Omega} v \Delta \varphi \, d\mathbf{x} \quad \text{for all } \varphi \in C_0^\infty(\Omega).$$

Therefore $w = -\Delta v$ and the operator L is closed.

Theorem 1.3 (Basic existence and uniqueness result) *Let V, W be Hilbert spaces and $L : D(L) \subset V \rightarrow W$ a closed linear operator. Assume that there exists a constant $C > 0$ such that*

$$\|Lv\|_W \geq C\|v\|_V \quad \text{for all } v \in D(L) \quad (1.21)$$

(this inequality sometimes is called the stability or coercivity estimate). If $R(L)^\perp = \{0\}$, then the operator equation $Lu = f$ has a unique solution.

Proof: First let us verify that $R(L)$ is closed. Let $\{w_n\}_{n=1}^\infty \subset R(L)$ such that $w_n \rightarrow w$. Then there is a sequence $\{v_n\}_{n=1}^\infty \subset D(L)$ such that $w_n = Lv_n$. The stability estimate (1.21) implies that $C\|v_n - v_m\|_V \leq \|w_n - w_m\|_W$, which means that $\{v_n\}_{n=1}^\infty$ is a Cauchy sequence in V . Completeness of the Hilbert space V yields existence of a $v \in V$ such that $v_n \rightarrow v$. Since L is closed, we obtain $v \in D(L)$ and $w = Lv \in R(L)$. Theorem 1.2 yields the existence of a solution. The uniqueness of the solution follows immediately from the stability estimate (1.21). ■

Now let us introduce the notion of monotonicity and show that strongly monotone linear operators satisfy the stability estimate (1.21):

Definition 1.4 (Monotonicity) *Let V be a Hilbert space and $L \in \mathcal{L}(V, V')$. The operator L is said to be monotone if*

$$\langle Lv, v \rangle \geq 0 \quad \text{for all } v \in V, \quad (1.22)$$

it is strictly monotone if

$$\langle Lv, v \rangle > 0 \quad \text{for all } 0 \neq v \in V, \quad (1.23)$$

and it is strongly monotone if there exists a constant $C_L > 0$ such that

$$\langle Lv, v \rangle \geq C_L \|v\|^2 \quad \text{for all } v \in V. \quad (1.24)$$

For every $u \in V$ the element $Lu \in V'$ is a linear form. The symbol $\langle Lv, v \rangle$, which means the application of Lu to $v \in V$, is called duality pairing.

The notion of monotonicity for linear operators is a special case of a more general definition applicable to nonlinear operators. An operator $T : V \rightarrow V'$ is said to be monotone if $\langle Tu - Tv, u - v \rangle \geq 0$ for all $u, v \in V$, it is strictly monotone if $\langle Tu - Tv, u - v \rangle > 0$ for all $u, v \in V$, $u \neq v$, and it is strongly monotone if there exists a positive constant C_L such that $\langle Tu - Tv, u - v \rangle \geq C_L \|u - v\|^2$ for all $u, v \in V$. The concept of monotonicity for operators is related to the standard notion of monotonicity of real functions: A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is monotone if the condition $x_1 < x_2$ implies that $f(x_1) \leq f(x_2)$. The same can be written as the condition $(f(x_1) - f(x_2))(x_1 - x_2) \geq 0$ for all $x_1, x_2 \in \mathbb{R}$.

Lemma 1.1 *Let V be a Hilbert space and $L \in \mathcal{L}(V, V')$ a continuous strongly monotone linear operator. Then there exists a constant $C > 0$ such that L satisfies the stability estimate (1.21).*

Proof: The strong monotonicity condition (1.24) implies

$$C_L \|v\|_V^2 \leq \langle Lv, v \rangle \leq \|Lv\|_{V'} \|v\|_V,$$

which means that

$$C_L \|v\|_V \leq \|Lv\|_{V'}$$

■

The following theorem presents an important abstract existence and uniqueness result for operator equations:

Theorem 1.4 (Existence and uniqueness of solution for strongly monotone operators) *Let V be a Hilbert space, $f \in V'$ and $L \in \mathcal{L}(V, V')$ a strongly monotone linear operator. Then for every $f \in V'$ the operator equation $Lu = f$ has a unique solution $u \in V$.*

Proof: According to Lemma 1.1 the operator L satisfies the stability estimate (1.21). Moreover, if $v \in R(L)^\perp$, then $\langle Lv, v \rangle = 0$ and

$$C \|v\|^2 \leq \langle Lv, v \rangle = 0.$$

Hence $R(L)^\perp = \{0\}$, and the conclusion follows from Theorem 1.3. ■

1.1.4 Exercises

Exercise 1.1 *Use Definition 1.3 to show that every continuous operator $L : V \rightarrow W$, where V and W are Banach spaces, is closed.*

Exercise 1.2 *Consider a second-order PDE in the form (1.1) with a nonsymmetric coefficient matrix $A(z)$. Symmetrize the coefficient matrix by defining $\tilde{A} = (A + A^T)/2$. Find out how the remaining coefficients b_i, c_i , and a_0 have to be adjusted so that the equation remains in the form (1.1). Hint: Write $a_{ij} = (a_{ij} + a_{ji})/2 + (a_{ij} - a_{ji})/2$.*

Exercise 1.3 *Consider a second-order PDE in the alternative form (1.3),*

$$-\sum_{i,j=1}^n \tilde{a}_{ij} \frac{\partial^2 u}{\partial z_i \partial z_j} + \sum_{i=1}^n \tilde{b}_i \frac{\partial u}{\partial z_i} + \tilde{a}_0 u = f.$$

where $\tilde{a}_{ij} = \tilde{a}_{ji}$ for all $1 \leq i, j \leq n$.

1. Turn the equation into the conventional form (1.1),

$$-\sum_{i,j=1}^n \frac{\partial}{\partial z_i} \left(a_{ij} \frac{\partial u}{\partial z_j} \right) + \sum_{i=1}^n \left(\frac{\partial}{\partial z_i} (b_i u) + c_i \frac{\partial u}{\partial z_i} \right) + a_0 u = f.$$

2. Write the relations of the coefficients a_{ij} , b_i , c_i , a_0 and \tilde{a}_{ij} , \tilde{b}_i , \tilde{c}_i , \tilde{a}_0 .

Exercise 1.4 Use Definition 1.1 to show that equation (1.8) from Example 1.1 is elliptic.

Exercise 1.5 Use Definition 1.1 to show that equation (1.9) from Example 1.2 is parabolic.

Exercise 1.6 Use Definition 1.1 to show that equation (1.10) from Example 1.3 is hyperbolic.

Exercise 1.7 Verify that the function $u(x, t)$ defined in $(0, \pi)$ by the relation (1.17) is the solution of the heat-transfer equation (1.15) with the boundary conditions $u(0, t) = u(\pi, t) = 0$ for all $t > 0$.

Exercise 1.8 In \mathbb{R}^3 consider the equation

$$\frac{\partial u}{\partial t} - (1 + x_1^2) \frac{\partial^2 u}{\partial x_1^2} - (1 + x_2^4) \frac{\partial^2 u}{\partial x_2^2} - (1 + x_3^6) \frac{\partial^2 u}{\partial x_3^2} + \sqrt{1 + |\mathbf{x}|^2} \frac{\partial u}{\partial x_3} = e^{-|\mathbf{x}|}$$

and decide if (and where in \mathbb{R}^3) it is elliptic, parabolic, or hyperbolic.

Exercise 1.9 In \mathbb{R}^2 consider the equation

$$\frac{\partial^2 u}{\partial t^2} + (1 - x_1^2) \frac{\partial^2 u}{\partial x_1^2} + (1 - x_2^2) \frac{\partial^2 u}{\partial x_2^2} - x_1 x_2 \frac{\partial u}{\partial x_1} = \sin(x_1 \pi) \cos(x_2 \pi).$$

and decide if (and where in \mathbb{R}^2) it is elliptic, parabolic, or hyperbolic.

Exercise 1.10 In \mathbb{R}^3 consider the equation

$$-\Delta u - 2 \frac{\partial^2 u}{\partial x_1 x_2} - 2 \frac{\partial^2 u}{\partial x_2 x_3} = f$$

and decide if (and where in \mathbb{R}^2) it is elliptic, parabolic, or hyperbolic.

Exercise 1.11 In \mathbb{R}^3 consider the equation

$$(1 - x_1^2 - x_2^2) \frac{\partial^2 u}{\partial t^2} - (1 + x_2^2) \frac{\partial^2 u}{\partial x_1^2} - (1 + x_1^4) \frac{\partial^2 u}{\partial x_2^2} + (1 + x_3^2) \frac{\partial u}{\partial x_3} = e^{-|\mathbf{x}|^2}$$

and decide if (and where in \mathbb{R}^3) it is elliptic, parabolic, or hyperbolic.

Exercise 1.12 In \mathbb{R}^2 consider the equation

$$\frac{\partial^2 u}{\partial t^2} - (1 - |\mathbf{x}|^2) \frac{\partial^2 u}{\partial x_1^2} - \left(1 - \frac{|\mathbf{x}|^2}{4} \right) \frac{\partial^2 u}{\partial x_2^2} = 0$$

and decide if (or where in \mathbb{R}^2) it is elliptic, parabolic, or hyperbolic.

1.2 SECOND-ORDER ELLIPTIC PROBLEMS

This section is devoted to the discussion of linear second-order elliptic problems. We begin by deriving the weak formulation of a model problem in Paragraph 1.2.1. Properties of bilinear forms arising in the weak formulation of linear elliptic problems are discussed in Paragraph 1.2.2. In Paragraph 1.2.3 we introduce the Lax–Milgram lemma, which is the basic tool for proving the existence and uniqueness of solution to linear elliptic problems. The weak formulations and solvability analysis of problems involving various types of boundary conditions are discussed in Paragraphs 1.2.5–1.2.8. Abstract energy of elliptic problems, which plays an important role in their numerical solution (error estimation, automatic adaptivity), is introduced in Paragraph 1.2.9. Finally, Paragraph 1.2.10 presents maximum principles for elliptic problems, which are used to prove their well-posedness.

1.2.1 Weak formulation of a model problem

Assume an open bounded set $\Omega \subset \mathbb{R}^d$ with Lipschitz-continuous boundary, and recall the general linear second-order equation (1.1),

$$-\sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial u}{\partial x_j} \right) + \sum_{i=1}^n \left(\frac{\partial}{\partial x_i} (b_i u) + c_i \frac{\partial u}{\partial x_i} \right) + a_0 u = f, \quad (1.25)$$

where the coefficients and the right-hand side satisfy the regularity assumptions formulated in Paragraph 1.1.1. In this case we put $n = d$. Equation (1.25) is elliptic if the symmetric coefficient matrix $A = \{a_{ij}\}_{i,j=1}^d$ is positive definite everywhere in Ω (Definition 1.1).

Consider the model equation

$$-\nabla \cdot (a_1 \nabla u) + a_0 u = f \quad \text{in } \Omega, \quad (1.26)$$

obtained from (1.25) by assuming $a_{ij}(\mathbf{x}) = a_1(\mathbf{x})\delta_{ij}$ and $\mathbf{b}(\mathbf{x}) = \mathbf{c}(\mathbf{x}) = 0$ in Ω . For the existence and uniqueness of solution we add another important assumption:

$$a_1(\mathbf{x}) \geq C_{min} > 0 \quad \text{and} \quad a_0(\mathbf{x}) \geq 0 \quad \text{in } \Omega. \quad (1.27)$$

The problem (1.26) is fairly general: Even with $a_0 \equiv 0$ it describes, for example, the following physical processes:

1. Stationary heat transfer (u is the temperature, a_1 is the thermal conductivity, and f are the heat sources),
2. electrostatics (u is the electrostatic potential, a_1 is the dielectric constant, and f is the charge density),
3. transverse deflection of a cable (u is the transverse deflection, a_1 is the axial tension, and f is the transversal load),
4. axial deformation of a bar (u is the axial displacement, $a_1 = EA$ is the product of the elasticity modulus and the cross-sectional area, and f is either the friction or contact force on the surface of the bar),
5. pipe flow (u is the hydrostatic pressure, $a_1 = \pi D^4/128\mu$, D is the diameter, μ is the viscosity and $f = 0$ represents zero flow sources),

6. laminar incompressible flow through a channel under constant pressure gradient (u is the velocity, a_1 is the viscosity, and f is the pressure gradient),
7. porous media flow (u is the fluid head, a_1 is the permeability coefficient, and f is the fluid flux).

To begin with, let (1.26) be endowed with homogeneous Dirichlet boundary conditions

$$u(\boldsymbol{x}) = 0 \quad \text{on } \partial\Omega. \quad (1.28)$$

This type of boundary conditions carries the name of a French mathematician Johann Peter Gustav Lejeune Dirichlet, who made substantial contributions to the solution of Fermat's Last Theorem, theory of polynomial functions, analytic and algebraic number theory, convergence of trigonometric series, and boundary-value problems for harmonic¹ functions.



Figure 1.3 Johann Peter Gustav Lejeune Dirichlet (1805–1859).

Classical solution to the problem (1.26), (1.28) is a function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ satisfying the equation (1.26) everywhere in Ω and fulfilling the boundary condition (1.28) at every $\boldsymbol{x} \in \partial\Omega$. Naturally, one has to assume that $f \in C(\Omega)$. However, neither this nor even stronger requirement $f \in C(\overline{\Omega})$ guarantees the solvability of the problem, for which still stronger smoothness of f is required.

Weak formulation In order to reduce the above-mentioned regularity restrictions, we introduce the weak formulation of the problem (1.26), (1.28). The derivation of the weak formulation of (1.26) consists of the following four standard steps:

1. Multiply (1.26) with a test function $v \in C_0^\infty(\Omega)$,

$$-\nabla \cdot (a_1 \nabla u)v + a_0 uv = fv.$$

2. Integrate over Ω ,

$${}^1\Delta u = 0$$

$$-\int_{\Omega} \nabla \cdot (a_1 \nabla u) v \, d\mathbf{x} + \int_{\Omega} a_0 uv \, d\mathbf{x} = \int_{\Omega} f v \, d\mathbf{x}.$$

3. Use the Green's formula (A.80) to reduce the maximum order of the partial derivatives present in the equation. The fact that v vanishes on the boundary $\partial\Omega$ removes the boundary term, and we have

$$\int_{\Omega} a_1 \nabla u \cdot \nabla v \, d\mathbf{x} + \int_{\Omega} a_0 uv \, d\mathbf{x} = \int_{\Omega} f v \, d\mathbf{x}. \quad (1.29)$$

4. Find the largest possible function spaces for u, v , and other functions in (1.29) where all integrals are finite. Originally, identity (1.29) was derived under very strong regularity assumptions $u \in C^2(\Omega) \cap C(\bar{\Omega})$ and $v \in C_0^\infty(\Omega)$. All integrals in (1.29) remain finite when these assumptions are weakened to

$$u, v \in H_0^1(\Omega), \quad f \in L^2(\Omega), \quad (1.30)$$

where $H_0^1(\Omega)$ is the Sobolev space $W_0^{1,2}(\Omega)$ defined in Section A.4. Similarly the regularity assumptions for the coefficients a_1 and a_0 can be reduced to

$$a_1, a_0 \in L^\infty(\Omega). \quad (1.31)$$

The weak form of the problem (1.26), (1.28) is stated as follows: Given $f \in L^2(\Omega)$, find a function $u \in H_0^1(\Omega)$ such that

$$\int_{\Omega} a_1 \nabla u \cdot \nabla v + a_0 uv \, d\mathbf{x} = \int_{\Omega} f v \, d\mathbf{x} \quad \text{for all } v \in H_0^1(\Omega). \quad (1.32)$$

The existence and uniqueness of solution will be discussed in Paragraph 1.2.4.

Let us mention that the assumption $f \in L^2(\Omega)$ can be further weakened to $f \in H^{-1}(\Omega)$, where $H^{-1}(\Omega)$, which is the dual space to $H_0^1(\Omega)$, is larger than $L^2(\Omega)$. Then the integral

$$\int_{\Omega} f v \, d\mathbf{x}$$

is interpreted as the duality pairing $\langle f, v \rangle$ between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

Equivalence of the strong and weak solutions Obviously the classical solution to the problem (1.26), (1.28) also solves the weak formulation (1.32). Conversely, if the weak solution of (1.32) is sufficiently regular, which in this case means $u \in C^2(\Omega) \cap C(\bar{\Omega})$, it also satisfies the classical formulation (1.26), (1.28).

In the language of linear forms Let $V = H_0^1(\Omega)$. We define a bilinear form $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$,

$$a(u, v) = \int_{\Omega} (a_1 \nabla u \cdot \nabla v + a_0 uv) \, d\mathbf{x},$$

and a linear form $l \in V'$,

$$l(v) = \langle l, v \rangle = \int_{\Omega} f v \, dx.$$

Then the weak formulation of the problem (1.26), (1.28) reads: Find a function $u \in V$ such that

$$a(u, v) = l(v) \quad \text{for all } v \in V. \quad (1.33)$$

This notation is common in the study of partial differential equations and finite element methods.

1.2.2 Bilinear forms, energy norm, and energetic inner product

In this paragraph we learn more about bilinear forms for elliptic problems, and introduce the notions of energy norm and energetic inner product. Every bilinear form $a : V \times V \rightarrow \mathbb{R}$ in a Banach space V is associated with a unique linear operator $A : V \rightarrow V'$ defined by

$$(Au)(v) = \langle Au, v \rangle = a(u, v) \quad \text{for all } u, v \in V. \quad (1.34)$$

Lemma 1.2 *Relation (1.34) defines a one-to-one correspondence between continuous bilinear forms $a : V \times V \rightarrow \mathbb{R}$ and linear continuous operators $A : V \rightarrow V'$.*

Proof: If $A \in \mathcal{L}(V, V')$, then the mapping $a : V \times V \rightarrow \mathbb{R}$ defined by (1.34) is bilinear and bounded,

$$|a(u, v)| \leq \|Au\|_{V'} \|v\|_V \leq \|A\| \|u\|_V \|v\|_V \quad \text{for all } u, v \in V.$$

Conversely, let $a(\cdot, \cdot)$ be a continuous bilinear form on $V \times V$. For any $u \in V$ the map $v \rightarrow a(u, v)$ defines a continuous linear operator on V . Hence there exists an element $Au \in V'$ such that (1.34) holds. The bilinearity and boundedness of $a(\cdot, \cdot)$ implies the linearity and boundedness of A . ■

Basic properties of bilinear forms in Hilbert spaces are introduced in Definition 1.5 and discussed in Lemma 1.3:

Definition 1.5 *Let V be a real Hilbert space, $a : V \times V \rightarrow \mathbb{R}$ a bilinear form and $A : V \rightarrow V'$ a linear operator related to $a(\cdot, \cdot)$ via (1.34). We say that*

1. a is bounded if there exists a constant $C_a > 0$ such that $|a(u, v)| \leq C_a \|u\| \|v\|$ for all $u, v \in V$,
2. a is positive if $a(v, v) \geq 0$ for all $v \in V$,
3. a is strictly positive if $a(v, v) > 0$ for all $0 \neq v \in V$,
4. a is V -elliptic (coercive) if there exists a constant $\tilde{C}_a > 0$ such that $a(v, v) \geq \tilde{C}_a \|v\|_V^2$ for all $v \in V$,
5. a is symmetric if $a(u, v) = a(v, u)$ for all $u, v \in V$.

Lemma 1.3 *Under the assumptions of Definition 1.5 it holds:*

1. *The bilinear form a is bounded if and only if the linear operator A is bounded.*
2. *The bilinear form a is positive if and only if the linear operator A is monotone.*
3. *The bilinear form a is strictly positive if and only if the linear operator A is strictly monotone.*
4. *The bilinear form a is V -elliptic if and only if the linear operator A is strongly monotone.*
5. *The bilinear form a is symmetric if and only if the linear operator A is symmetric (i.e., if $\langle Au, v \rangle = \langle Av, u \rangle$ for all $u, v \in V$).*

Proof: Left to the reader as an exercise. ■

Definition 1.6 (Energetic inner product, energy norm) *Let V be a Hilbert space and $a : V \times V \rightarrow \mathbb{R}$ a bounded symmetric V -elliptic bilinear form. The bilinear form defines an inner product*

$$(u, v)_e = a(u, v) \quad (1.35)$$

in V , called energetic inner product. The norm induced by the energetic inner product,

$$\|u\|_e = \sqrt{(u, u)_e}, \quad (1.36)$$

is called energy norm.

It is easy to verify that $\|\cdot\|_e$ and $(\cdot, \cdot)_e$ fulfill all properties of norm and inner product (use Definitions A.24 and A.41).

Lemma 1.4 *Let V be a Hilbert space and $a : V \times V \rightarrow \mathbb{R}$ a bounded symmetric V -elliptic bilinear form. The energy norm induced by a is equivalent to the original norm in V ,*

$$C_1 \|u\|_V \leq \|u\|_e \leq C_2 \|u\|_V \quad \text{for all } u \in V, \quad (1.37)$$

where $C_1, C_2 > 0$ are some real constants.

Proof: Left to the reader as an exercise. ■

If the V -elliptic bilinear form $a(\cdot, \cdot)$ is not symmetric, it does not represent an inner product, but still it induces an energy norm. If $a : V \times V \rightarrow \mathbb{C}$, then the symmetry requirement $a(u, v) = a(v, u)$ is replaced with the sesquilinearity requirement $a(u, v) = \overline{a(v, u)}$.

Both the energetic inner product $(\cdot, \cdot)_e$ and the energy norm $\|\cdot\|_e$ represent important tools in the error analysis and numerical solution of elliptic PDEs. They are used to derive both a-priori and a-posteriori error estimates, to guide refinement strategies for adaptive finite element methods, and for other purposes. We will return to this topic later, after introducing the finite element discretization in Chapter 2.

1.2.3 The Lax–Milgram lemma

The Lax–Milgram lemma is the basic and most important tool for proving the existence and uniqueness of solution to elliptic problems.

Theorem 1.5 (Lax–Milgram lemma) *Let V be a Hilbert space, $a : V \times V \rightarrow \mathbb{R}$ a bounded V -elliptic bilinear form and $l \in V'$. Then there exists a unique solution to the problem*

$$a(u, v) = l(v) \quad \text{for all } v \in V. \quad (1.38)$$

Remark 1.4 (Lax–Milgram vs. Riesz) *If the bilinear form $a(\cdot, \cdot)$ is symmetric, then the unique solution $u \in V$ of equation (1.38) is nothing else than the unique representant of the linear form $l \in V'$ with respect to the energetic inner product $(\cdot, \cdot)_e = a(\cdot, \cdot)$. In this sense the Lax–Milgram lemma is a special case of the Riesz representation theorem (Theorem A.15).*

Proof: The uniqueness of solution follows immediately from the V -ellipticity of the bilinear form a . We will use Theorem 1.2 to verify the existence. Let $A : V \rightarrow V'$ be the linear operator associated with the bilinear form a via (1.34). Then A is bounded and strongly monotone. By $L = \mathcal{J}A : V \rightarrow V$ denote the isometric dual mapping from the Riesz theorem,

$$a(u, v) = \langle Au, v \rangle = (\mathcal{J}Au, v) \quad \text{for all } u, v \in V.$$

Recall that $R(L) = V$ if and only if $R(L)$ is closed and $R(L)^\perp = \{0\}$. To show that $R(L)$ is closed, let $\{u_n\}_{n=1}^\infty \subset R(L)$ be a sequence converging to some function u . Then $u_n = \mathcal{J}Aw_n$ where $\{w_n\}_{n=1}^\infty \subset V$. Lemma 1.1 yields the existence of a constant $C > 0$ such that

$$\|u_n - u_m\| = \|\mathcal{J}A(w_n - w_m)\| = \|A(w_n - w_m)\| \geq C\|w_n - w_m\|.$$

Hence $\{w_n\}_{n=1}^\infty$ is a Cauchy sequence that has a limit $w \in V$. It holds

$$\|u_n - \mathcal{J}Aw\| = \|\mathcal{J}A(w_n - w)\| = \|A(w_n - w)\| \leq C_a\|w_n - w\| \rightarrow 0.$$

Therefore $u = \mathcal{J}Aw \in R(L)$ and $R(L)$ is closed. To prove that $R(L)^\perp = \{0\}$, take an arbitrary $u \in R(L)^\perp$. Then for any $v \in V$ it is

$$0 = (\mathcal{J}Av, u) = a(v, u).$$

Putting $v = u$, we obtain that the energy norm $\|u\|_e = 0$ and thus that $u = 0$. ■

1.2.4 Unique solvability of the model problem

The existence and uniqueness of solution to the model problem (1.33) can be proved using the Lax–Milgram lemma (Theorem 1.5) under the following assumptions:

Lemma 1.5 Assume that $a_1(\mathbf{x}) \geq C_{min} > 0$ and $a_0(\mathbf{x}) \geq 0$ a.e. in Ω . Then the weak problem (1.33) has a unique solution $u \in V$.

Proof: Since $a_1, a_0 \in L^\infty(\Omega)$, there exists a $C_{max} < \infty$ such that $|a_1(\mathbf{x})| \leq C_{max}$ and $|a_0(\mathbf{x})| \leq C_{max}$ a.e. in Ω . Then,

$$|a(u, v)| \leq \int_{\Omega} (a_1 |\nabla u \cdot \nabla v| + a_0 |uv|) \, d\mathbf{x} \leq C_{max} \int_{\Omega} (|\nabla u \cdot \nabla v| + |uv|) \, d\mathbf{x}. \quad (1.39)$$

Since $\nabla u, \nabla v \in [L^2(\Omega)]^d$, the Hölder inequality (A.50) yields

$$\int_{\Omega} |\nabla u \cdot \nabla v| \, d\mathbf{x} \leq \left(\int_{\Omega} |\nabla u|^2 \, d\mathbf{x} \right)^{\frac{1}{2}} \left(\int_{\Omega} |\nabla v|^2 \, d\mathbf{x} \right)^{\frac{1}{2}} = |u|_{1,2} |v|_{1,2}. \quad (1.40)$$

Analogously, for the product $|uv|$ one obtains

$$\int_{\Omega} |uv| \, d\mathbf{x} \leq \left(\int_{\Omega} u^2 \, d\mathbf{x} \right)^{\frac{1}{2}} \left(\int_{\Omega} v^2 \, d\mathbf{x} \right)^{\frac{1}{2}} = \|u\|_{L^2} \|v\|_{L^2}. \quad (1.41)$$

The norm $\|\cdot\|_{1,2}$ is obtained by adding a nonnegative term to the seminorm $|\cdot|_{1,2}$,

$$|u|_{1,2} |v|_{1,2} \leq \|u\|_{1,2} \|v\|_{1,2}. \quad (1.42)$$

Similarly for the L^2 -norm,

$$\|u\|_{L^2} \|v\|_{L^2} \leq \|u\|_{1,2} \|v\|_{1,2}. \quad (1.43)$$

Finally, relations (1.39) to (1.43) together yield

$$|a(u, v)| \leq 2C_{max} \|u\|_{1,2} \|v\|_{1,2},$$

which means that the bilinear form is bounded with the constant $C_a = 2C_{max}$. Next let us prove the V -ellipticity of $a(\cdot, \cdot)$. Using the Poincaré–Friedrichs’ inequality (Theorem A.26) in the space $V = H_0^1(\Omega)$, together with the nonnegativity of a_0 and strict positivity of a_1 , we obtain that there exists a constant $C_{pf} > 0$ such that

$$\begin{aligned} a(v, v) &= \int_{\Omega} a_1 |\nabla v|^2 + a_0 v^2 \, d\mathbf{x} \geq \int_{\Omega} a_1 |\nabla v|^2 \, d\mathbf{x} \\ &\geq C_{min} \int_{\Omega} |\nabla v|^2 \, d\mathbf{x} = C_{min} |v|_{1,2}^2 \geq C_{min} C_{pf}^2 \|v\|_{1,2}^2 \quad \text{for all } v \in V. \end{aligned}$$

Thus the bilinear form $a(\cdot, \cdot)$ is bounded and V -elliptic, and the Lax–Milgram lemma yields the existence and uniqueness of solution for every $f \in L^2(\Omega)$. ■

Discussion of the existence and uniqueness of solution for elliptic operators of the general form (1.25) can be found, e.g., in [93].

1.2.5 Nonhomogeneous Dirichlet boundary conditions

In this paragraph we consider the model equation (1.26) endowed with more general Dirichlet boundary conditions of the form

$$u(\mathbf{x}) = g(\mathbf{x}) \quad \text{on } \partial\Omega, \quad (1.44)$$

where $g \in C(\partial\Omega)$. For the purpose of the weak formulation we consider a function $G \in C^2(\Omega) \cap C(\bar{\Omega})$ such that $G = g$ on $\partial\Omega$ (the so-called Dirichlet lift of g). Notice that G is not unique, but we will show later that the solution is invariant under its choice. Writing $u = G + U$, the problem (1.26), (1.44) can be reformulated to:

Find $U \in C_0^2(\Omega)$ such that

$$\begin{aligned} -\nabla \cdot [a_1 \nabla(U + G)] + a_0(U + G) &= f \quad \text{in } \Omega, \\ U + G &= g \quad \text{on } \partial\Omega, \end{aligned}$$

or, equivalently,

$$-\nabla \cdot (a_1 \nabla U) + a_0 U = f + \nabla \cdot (a_1 \nabla G) - a_0 G \quad \text{in } \Omega, \quad (1.45)$$

$$U = 0 \quad \text{on } \partial\Omega, \quad (1.46)$$

Except for an adjusted right-hand side, this problem is identical to the model problem (1.26), (1.28). We proceed analogously as in Paragraph 1.2.1 to derive its weak formulation:

Find $U \in V = H_0^1(\Omega)$ such that

$$a(U, v) = l(v) \quad \text{for all } v \in V \quad (1.47)$$

with

$$\begin{aligned} a(U, v) &= \int_{\Omega} (a_1 \nabla U \cdot \nabla v + a_0 U v) \, d\mathbf{x}, \quad v \in V, \\ l(v) &= \int_{\Omega} (f v - a_1 \nabla G \cdot \nabla v - a_0 G v) \, d\mathbf{x}, \quad v \in V, \end{aligned}$$

This weak formulation is defined under much weaker assumptions on f , g , and G . In particular, we can assume that $f \in L^2(\Omega)$ and $G \in H^1(\Omega)$ with the trace $g \in H^{\frac{1}{2}}(\partial\Omega)$.

We have seen in Paragraph 1.2.4 that the bilinear form $a(\cdot, \cdot)$ is bounded and V -elliptic. In other words, the Lax–Milgram lemma yields the existence and uniqueness of solution to (1.47) for every Dirichlet lift G .

Independence of the solution $u = U + G$ on the Dirichlet lift G : Assume that $U_1 + G_1 = u_1 \in H^1(\Omega)$ and $U_2 + G_2 = u_2 \in H^1(\Omega)$ are two weak solutions. By (1.47) the difference $u_1 - u_2 \in V = H_0^1(\Omega)$ satisfies

$$a(u_1 - u_2, v) = 0 \quad \text{for all } v \in V.$$

Taking $u_1 - u_2$ for v and using the V -ellipticity of the bilinear form a , we obtain

$$0 = a(u_1 - u_2, u_1 - u_2) \geq C_{el} \|u_1 - u_2\|_V^2.$$

This means that

$$\|u_1 - u_2\|_V = 0,$$

i.e., that $u_1 = u_2$ a.e. in Ω .

1.2.6 Neumann boundary conditions

Consider the model equation (1.26) with Neumann boundary conditions of the form

$$\frac{\partial u}{\partial \boldsymbol{\nu}} = g \quad \text{on } \partial\Omega, \quad (1.48)$$

where $g \in C(\partial\Omega)$. This time we have to strengthen the positivity assumption on the coefficient a_0 to

$$a_0(\mathbf{x}) \geq \hat{C}_{min} > 0 \quad \text{in } \Omega. \quad (1.49)$$

The weak formulation of the problem (1.26), (1.48) is derived as follows: Assume that $u \in C^\infty(\Omega) \cap C^1(\bar{\Omega})$. Multiply (1.26) with a test function $v \in C^\infty(\Omega) \cap C^1(\bar{\Omega})$, integrate over Ω , and use the Green's theorem to reduce the maximum order of the partial derivatives. The boundary integrals do not vanish as they did in the homogeneous Dirichlet case, and we get an extra boundary term,

$$\int_{\Omega} (a_1 \nabla u \cdot \nabla v + a_0 uv) \, d\mathbf{x} - \int_{\partial\Omega} a_1 \frac{\partial u}{\partial \boldsymbol{\nu}} v \, d\mathbf{S} = \int_{\Omega} f v \, d\mathbf{x}.$$

Here $\boldsymbol{\nu}$ is the unit outer normal vector to $\partial\Omega$ and $\partial u / \partial \boldsymbol{\nu} = \nabla u \cdot \boldsymbol{\nu}$. Substituting the boundary condition (1.48) into the boundary integral, and weakening the regularity assumptions, we obtain the following weak formulation:

Given $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$, find $u \in V = H^1(\Omega)$ such that

$$\int_{\Omega} (a_1 \nabla u \cdot \nabla v + a_0 uv) \, d\mathbf{x} = \int_{\Omega} f v \, d\mathbf{x} + \int_{\partial\Omega} a_1 g v \, d\mathbf{S} \quad \text{for all } v \in V.$$

Stated in the language of linear forms, one has to find a function $u \in V$ such that

$$a(u, v) = l(v) \quad \text{for all } v \in V, \quad (1.50)$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} a_1 \nabla u \cdot \nabla v + a_0 uv \, d\mathbf{x} \quad \text{for all } u, v \in V, \\ l(v) &= \int_{\Omega} f v \, d\mathbf{x} + \int_{\partial\Omega} a_1 g v \, d\mathbf{S} \quad \text{for all } v \in V. \end{aligned}$$

Notice that although the bilinear form $a(\cdot, \cdot)$ is given by the same formula as in the case of Dirichlet boundary conditions, it is different since the space V changed.

The boundedness of the bilinear form $a(\cdot, \cdot)$ in $V \times V$ can be shown analogously to the proof of Lemma 1.5. Notice, however, that one cannot use the Poincaré–Friedrichs' inequality to prove the V -ellipticity of $a(\cdot, \cdot)$, since now the solution is not zero on the boundary. Here the additional assumption (1.49) comes into the play, and we obtain

$$a(v, v) \geq \min(C_{min}, \hat{C}_{min}) \|v\|_V^2.$$

The Lax–Milgram lemma guarantees that the problem (1.50) has a unique solution $u \in V$.

Remark 1.5 (Neumann problem without the assumption (1.49)) *The assumption (1.49) guarantees the presence of a nonzero L^2 -term in the bilinear form. Without this term, neither the classical nor the weak formulation has a unique solution in Sobolev spaces. For example, if u is a solution of $-\Delta u = f$ with Neumann boundary conditions, then also $u + C$, where C is an arbitrary constant, is a solution. Let us formulate this problem in the weak sense:*

Find $u \in H^1(\Omega)$ such that

$$\int_{\Omega} \nabla u \cdot \nabla v \, d\mathbf{x} = \int_{\Omega} f v \, d\mathbf{x} + \int_{\partial\Omega} g v \, d\mathbf{S} \quad \text{for all } v \in H^1(\Omega). \quad (1.51)$$

Using the test function $v = 1 \in H^1(\Omega)$, one finds that a necessary condition for (1.51) to have a solution at all is

$$\int_{\Omega} f \, d\mathbf{x} + \int_{\partial\Omega} g \, d\mathbf{S} = 0. \quad (1.52)$$

It follows from a deeper analysis in the quotient space $H^1(\Omega)/\mathbb{R}$ that condition (1.52) is sufficient for the existence and uniqueness of solution in $H^1(\Omega)/\mathbb{R}$ (see, e.g., [6]).

Remark 1.6 (Essential and natural boundary conditions) *Dirichlet boundary conditions are sometimes called essential since they essentially influence the weak formulation: They determine the function space in which the solution is sought. On the other hand, Neumann boundary conditions do not influence the function space and can be naturally incorporated into the boundary integrals. Therefore they are called natural.*

1.2.7 Newton (Robin) boundary conditions

Another frequently used type of natural boundary conditions involves a combination of function values and normal derivatives. Consider the model equation (1.26) equipped with such boundary conditions,

$$-\nabla \cdot (a_1 \nabla u) + a_0 u = f \quad \text{in } \Omega, \quad (1.53)$$

$$c_1 u + c_2 \frac{\partial u}{\partial \nu} = g \quad \text{on } \partial\Omega, \quad (1.54)$$

where $f \in C(\Omega)$, $g \in C(\partial\Omega)$, and $c_1, c_2 \in C(\partial\Omega)$ are such that $c_1 c_2 > 0$ and $0 < \epsilon \leq |c_2|$ on $\partial\Omega$. The positivity assumptions (1.27) and (1.49) on the coefficients a_0, a_1 apply.

For a sufficiently regular function $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$, the weak identity

$$\int_{\Omega} a_1 \nabla u \cdot \nabla v + a_0 u v \, d\mathbf{x} - \int_{\partial\Omega} a_1 \frac{\partial u}{\partial \nu} v \, d\mathbf{S} = \int_{\Omega} f v \, d\mathbf{x}$$

is derived analogously to the Neumann case. Using the boundary condition (1.54), we obtain the following weak formulation:

Given $f \in L^2(\Omega)$, $g \in L^2(\partial\Omega)$, and $a_0, a_1 \in L^\infty(\Omega)$, find $u \in V = H^1(\Omega)$ such that

$$\int_{\Omega} a_1 \nabla u \cdot \nabla v + a_0 u v \, d\mathbf{x} + \int_{\partial\Omega} \frac{a_1 c_1}{c_2} u v \, d\mathbf{S} = \int_{\Omega} f v \, d\mathbf{x} + \int_{\partial\Omega} \frac{a_1 g}{c_2} v \, d\mathbf{S} \quad \text{for all } v \in V.$$

In other words, it is our task to find $u \in V$ such that

$$a(u, v) = l(v) \quad \text{for all } v \in V, \quad (1.55)$$

where

$$\begin{aligned} a(u, v) &= \int_{\Omega} a_1 \nabla u \cdot \nabla v + a_0 uv \, d\mathbf{x} + \int_{\partial\Omega} \frac{a_1 c_1}{c_2} uv \, d\mathbf{S} \quad \text{for all } u, v \in V, \\ l(v) &= \int_{\Omega} f v \, d\mathbf{x} + \int_{\partial\Omega} \frac{a_1 g}{c_2} v \, d\mathbf{S} \quad \text{for all } v \in V. \end{aligned}$$

Since the bilinear form $a(\cdot, \cdot)$ is both bounded and V -elliptic (use Theorem A.28), the Lax–Milgram lemma implies that problem (1.55) has a unique solution $u \in V$.

1.2.8 Combining essential and natural boundary conditions

What remains to be discussed is the combination of essential and natural boundary conditions. Let us choose, for example, the Dirichlet and Neumann conditions for this purpose. Hence, let the boundary $\partial\Omega$ be split into two nonempty disjoint open parts Γ_D and Γ_N , and consider the problem

$$-\nabla \cdot (a_1 \nabla u) + a_0 u = f \quad \text{in } \Omega, \quad (1.56)$$

$$u = g_D \quad \text{on } \Gamma_D, \quad (1.57)$$

$$\frac{\partial u}{\partial \nu} = g_N \quad \text{on } \Gamma_N. \quad (1.58)$$

The weak formulation is derived as follows: First extend the function $g_D \in C(\Gamma_D)$ to the rest of the boundary $\partial\Omega$ by introducing a function $\tilde{g}_D \in C(\partial\Omega)$ such that $\tilde{g}_D \equiv g_D$ on Γ_D . The nonuniqueness of this extension is not going to cause any problems. Next find some Dirichlet lift $G \in C^2(\Omega) \cap C(\bar{\Omega})$ of \tilde{g}_D (i.e., $G \equiv \tilde{g}_D$ on $\partial\Omega$). The solution u is sought in the form $u = U + G$ analogously to the pure Dirichlet case. The equations

$$-\nabla \cdot [a_1 \nabla (U + G)] + a_0 (U + G) = f \quad \text{in } \Omega, \quad (1.59)$$

$$(U + G) = g_D \quad \text{on } \Gamma_D, \quad (1.60)$$

$$\frac{\partial (U + G)}{\partial \nu} = g_N \quad \text{on } \Gamma_N, \quad (1.61)$$

yield

$$-\nabla \cdot (a_1 \nabla U) + a_0 U = f + \nabla \cdot (a_1 \nabla G) - a_0 G \quad \text{in } \Omega, \quad (1.62)$$

$$U = 0 \quad \text{on } \Gamma_D, \quad (1.63)$$

$$\frac{\partial (U + G)}{\partial \nu} = g_N \quad \text{on } \Gamma_N. \quad (1.64)$$

The appropriate space for the function U is

$$V = \{u \in H^1(\Omega); u = 0 \text{ on } \Gamma_D\}. \quad (1.65)$$

Applying the standard procedure that we went through several times, we arrive at the weak identity

$$\begin{aligned} & \int_{\Omega} (a_1 \nabla U \cdot \nabla v + a_0 U v) \, d\mathbf{x} \\ &= \int_{\Omega} (f v - a_1 \nabla G \cdot \nabla v - a_0 G v) \, d\mathbf{x} + \int_{\Gamma_N} \left(a_1 \frac{\partial(U+G)}{\partial \nu} v \right) \, d\mathbf{S} \quad \text{for all } v \in V. \end{aligned}$$

Using the Neumann boundary condition (1.64) on Γ_N , we finally obtain the following weak problem:

Find a function U in the space V such that

$$a(U, v) = l(v) \quad \text{for all } v \in V, \quad (1.66)$$

where

$$\begin{aligned} a(U, v) &= \int_{\Omega} (a_1 \nabla U \cdot \nabla v + a_0 U v) \, d\mathbf{x}, \quad U, v \in V, \\ l(v) &= \int_{\Omega} (f v - a_1 \nabla G \cdot \nabla v - a_0 G v) \, d\mathbf{x} + \int_{\Gamma_N} a_1 g_N v \, d\mathbf{S} \quad \text{for all } v \in V. \end{aligned} \quad (1.67)$$

The bilinear form $a(\cdot, \cdot)$ is bounded and V -elliptic (the proof is analogous to Paragraph 1.2.4). The Poincaré–Friedrichs' inequality holds in V due to the zero boundary condition for U on Γ_D (see Remark A.8). Therefore the Lax–Milgram lemma implies that problem (1.66) has a unique solution $U \in V$. As usual, the final solution satisfying both the essential and natural boundary conditions is $u = U + G$.

1.2.9 Energy of elliptic problems

It was mentioned in Paragraph 1.1.1 that elliptic problems usually describe some equilibrium or minimum-energy state of a system. In this paragraph we introduce the explicit form of the abstract energy, at least for symmetric problems. The most important numerical scheme based on the minimization of the abstract energy, the Ritz method, will be discussed later in Chapter 2.

Theorem 1.6 *Let V be a linear space, $a : V \times V \rightarrow \mathbb{R}$ a symmetric V -elliptic bilinear form and $l \in V'$. Then the functional of abstract energy,*

$$E(v) = \frac{1}{2} a(v, v) - l(v), \quad (1.68)$$

attains its minimum in V at an element $u \in V$ if and only if

$$a(u, v) = l(v) \quad \text{for all } v \in V. \quad (1.69)$$

Moreover, the minimizer $u \in V$ is unique.

Proof: Let (1.69) hold. Then

$$\begin{aligned} E(u + tv) &= \frac{1}{2} a(u + tv, u + tv) - l(u + tv) \\ &= E(u) + t(a(u, v) - l(v)) + \frac{1}{2} t^2 a(v, v) \end{aligned} \quad (1.70)$$

for all $u, v \in V$ and $t \in \mathbb{R}$. If $u \in V$ satisfies (1.69), then the last equation with $t = 1$ implies

$$E(u + v) = E(u) + \frac{1}{2}a(v, v) > E(u) \quad \text{for all } 0 \neq v \in V.$$

Thus $u \in V$ is a unique minimizer of (1.68).

Conversely, if E has a minimum at $u \in V$, then for every $v \in V$ the derivative of the quadratic function $\phi(t) = E(u + tv)$ must vanish at $t = 0$. By (1.70),

$$0 = \phi'(0) = a(u, v) - l(v),$$

and (1.69) holds. ■

Another interesting theoretical application of the energy-minimization concept is an alternative proof of the Lax–Milgram lemma for symmetric elliptic problems in convex sets:

Theorem 1.7 (Lax–Milgram lemma for convex sets) *Let W be a closed convex set in a Hilbert space V and $a : V \times V \rightarrow \mathbb{R}$ a bounded V -elliptic bilinear form. Then for every $l \in V'$ there exists a unique $u \in W$ such that $E(u) = \inf\{E(v); v \in W\}$, where*

$$E(v) = \frac{1}{2}a(v, v) - l(v).$$

Proof: The functional E is bounded from below since

$$E(v) \geq \frac{1}{2}C_a\|v\|^2 - \|l\|\|v\| = \frac{1}{2C_a}(C_a\|v\| - \|l\|)^2 - \frac{\|l\|^2}{2C_a} \geq -\frac{\|l\|^2}{2C_a}.$$

Let $e_0 = \inf\{E(v); v \in W\}$ and let $\{v_n\}_{n=1}^\infty$ be a minimizing sequence, i.e.,

$$\lim_{n \rightarrow \infty} E(v_n) = e_0.$$

Then

$$\begin{aligned} C_a\|v_n - v_m\|^2 &\leq a(v_n - v_m, v_n - v_m) \\ &= 2a(v_n, v_n) + 2a(v_m, v_m) - a(v_n + v_m, v_n + v_m) \\ &= 4E(v_n) + 4E(v_m) - 8E\left(\frac{v_n + v_m}{2}\right) \\ &\leq 4E(v_n) + 4E(v_m) - 8e_0, \end{aligned}$$

where $\frac{1}{2}(v_n + v_m) \in W$ thanks to the convexity of W . Now $E(v_n), E(v_m) \rightarrow e_0$ implies $\|v_n - v_m\| \rightarrow 0$ as $n, m \rightarrow \infty$. Thus $\{v_n\}_{n=1}^\infty$ is a Cauchy sequence in V and there exists a limit $u \in V$, $v_n \rightarrow u$. Since W is closed, we also have $u \in W$. The continuity of E implies

$$E(u) = \lim_{n \rightarrow \infty} E(v_n) = \inf_{v \in W} E(v).$$

Let us show that the solution $u \in W$ is unique. Suppose that both u_1 and u_2 are solutions. Clearly the sequence $u_1, u_2, u_1, u_2, \dots$ is a minimizing sequence. Above we saw that every minimizing sequence has to be a Cauchy sequence. Thus $u_1 = u_2$. ■

1.2.10 Maximum principles and well-posedness

Another important aspect of elliptic problems is the existence of maximum principles. We find it useful to present several of them here and illustrate how they imply the well-posedness of elliptic problems. The counterpart of the maximum principles on the numerical level are the discrete maximum principles (see, e.g., [11, 14, 19, 31, 57, 67] and [112]), which find particularly important application in problems where physically nonnegative quantities like the temperature, density, or concentration are computed.

Theorem 1.8 (Basic maximum principle) *Consider an open bounded set $\Omega \subset \mathbb{R}^d$ and a symmetric elliptic operator of the form*

$$Lu = - \sum_{i,j=1}^d a_{ij}(\mathbf{x}) \frac{\partial^2 u}{\partial x_i \partial x_j}, \quad (1.71)$$

where $a_{ij} \in C(\Omega)$. Let $u \in C^2(\Omega) \cap C(\overline{\Omega})$ be the solution of the equation $Lu = f$, where $f \in C(\Omega)$ and

$$f \leq 0 \quad \text{in } \Omega.$$

Then the maximum of u in $\overline{\Omega}$ is attained on the boundary $\partial\Omega$. Furthermore it holds that if the maximum is attained at an interior point of Ω , then the function u is constant.

This result remains true under less restrictive assumptions on the coefficients a_{ij} .

Proof: Recall that L is elliptic if the coefficient matrix $A(\mathbf{x}) = \{a_{ij}\}_{i,j=1}^d$ is positive definite in Ω . First we carry out the proof under a stronger assumption that $f < 0$ in Ω . Suppose that there exists some $\tilde{\mathbf{x}} \in \Omega$ such that

$$u(\tilde{\mathbf{x}}) = \sup_{\mathbf{x} \in \Omega} u(\mathbf{x}) > \sup_{\mathbf{x} \in \partial\Omega} u(\mathbf{x}). \quad (1.72)$$

Since $A(\tilde{\mathbf{x}}) = \{a_{ij}(\tilde{\mathbf{x}})\}_{i,j=1}^d$ is symmetric and positive definite, it is diagonalizable and has positive real eigenvalues $\lambda_1(\tilde{\mathbf{x}}), \lambda_2(\tilde{\mathbf{x}}), \dots, \lambda_d(\tilde{\mathbf{x}})$. Thus there exists a nonsingular $d \times d$ matrix C such that

$$\Lambda = C^{-1}A(\tilde{\mathbf{x}})C,$$

where $\Lambda = \text{diag}(\lambda_1(\tilde{\mathbf{x}}), \lambda_2(\tilde{\mathbf{x}}), \dots, \lambda_d(\tilde{\mathbf{x}}))$. In a new coordinate system defined by

$$\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x}) = C\mathbf{x}$$

we have that

$$\begin{aligned} 0 &> f(\tilde{\mathbf{x}}) = (Lu)(\tilde{\mathbf{x}}) \\ &= - \sum_{i,j=1}^d (C^{-1}A(\tilde{\mathbf{x}})C)_{ij} \frac{\partial^2 u}{\partial \xi_i \partial \xi_j}(\tilde{\mathbf{x}}) \\ &= - \sum_{i=1}^d \lambda_i(\tilde{\mathbf{x}}) \frac{\partial^2 u}{\partial \xi_i^2}(\tilde{\mathbf{x}}), \end{aligned} \quad (1.73)$$

which is a contradiction since $\lambda_i(\tilde{\mathbf{x}}) > 0$ for all $1 \leq i \leq d$, and $\tilde{\mathbf{x}} \in \Omega \setminus \partial\Omega$ is a maximum point of u , meaning that

$$\frac{\partial^2 u}{\partial \xi_i^2}(\tilde{\mathbf{x}}) \leq 0 \quad \text{for all } 1 \leq i \leq d.$$

Next let us prove the result for the weaker assumption $f \leq 0$ in Ω . Again, suppose that there exists some $\tilde{x} \in \Omega$ satisfying (1.72). Consider the function

$$h(\mathbf{x}) = \sum_{i=1}^d (x_i - \tilde{x}_i)^2.$$

Since the maximum point \tilde{x} of u lies in the interior of Ω and $h(\mathbf{x})$ is bounded in Ω , for a sufficiently small $\beta > 0$ the function $w(\mathbf{x}) = u(\mathbf{x}) + \beta h(\mathbf{x})$ attains its maximum at some interior point $\mathbf{x}_0 \in \Omega$. Since

$$\frac{\partial^2 h}{\partial x_i \partial x_j}(\mathbf{x}) = 2\delta_{ij} \quad \text{for all } \mathbf{x} \in \Omega,$$

we have

$$(Lw)(\mathbf{x}) = (Lu)(\mathbf{x}) + \beta(Lh)(\mathbf{x}) = f(\mathbf{x}) - 2\beta \sum_{i=1}^d a_{ii}(\mathbf{x}) = \tilde{f}(\mathbf{x}) < 0 \quad \text{in } \Omega.$$

Thus we can apply the result of the first part of the proof. ■

■ EXAMPLE 1.7 (Maximum principle)

Consider an open bounded set $\Omega = (-1, 1)^2 \subset \mathbb{R}^2$ and the Poisson equation

$$-\Delta u = -4 \quad \text{in } \Omega \tag{1.74}$$

($L = -\Delta$ is obtained from (1.71) putting $a_{ij} = \delta_{ij}$). The solution u has the form

$$u(x_1, x_2) = x_1^2 + x_2^2 + C,$$

where $C \in \mathbb{R}$ is an arbitrary constant to be determined from the boundary conditions. Since $f \leq 0$ in Ω , the maximum principle (Theorem 1.8) implies that u attains its maximum on the boundary $\partial\Omega$. This indeed is true, as shown in Figure 1.4.

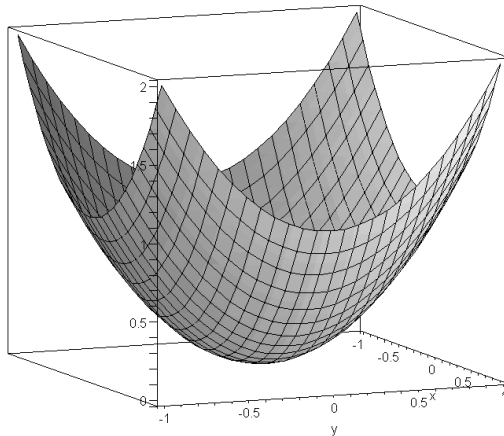


Figure 1.4 Maximum principle for the Poisson equation in 2D.

Immediate consequences of the maximum principle are the minimum principle, comparison principle, and the continuous dependence of the solution on boundary and initial data. Most of these results are straightforward consequences of Theorem 1.8. We encourage the reader to perform the proofs using the hints given.

Corollary 1.1 (Minimum principle) *Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and L an elliptic operator of the form (1.71). If $Lu = f \geq 0$ in Ω , then u attains its minimum on the boundary $\partial\Omega$.*

Proof: Apply Theorem 1.8 to $\tilde{u} := -u$. ■

Corollary 1.2 (Comparison principle) *Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and L an elliptic operator of the form (1.71). Suppose that functions $u, v \in C^2(\Omega) \cap C(\overline{\Omega})$ solve the equations $Lu = f_u$ and $Lv = f_v$, respectively, and*

$$\begin{aligned} f_u &\leq f_v \quad \text{in } \Omega, \\ u &\leq v \quad \text{on } \partial\Omega. \end{aligned}$$

Then $u \leq v$ in Ω .

Proof: Apply the minimum principle to $w := v - u$. ■

Corollary 1.3 (Continuous dependence on boundary data) *Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and L an elliptic operator of the form (1.71). Suppose that u_1 and u_2 solve the equation $Lu = f$ with different Dirichlet boundary data. Then*

$$\sup_{\mathbf{x} \in \Omega} |u_1(\mathbf{x}) - u_2(\mathbf{x})| = \sup_{\mathbf{x} \in \partial\Omega} |u_1(\mathbf{x}) - u_2(\mathbf{x})|.$$

Proof: The function $w = u_1 - u_2$ satisfies the homogeneous equation $Lw = 0$ in Ω . Apply both the maximum and minimum principles to obtain the result. ■

Before introducing the continuous dependence of solution on the right-hand side, we need to define the notion of uniform ellipticity:

Definition 1.7 (Uniform ellipticity) *A linear elliptic operator L of the form (1.4) is said to be uniformly elliptic in an open set $\Omega \subset \mathbb{R}^d$ if there exists a constant $C_A > 0$ such that*

$$\boldsymbol{\xi}^T A(\mathbf{x}) \boldsymbol{\xi} \geq C_A \|\boldsymbol{\xi}\|^2 \quad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^d,$$

and all $\mathbf{x} \in \Omega$, where $A(\mathbf{x})$ is the corresponding coefficient matrix.

Corollary 1.4 (Continuous dependence on the right-hand side) *Let $\Omega \subset \mathbb{R}^d$ be an open bounded set and L an elliptic operator of the form (1.71). Moreover, assume that L is uniformly elliptic in Ω . Then there exists a constant C only depending on the set Ω and the uniform ellipticity constant C_A , such that*

$$|u(\mathbf{x})| \leq \sup_{\mathbf{y} \in \partial\Omega} |u(\mathbf{y})| + C \sup_{\mathbf{y} \in \Omega} |f(\mathbf{y})| \quad (1.75)$$

for all $\mathbf{x} \in \Omega$.

Proof: Since Ω is bounded, it is contained in some open ball $B(0, r)$. Let

$$w(\mathbf{x}) = r^2 - \sum_{i=1}^d x_i^2.$$

Clearly $0 \leq w \leq r^2$ in Ω . Since

$$\frac{\partial^2 w}{\partial x_i \partial x_j} = -2\delta_{ij},$$

it is $Lw \geq 2dC_A$, where C_A is the uniform ellipticity constant of L . Let

$$v(\mathbf{x}) := \sup_{\mathbf{y} \in \partial\Omega} |u(\mathbf{y})| + w(\mathbf{x}) \frac{1}{2dC_A} \sup_{\mathbf{y} \in \partial\Omega} |Lu(\mathbf{y})|.$$

Then $Lv \geq |Lu|$ in Ω and $v \geq |u|$ on $\partial\Omega$. The comparison principle implies that $-v(\mathbf{x}) \leq u(\mathbf{x}) \leq v(\mathbf{x})$ in Ω . Since $w \leq r^2$, (1.75) holds with $C = r^2/(2dC_A)$. ■

Corollary 1.5 (Elliptic operator with a Helmholtz term) *Consider an elliptic operator L of the form*

$$Lu = - \sum_{i,j=1}^d a_{ij}(\mathbf{x}) \frac{\partial^2 u}{\partial x_i \partial x_j} + a_0(\mathbf{x})u$$

with $a_0(\mathbf{x}) \geq 0$ in Ω . Then $Lu \leq 0$ in Ω implies that

$$\sup_{\mathbf{x} \in \Omega} u(\mathbf{x}) \leq \max\{0, \sup_{\mathbf{x} \in \partial\Omega} u(\mathbf{x})\}.$$

Proof: Without loss of generality, let $\mathbf{x}_0 \in \Omega$ be such that

$$u(\mathbf{x}_0) = \sup_{\mathbf{y} \in \Omega} u(\mathbf{y}) > 0.$$

Then $(Lu)(\mathbf{x}_0) - a_0(\mathbf{x}_0)u(\mathbf{x}_0) \leq (Lu)(\mathbf{x}_0) \leq 0$, and the principal part $Lu - a_0u$ defines an elliptic operator of the form (1.71). The conclusion follows from Theorem 1.8. ■

1.2.11 Exercises

Exercise 1.13 Show that the bilinear form $a(\cdot, \cdot)$ from (1.55) is bounded and V -elliptic.

Exercise 1.14 Show that relation (1.35) in Lemma 1.4 defines an inner product. Further show that the energy norm (1.36) induced by this inner product satisfies the relation (1.37) (i.e., that it is equivalent to the norm $\|\cdot\|_V$).

Exercise 1.15 Let $\Omega \subset \mathbb{R}^d$ be an open bounded set with Lipschitz-continuous boundary. Let the boundary $\partial\Omega$ be split into two nonempty disjoint open parts Γ_N and Γ_D such that $\overline{\Gamma_N} \cup \overline{\Gamma_D} = \partial\Omega$. Consider boundary data (real functions) g_N, g_D defined on Γ_N and Γ_D , respectively. Write the weak formulation of the boundary value problem for the Poisson equation

$$-\Delta u = f,$$

equipped with boundary conditions

$$\frac{\partial u}{\partial \nu}(\mathbf{x}) + u(\mathbf{x}) = g_N(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N,$$

and

$$u(\mathbf{x}) = g_D(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D,$$

where f is a real-valued load function defined in Ω . Identify the largest function spaces where the solution u as well as the test functions v and data g_N, g_D , and f must lie in order that all integrals in the weak formulation be defined.

Exercise 1.16 Prove Corollary 1.1.

Exercise 1.17 Prove Corollary 1.2.

1.3 SECOND-ORDER PARABOLIC PROBLEMS

Next let us turn our attention to linear parabolic problems (the notion of parabolicity was introduced in Definition 1.1). Let $\Omega \subset \mathbb{R}^d$ be an open set with Lipschitz-continuous boundary. We will study a class of linear parabolic equations

$$\frac{\partial u}{\partial t} + Lu = f \quad \text{in } \Omega, \quad (1.76)$$

where t is the time, $u = u(\mathbf{x}, t)$, $f = f(\mathbf{x}, t)$ and L is an elliptic operator of the form (1.1) with time-independent coefficients. The equation (1.76) is considered in a space-time cylinder $Q_T = \Omega \times (0, T)$, where $T > 0$.

1.3.1 Initial and boundary conditions

Boundary conditions for parabolic problems are analogous to the elliptic case: Dirichlet, Neumann, Newton, and combined (see Section 1.2). For simplicity, let us denote them by

$$(Bu)(\mathbf{x}, t) = g(\mathbf{x}, t) \quad \text{for all } (\mathbf{x}, t) \in \partial\Omega \times (0, T). \quad (1.77)$$

Parabolic problems describe evolutionary processes, and thus one needs to provide an initial condition of the form

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega. \quad (1.78)$$

If the problem is considered in the classical sense, then the initial condition $u_0(\mathbf{x})$ must moreover satisfy the boundary conditions (this is known as compatibility condition).

1.3.2 Weak formulation

At every time instant the solution is sought in a closed subspace $V \subset H^1(\Omega)$ such that $H_0^1(\Omega) \subset V$. The form of the space V depends on the boundary conditions analogously to the elliptic case (see Paragraphs 1.2.5–1.2.8).

For the analysis of existence and uniqueness of solution we need to introduce function spaces and norms for time-dependent functions:

Definition 1.8 First by $L^q(0, T; W^{k,p}(\Omega))$ we denote the space

$$L^q(0, T; W^{k,p}(\Omega)) = \left\{ u : (0, T) \rightarrow W^{k,p}(\Omega); \right. \\ \left. u \text{ is measurable and } \int_0^T \|u(t)\|_{k,p,\Omega}^q dt < \infty \right\},$$

endowed with the norm

$$\|u\|_{L^q(0,T;W^{k,p}(\Omega))} = \left(\int_0^T \|u(t)\|_{k,p,\Omega}^q dt \right)^{\frac{1}{q}}. \quad (1.79)$$

The symbol $u(t)$ stands for a function of \mathbf{x} such that $u(t) : \mathbf{x} \rightarrow u(\mathbf{x}, t)$. Further we define the space

$$C([0, T]; L^p(\Omega)) = \{u : [0, T] \rightarrow L^p(\Omega); \|u(t)\|_{p,\Omega} \text{ is continuous in } [0, T]\}. \quad (1.80)$$

Analogously we use the $W^{k,p}$ -norm in Ω to define the space

$$C([0, T]; W^{k,p}(\Omega)) = \{u : [0, T] \rightarrow W^{k,p}(\Omega); \|u(t)\|_{k,p,\Omega} \text{ is continuous in } [0, T]\}. \quad (1.81)$$

Weak formulation The weak formulation of parabolic problems is derived using a procedure analogous to elliptic equations. For example, in the case of homogeneous Dirichlet boundary conditions the weak formulation of the problem (1.76), (1.77), (1.78) reads:

Given $f \in L^2(Q_T)$ and $u_0 \in V = H_0^1(\Omega)$, find $u \in L^2(0, T; V) \cap C([0, T]; L^2(\Omega))$ such that

$$\frac{d}{dt}(u(t), v)_{L^2} + a(u(t), v) = (f(t), v)_{L^2} \quad \text{for all } v \in V, t \in (0, T), \quad (1.82)$$

$$u(0) = u_0, \quad (1.83)$$

where the bilinear form $a(\cdot, \cdot)$ corresponds to the elliptic operator (1.1),

$$a(u, v) = \int_{\Omega} \left[\sum_{i,j=1}^d a_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} - \sum_{i=1}^d \left(b_i u \frac{\partial v}{\partial x_i} - c_i v \frac{\partial u}{\partial x_i} \right) + a_0 uv \right] dx. \quad (1.84)$$

The other types of boundary conditions are handled analogously to elliptic problems.

1.3.3 Existence and uniqueness of solution

Since the difficulty of the proof of existence and uniqueness of solution to the problem (1.82), (1.83) exceeds the scope of this text, we restrict ourselves to formulating the principal theoretical result, and providing appropriate references.

We need to introduce the notion of weak coercivity of the form $a(u, v)$ in the space V : There exist two constants $c_{1,2} > 0$ and $c_2 \geq 0$ such that

$$a(u, u) + c_2 \|u\|_{L^2}^2 \geq c_{1,2} \|u\|_{W^{1,2}}^2 \quad \text{for all } u \in V. \quad (1.85)$$

If the form $a(u, v)$ is V -elliptic (coercive), then this inequality holds with $c_2 = 0$. This is the case, for example, for the heat transfer equation $\partial u / \partial t - \Delta u = f$ with homogeneous Dirichlet boundary conditions. In general, condition (1.85) is satisfied for all types of boundary value problems we deal with, provided that all coefficients a_{ij}, b_i, c_i , and a_0 of the operator (1.4) belong to $L^\infty(\Omega)$.

Before introducing the existence and uniqueness theorem, let us show an interesting trick that turns the weakly coercive bilinear form $a(\cdot, \cdot)$ into a coercive one. Applying the substitution

$$\tilde{u}(\mathbf{x}, t) = e^{-c_2 t} u(\mathbf{x}, t),$$

equation (1.76) comes over to the form

$$\frac{\partial \tilde{u}}{\partial t} + L\tilde{u} + c_2 \tilde{u} = e^{-c_2 t} f \quad \text{in } Q_T.$$

Defining $\tilde{f} := e^{-c_2 t} f$ and $\tilde{L} := (L + c_2 I)$, where I stands for the identity operator, the equation returns to the form (1.76). However, if the original bilinear form $a(u, v)$ is weakly coercive, the bilinear form $\tilde{a}(u, v) = a(u, v) + c_2(u, v)$ is coercive. This technique is used in the analysis of parabolic PDEs quite frequently. Now let us formulate the promised existence and uniqueness result:

Theorem 1.9 (Existence and uniqueness of solution) *Let the bilinear form $a(\cdot, \cdot)$ be continuous in $V \times V$ and weakly coercive. Given $f \in L^2(Q_T)$ and $u_0 \in V$, there exists a unique solution $u \in L^2(0, T; V) \cap C([0, T]; L^2(\Omega))$ to the system (1.82), (1.83). Moreover, $\partial u / \partial t \in L^2(0, T; V')$ and the energy estimate*

$$\max_{t \in [0, T]} \|u(t)\|_{L^2}^2 + c_{1,2} \int_0^T \|u(t)\|_{W^{1,2}}^2 \leq \|u(0)\|_{L^2}^2 + \frac{1}{c_{1,2}} \int_0^T \|f(t)\|_2^2 \quad (1.86)$$

holds.

Proof: See, e.g., [93], pages 366 to 369. ■

1.3.4 Exercises

Exercise 1.18 *Let $Q_T = \Omega \times (0, T)$, where $\Omega \subset \mathbb{R}^d$ is an open bounded set with Lipschitz-continuous boundary. Consider the heat-transfer equation*

$$\frac{\partial u}{\partial t} - \Delta u = f \quad \text{in } \Omega, \quad (1.87)$$

$f \in L^2(Q_T)$, equipped with some initial condition $u(\mathbf{x}, 0) = u_0(\mathbf{x})$, $u_0 \in H^1(\Omega)$, and Neumann boundary conditions

$$\frac{\partial u}{\partial \nu} = g \quad \text{on } \partial\Omega, \quad (1.88)$$

$g \in C(\partial\Omega)$.

1. What is the space V in this case?
2. Verify in detail all assumptions of Theorem 1.9 and use it to show the unique solvability of this problem.
3. Consider the elliptic problem $-\Delta u = f$ in Ω , which is the stationary version of equation (1.87), equipped with the pure Neumann boundary conditions (1.88). Does this problem have a unique solution?
4. Explain the difference between the V -ellipticity condition (Definition 1.5) and condition (1.85). What does this difference imply for the unique solvability of elliptic and parabolic problems?

1.4 SECOND-ORDER HYPERBOLIC PROBLEMS

In this section we study linear second-order hyperbolic problems. A model equation with appropriate boundary and initial conditions is formulated in Paragraph 1.4.1. In Paragraph 1.4.2 we derive its weak formulation and present a basic existence and uniqueness result. In Paragraph 1.4.3 we show the link between the second-order hyperbolic equations and first-order hyperbolic systems.

1.4.1 Initial and boundary conditions

The notion of hyperbolicity was first introduced in Definition 1.1. Consider the model equation

$$\frac{\partial^2 u}{\partial t^2} + Lu = f, \quad (1.89)$$

where L is an elliptic operator of the form

$$L = \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial}{\partial x_j} \right) \quad (1.90)$$

with time-independent coefficients. We are interested in solving equation (1.89) in a space-time cylinder $Q_T = \Omega \times (0, T)$, where $\Omega \subset \mathbb{R}^d$ is some open bounded set with Lipschitz-continuous boundary, and $T > 0$.

Let the boundary $\partial\Omega$ be split into two open parts $\Gamma_D, \Gamma_N \subset \partial\Omega$ such that $\Gamma_D \cap \Gamma_N = \emptyset$ and $\overline{\Gamma_D} \cup \overline{\Gamma_N} = \partial\Omega$. We prescribe a Dirichlet boundary condition

$$u(\mathbf{x}, t) = g_D(\mathbf{x}, t) \quad \text{for all } (\mathbf{x}, t) \in \Gamma_D \times (0, T), \quad (1.91)$$

and a Neumann boundary condition

$$\frac{\partial u}{\partial \boldsymbol{\nu}_L}(\mathbf{x}, t) = g_N(\mathbf{x}, t) \quad \text{for all } (\mathbf{x}, t) \in \Gamma_N \times (0, T). \quad (1.92)$$

Here

$$\frac{\partial u}{\partial \boldsymbol{\nu}_L} = \sum_{i,j=1}^d a_{ij} \frac{\partial u}{\partial x_j} n_i$$

is the conormal derivative to $\partial\Omega$, $\boldsymbol{\nu} = (n_1, n_2, \dots, n_d)^T$ being the unit outer normal vector to $\partial\Omega$.

Since the equation is of second-order in time, one has to prescribe initial boundary conditions for both the function values,

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega, \quad (1.93)$$

and the temporal derivative,

$$\frac{\partial u}{\partial t}(\mathbf{x}, 0) = u_1(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega. \quad (1.94)$$

1.4.2 Weak formulation and unique solvability

To avoid complications related to the Dirichlet lift, for simplicity consider homogeneous boundary conditions on $\partial\Omega$. Then $V = H_0^1(\Omega)$, and the weak formulation of the problem (1.89)–(1.94) reads:

Given some right-hand side $f \in L^2(Q_T)$ and initial conditions $u_0 \in V$ and $u_1 \in L^2(\Omega)$, find a function $u \in C([0, T]; V) \cap C^1([0, T]; L^2(\Omega))$ such that

$$\frac{d^2}{dt^2}(u(t), v)_{L^2} + a(u(t), v) = (f(t), v)_{L^2} \quad \text{for all } v \in V, t \in (0, T), \quad (1.95)$$

$$u(0) = u_0, \quad (1.96)$$

$$\frac{du}{dt}(0) = u_1, \quad (1.97)$$

where the bilinear form $a(\cdot, \cdot)$ corresponds to the elliptic operator (1.90).

Theorem 1.10 *Under the above assumptions on the data, the problem (1.95)–(1.97) has a unique solution.*

Proof: The technicality of the proof exceeds the scope of this text. We refer the reader, e.g., to [79] and [94]. ■

1.4.3 The wave equation

Sometimes it is practical to abbreviate the notation for partial derivatives using a subscript, for example, $\partial u / \partial x = u_x$, $\partial u / \partial t = u_t$, $\partial^2 u / \partial x^2 = u_{xx}$, etc. We shall take advantage of this notation in what follows. One of the simplest examples of a second-order hyperbolic equation is the one-dimensional wave equation

$$u_{tt} = c^2 u_{xx}, \quad (1.98)$$

to be satisfied for all $(x, t) \in \mathbb{R} \times (0, T)$. The positive constant $c > 0$ is the wave speed. The equation (1.98) does not require boundary conditions since it is defined in \mathbb{R} , but it has to be supplemented with some initial conditions of the form

$$\begin{aligned} u(x, 0) &= u_0(x), \\ u_t(x, 0) &= u_1(x). \end{aligned} \quad (1.99)$$

Using the substitution

$$v = u_x \quad \text{and} \quad w = u_t,$$

the equation (1.98) comes over to a system of two first-order equations

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} + \begin{pmatrix} -w_x \\ -c^2 v_x \end{pmatrix} = \mathbf{0},$$

which can be written in the matrix form

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ -c^2 & 0 \end{pmatrix} \begin{pmatrix} v_x \\ w_x \end{pmatrix} = \begin{pmatrix} v_t \\ w_t \end{pmatrix} + \mathbf{A} \begin{pmatrix} v_x \\ w_x \end{pmatrix} = \mathbf{0}. \quad (1.100)$$

The initial conditions to (1.100) are

$$\begin{aligned} v(x, 0) &= u'_0(x), \\ w(x, 0) &= u_1(x). \end{aligned} \quad (1.101)$$

This problem belongs to the class of first-order hyperbolic conservation laws that will be studied in Section 1.5. There the reader will learn how to derive the exact solution to (1.98), (1.99) in the form

$$u(x, t) = \frac{1}{2} \left[u_0(x - ct) + u_0(x + ct) - \frac{1}{c} U_1(x - ct) + \frac{1}{c} U_1(x + ct) \right], \quad (1.102)$$

where $U_1(x)$ is a primitive function to $u_1(x)$.

1.4.4 Exercises

Exercise 1.19 Can equation (1.89), when equipped with a Neumann boundary condition on the whole boundary $\partial\Omega$, have a unique solution? How would this change in the stationary case $Lu = f$?

Exercise 1.20 Calculate the eigenvalues and eigenvectors of the matrix \mathbf{A} in (1.100).

Exercise 1.21 Verify that the function $u(x, t)$ defined in (1.102) is the exact solution of the 1D wave equation (1.98) with the initial conditions (1.99).

1.5 FIRST-ORDER HYPERBOLIC PROBLEMS

This section is devoted to first-order hyperbolic problems of the form

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) + \operatorname{div} \mathbf{f}(\mathbf{u}(x, t)) = 0. \quad (1.103)$$

These equations differ from the previously studied second-order PDEs significantly and methods other than FEM are usually used for their numerical solution. PDEs of the form (1.103) are referred to as conservation laws, and they play an important role in the continuum mechanics and fluid dynamics.

The (generally nonlinear) flux function $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_d)^T$, where d is the spatial dimension, consists of d directional fluxes $\mathbf{f}_i : \mathbb{R}^m \rightarrow \mathbb{R}^m$ that describe the transport of the solution in the axial directions x_i . The equation (1.103) is equipped with an initial condition

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}).$$

Boundary conditions are not required if the problem is stated in $\Omega = \mathbb{R}^d$, otherwise suitable conditions on the boundary have to be imposed. An example of a conservation law are the Euler equations of compressible inviscid flow, which consist of the law of conservation of mass (continuity equation), law of conservation of momentum (Euler momentum equations), and the law of conservation of energy. For the analysis and numerical solution of the compressible Euler equations see, e.g., [52] and the references therein.

After a brief general introduction in Paragraph 1.5.1 we begin with the study of scalar and vector-valued linear conservation laws in one spatial dimension. Due to the existence of characteristics, the solutions of conservation laws have a unique structure. Characteristics are space-time curves that distribute the information from the initial and boundary conditions through the space-time cylinder $Q_T = \Omega \times (0, T)$. We will define and study the characteristics in Paragraph 1.5.2, and consequently utilize them to construct the exact solutions to a general one-dimensional linear first-order system in Paragraph 1.5.3.

Exciting things happen when the flux function \mathbf{f} is nonlinear. Nonlinear hyperbolic systems exhibit discontinuous solutions, a feature unknown in elliptic and parabolic problems. The discontinuities, which may arise at finite times and even in problems with infinitely smooth initial and boundary data, banish the solution from Sobolev spaces and pose serious difficulties to both the analysis and numerical solution of hyperbolic problems. In Paragraph 1.5.5 we exploit the characteristics introduced in Paragraph 1.5.2 to understand the mechanism of creation of discontinuities in solutions to nonlinear hyperbolic problems.

1.5.1 Conservation laws

In one spatial dimension the conservation law (1.103) takes the form

$$\frac{\partial}{\partial t} \mathbf{u}(x, t) + \frac{\partial}{\partial x} \mathbf{f}(\mathbf{u}(x, t)) = 0, \quad (1.104)$$

where $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the flux function and $\mathbf{u} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^m$ is m -dimensional vector of conserved quantities (state variables) such as, e.g., the mass, momentum or energy. When we say that a quantity $\mathbf{u}(x, t)$ is conserved, we mean that all its components satisfy

$$\int_{\mathbb{R}} u_i(x, t) \, dx = \text{const}_i, \quad (1.105)$$

or,

$$\frac{d}{dt} \int_{\mathbb{R}} u_i(x, t) \, dx = 0. \quad (1.106)$$

Notice that while satisfying (1.106), the functions u_i themselves may change in time. Moreover notice that (1.104) implies (1.106).

Definition 1.9 (Cauchy problem) *By Cauchy problem we mean the pure initial-value problem where one requires that (1.104) holds for all $x \in \mathbb{R}$ and all $t \geq 0$. In this case one has to specify the initial condition only,*

$$\mathbf{u}(x, 0) = \mathbf{u}_0(x), \quad x \in \mathbb{R}.$$

Of particular interest are conservation laws (1.104) which are hyperbolic:

Definition 1.10 (Hyperbolicity) *The system (1.104) is said to be hyperbolic if the flux function \mathbf{f} is continuously differentiable and the $m \times m$ Jacobi matrix $D\mathbf{f}/D\mathbf{u}$ is diagonalizable and has real eigenvalues only.*

Recall that a square $m \times m$ matrix is diagonalizable if and only if it is similar to a diagonal matrix (Definition A.20). It is worth mentioning that the first-order system (1.100) associated with the second-order wave equation (1.98) was a hyperbolic conservation law: The flux function was linear, $\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$, and the eigenvalues of its Jacobi matrix $D\mathbf{f}/D\mathbf{u} = \mathbf{A}$ were real numbers $\pm c$.

More generally, in \mathbb{R}^d the conservation law (1.103) takes the form

$$\frac{\partial}{\partial t} \mathbf{u}(\mathbf{x}, t) + \sum_{i=1}^3 \frac{\partial}{\partial x_i} \mathbf{f}_i(\mathbf{u}(\mathbf{x}, t)) = 0, \quad (1.107)$$

where $\mathbf{u} : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^m$, and $\mathbf{f}_1, \dots, \mathbf{f}_d : \mathbb{R}^m \rightarrow \mathbb{R}^m$ are flux functions in the directions x_1, \dots, x_d . Equation (1.107) is said to be hyperbolic if every linear combination of the Jacobi matrices

$$\sum_{i=1}^d a_i \frac{D\mathbf{f}_i}{D\mathbf{u}}, \quad (1.108)$$

where $a_i \in \mathbb{R}$ are arbitrary constants, is diagonalizable and has real eigenvalues only.

The Reynolds' transport theorem Conservation laws come from physics, where in most cases they are stated in integral form. For example, the law of mass conservation in fluids holds in the integral form

$$\frac{d}{dt} \int_{\sigma(t)} \varrho(\mathbf{x}, t) \, d\mathbf{x} = 0, \quad (1.109)$$

where $\sigma(t)$ is an arbitrary control volume. Control volume is a volume of fluid that is formed by the same particles at all times, and the integral of the density ϱ over $\sigma(t)$ yields the mass of $\sigma(t)$.

Since the integral formulations of conservation laws are very difficult to handle numerically, it is customary to use the Reynolds' transport theorem to convert them into PDEs. For a general density function $\mathcal{D}(\mathbf{x}, t)$ and under suitable regularity assumptions (see, e.g., [52]) the Reynolds' transport theorem says

$$\frac{d}{dt} \int_{\sigma(t)} \mathcal{D} \, d\mathbf{x} = \int_{\sigma(t)} \left(\frac{\partial \mathcal{D}}{\partial t} + \operatorname{div}(\mathcal{D}\mathbf{v}) \right) d\mathbf{x}, \quad (1.110)$$

where $\mathbf{v}(\mathbf{x}, t)$ is the fluid velocity. Applying (1.110)–(1.109) with $\mathcal{D} = \varrho$, we obtain

$$0 = \frac{d}{dt} \int_{\sigma(t)} \varrho(\mathbf{x}, t) \, d\mathbf{x} = \int_{\sigma(t)} \left(\frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho\mathbf{v}) \right) d\mathbf{x}. \quad (1.111)$$

Since the control volume $\sigma(t) \subset \Omega$ in (1.111) is arbitrary, the standard localization theorem says that the integrand has to be zero almost everywhere in Ω . Thus (1.111) yields the continuity equation,

$$\frac{\partial \varrho}{\partial t} + \operatorname{div}(\varrho\mathbf{v}) = 0 \quad \text{a.e. in } Q_T = \Omega \times (0, T). \quad (1.112)$$

The localization theorem is intuitively clear and its proof straightforward. In particular, if the function ϱ is continuous, (1.112) holds everywhere in Q_T . For $\varrho \in H^1(\Omega)$ one proceeds by the density argument (see the end of Paragraph A.2.10).

Standard difficulties related to conservation laws The transformation of an integral equation to a PDE is not an equivalent operation. Usually the PDE is less general, undefined on discontinuities (shocks) where the integral form holds. Therefore one has to go back to the integral equation and derive suitable jump conditions to hold at the discontinuities and incorporate them back into the weak formulation of the PDE.

The weak solution usually admits more solutions than the unique physically admissible solution corresponding to the integral form of the conservation law. Therefore one has to impose some selection principle that excludes nonphysical solutions. For fluid dynamics problems one can appeal the second law of thermodynamics which states that the entropy is not decreasing. In particular, as molecules of a fluid pass through a shock, their entropy must increase. It turns out that this condition is sufficient to reliably distinguish between physically correct and incorrect discontinuities. Generally, such conditions are called entropy conditions.

1.5.2 Characteristics

The existence of characteristics (characteristic curves) is a unique aspect of hyperbolic PDEs. These space-time curves determine how the values of the initial and boundary conditions are distributed through the space-time cylinder $Q_T = \Omega \times (0, T)$.

To begin with, consider a constant $a \in \mathbb{R}$ and the Cauchy problem for a scalar hyperbolic equation with the linear flux function $f(u) = au$,

$$u_t + au_x = 0 \quad \text{for all } x \in \mathbb{R}, t > 0, \quad (1.113)$$

equipped with the initial condition

$$u(x, 0) = u_0(x) \quad \text{for all } x \in \mathbb{R}. \quad (1.114)$$

Definition 1.11 (Characteristics) *Characteristic curve of equation (1.113), passing through the point $(x_0, 0)$, $x_0 \in \mathbb{R}$, is the graph of the solution of the ordinary differential equation*

$$\begin{aligned} x'(t) &= a \quad \text{for all } t > 0, \\ x(0) &= x_0. \end{aligned} \quad (1.115)$$

Lemma 1.6 *The solution of (1.113), (1.114) is constant along the characteristics $x(t)$, and thus it is fully determined by the initial data,*

$$u(x, t) = u_0(x - at). \quad (1.116)$$

Proof: Since $a \in \mathbb{R}$ is constant, by (1.115) the characteristics are straight lines,

$$x(t) = at + x_0.$$

Consider the solution along these lines, $u(x(t), t)$, and take its derivative in time. Using the original equation (1.113), we obtain

$$\frac{d}{dt}u(at + x_0, t) = a \frac{\partial u}{\partial x}(x(t), t) + \frac{\partial}{\partial t}u(x(t), t) = 0.$$

For an arbitrary $(x, t) \in \mathbb{R} \times (0, T)$, the characteristics $x(t)$ passing through this point intersects with the real axis at $x(0) = x - at$, where it takes the value $u(x, t) = u(x - at, 0) = u_0(x - at)$. ■

Remark 1.7 (Equation (1.113) describes “flow”) *Equation (1.113) does not generate any new information, it only shifts the initial condition u_0 in time. The initial condition moves to the right if $a > 0$ and to the left if $a < 0$. In the degenerated case of $a = 0$ the equation reduces to $\partial u / \partial t = 0$, i.e., the solution is constant in time, which is compatible with the fact that the characteristics have the form $x(t) = x_0$.*

1.5.3 Exact solution to linear first-order systems

The next natural step to take is to analyze linear vector-valued problems in one spatial dimension. Hence, for $m \geq 1$ consider the hyperbolic conservation law (1.104) with a linear flux function $\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$,

$$\mathbf{u}_t(x, t) + \mathbf{A}\mathbf{u}_x(x, t) = 0, \quad (1.117)$$

$$\mathbf{u}(x, 0) = \mathbf{u}_0(x), \quad (1.118)$$

where $\mathbf{u} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^m \times \mathbb{R}^m$ is a constant matrix. By the hyperbolicity of the problem the matrix \mathbf{A} is diagonalizable with real eigenvalues, i.e., there exists a nonsingular $m \times m$ matrix \mathbf{R} such that

$$\mathbf{A} = \mathbf{R}\mathbf{\Lambda}\mathbf{R}^{-1}. \quad (1.119)$$

Here $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ is a diagonal eigenvalue matrix, and it is worth mentioning that the matrix \mathbf{R} contains the right eigenvectors of \mathbf{A} in its columns. Thus for the columns of \mathbf{R} we have

$$\mathbf{A}\mathbf{r}_i = \lambda_i\mathbf{r}_i \quad \text{for all } 1 \leq i \leq m.$$

Let us introduce the notion of strict hyperbolicity for reference:

Definition 1.12 (Strictly hyperbolic system) *The system (1.117), (1.118) is called strictly hyperbolic if the eigenvalues λ_i , $1 \leq i \leq m$, are distinct.*

Characteristic variables One can solve (1.117), (1.118) by switching to characteristic variables

$$\mathbf{v} = \mathbf{R}^{-1}\mathbf{u}.$$

Multiplying (1.117) by \mathbf{R}^{-1} and using (1.119), one obtains

$$\mathbf{R}^{-1}\mathbf{u}_t + \Lambda\mathbf{R}^{-1}\mathbf{u}_x = \mathbf{0},$$

which further yields

$$\mathbf{v}_t + \Lambda\mathbf{v}_x = 0. \quad (1.120)$$

By the diagonality of Λ , this is a system of m independent linear advection equations for the components of \mathbf{v} ,

$$\begin{aligned} (v_i)_t + \lambda_i(v_i)_x &= 0, \\ v_i(0) &= v_{0,i}, \end{aligned}$$

$i = 1, 2, \dots, m$. The initial condition for v_i is the i th component of the vector $\mathbf{R}^{-1}\mathbf{u}_0$. Using what we learned in Paragraph 1.5.2, for each $1 \leq i \leq m$ the solution is

$$v_i(x, t) = v_i(x - \lambda_i t, 0) = v_{0,i}(x - \lambda_i t).$$

The solution \mathbf{u} is finally recovered using the relation

$$\mathbf{u}(x, t) = \mathbf{R}\mathbf{v}(x, t) = \sum_{i=1}^m v_i(x, t)\mathbf{r}_i, \quad (1.121)$$

which yields

$$\mathbf{u}(x, t) = \sum_{i=1}^m v_{0,i}(x - \lambda_i t)\mathbf{r}_i. \quad (1.122)$$

Simple waves The solution (1.122) is the superposition of m independently advected linear waves. The i th wave has the form

$$v_i(x, 0)\mathbf{r}_i,$$

and propagates at the wave speed λ_i .

1.5.4 Riemann problem

The solution of the Riemann problem plays an important role in the design of finite volume methods for the approximate solution of nonlinear conservation laws.



Figure 1.5 Georg Friedrich Bernhard Riemann (1826–1866).

G.F.B. Riemann was a German mathematician who, besides other important achievements, introduced topological methods into the theory of complex functions, studied the representation of functions by trigonometric series, and established new foundations of geometry which were used later in relativity and cosmology. The Riemann hypothesis, related to the prime number theory, remains one of the most famous unsolved problems of modern mathematics.

Consider the one-dimensional linear hyperbolic equation (1.117),

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = 0, \quad (1.123)$$

with a piecewise-constant initial condition consisting of two different states $\mathbf{u}_L, \mathbf{u}_R \in \mathbb{R}^m$ on the negative and positive half of the real line, respectively,

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_L & x \leq 0, \\ \mathbf{u}_R & x > 0. \end{cases} \quad (1.124)$$

For simplicity we assume that the problem (1.123) is strictly hyperbolic. This means that the matrix \mathbf{A} has m eigenvalues which are real and distinct. They can be denoted as follows,

$$\lambda_1 < \lambda_2 < \dots < \lambda_m.$$

Exact solution in characteristic variables Recall that the exact solution to (1.123) is given by (1.122). We can simplify the situation by expressing the initial states \mathbf{u}_L and \mathbf{u}_R in terms of eigenvectors of the matrix \mathbf{A} ,

$$\mathbf{u}_L = \sum_{i=1}^m \alpha_i \mathbf{r}_i, \quad \mathbf{u}_R = \sum_{i=1}^m \beta_i \mathbf{r}_i.$$

Then

$$v_i(x, 0) = \begin{cases} \alpha_i & x \leq 0, \\ \beta_i & x > 0, \end{cases}$$

and the problem (1.123) decouples into m independent scalar Riemann problems

$$(v_i)_t + \lambda_i(v_i)_x = 0, \quad (1.125)$$

$$v_i(x, 0) = \begin{cases} \alpha_i & x \leq 0, \\ \beta_i & x > 0. \end{cases}$$

For i th scalar problem, the initial discontinuity $[\beta_i - \alpha_i]$ at $x = 0$ propagates into the space-time domain along the characteristics $x_i(t) = \lambda_i t$, as illustrated in Figure 1.6.

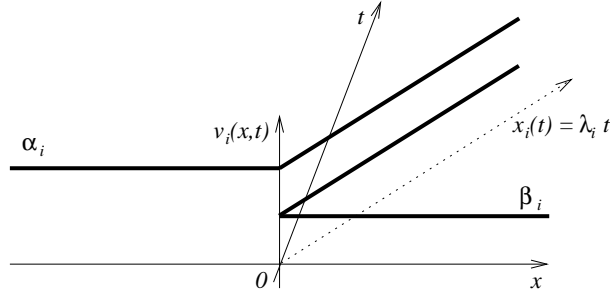


Figure 1.6 Propagation of the jump $[\beta_i - \alpha_i]$ in the i th characteristic variable $v_i(x, t)$ along the i th zero characteristics $x_i(t) = \lambda_i t$.

Solution at $x = 0$ Finite volume schemes are based on the value of the solution $\mathbf{u}(0, t)$, which is constant in time. It is defined if $\lambda_i \neq 0$ for all i (i.e., if no jump is propagated along the temporal axis $x = 0$). It is easy to see that the characteristic variable v_i satisfies

$$v_i(0, t) = \begin{cases} \alpha_i & \lambda_i \geq 0, \\ \beta_i & \lambda_i < 0. \end{cases}$$

Equation (1.121) then yields

$$\mathbf{u}(0, t) = \mathbf{R}\mathbf{v}(0, t) = \sum_{i=1}^m v_i(0, t)\mathbf{r}_i.$$

Let the first m_0 eigenvalues λ_i be negative and the rest positive. Then the exact solution at $x = 0$ can be expressed as

$$\mathbf{u}(0, t) = \sum_{i=1}^{m_0} \beta_i \mathbf{r}_i + \sum_{i=m_0+1}^m \alpha_i \mathbf{r}_i.$$

An important quantity is the (also time-independent) value of $\mathbf{A}\mathbf{u}(0, t)$ that represents the linear flux across the interface $x = 0$,

$$\begin{aligned} \mathbf{A}\mathbf{u}(0, t) &= \mathbf{A} \sum_{i=1}^{m_0} \beta_i \mathbf{r}_i + \mathbf{A} \sum_{i=m_0+1}^m \alpha_i \mathbf{r}_i & (1.126) \\ &= \sum_{i=1}^{m_0} \beta_i \lambda_i \mathbf{r}_i + \sum_{i=m_0+1}^m \alpha_i \lambda_i \mathbf{r}_i \\ &= \sum_{i=1}^{m_0} \beta_i \lambda_i^- \mathbf{r}_i + \sum_{i=1}^m \alpha_i \lambda_i^+ \mathbf{r}_i \\ &= \mathbf{A}^- \mathbf{u}_R + \mathbf{A}^+ \mathbf{u}_L. \end{aligned}$$

Here

$$\begin{aligned}\lambda_i^- &= \min(\lambda_i, 0), \\ \lambda_i^+ &= \max(\lambda_i, 0).\end{aligned}$$

The matrices \mathbf{A}^- , \mathbf{A}^+ are the negative and positive parts of the matrix \mathbf{A} , defined using the decomposition $\mathbf{A} = \mathbf{R}\mathbf{\Lambda}\mathbf{R}^{-1}$ and $\lambda_i = \lambda_i^- + \lambda_i^+$, as

$$\begin{aligned}\mathbf{A}^- &= \mathbf{R}\mathbf{\Lambda}^-\mathbf{R}^{-1}, \\ \mathbf{A}^+ &= \mathbf{R}\mathbf{\Lambda}^+\mathbf{R}^{-1}.\end{aligned}$$

Here $\mathbf{\Lambda}^- = \text{diag}(\lambda_1^-, \lambda_2^-, \dots, \lambda_m^-)$ and $\mathbf{\Lambda}^+ = \text{diag}(\lambda_1^+, \lambda_2^+, \dots, \lambda_m^+)$. Obviously, $\mathbf{A} = \mathbf{A}^- + \mathbf{A}^+$. Analogously we define the absolute value of the matrix \mathbf{A} , $|\mathbf{A}| = \mathbf{A}^+ - \mathbf{A}^- = \mathbf{R}|\mathbf{\Lambda}|\mathbf{R}^{-1}$, where $|\mathbf{\Lambda}| = \text{diag}(|\lambda_1|, |\lambda_2|, \dots, |\lambda_m|)$.

Application to nonlinear conservation laws The matrices \mathbf{A}^+ , \mathbf{A}^- , $|\mathbf{A}|$ are used by several popular finite volume schemes for the solution of nonlinear hyperbolic conservation laws, including the compressible Euler equations. The basic idea of the approximation consists in the linearization of the nonlinear flux functions (their replacement with their Jacobi matrices) and consequent application of the above-described procedure for the linear Riemann problem. The approximation of the time-independent value $\mathbf{A}u(0, t)$ plays a key role in the finite volume schemes. Let us stop the comment at this point, since the finite volume method lies beyond the scope of this text. There is a vast literature devoted to this topic. We refer the reader, e.g., to [52, 54, 77] and [78].

1.5.5 Nonlinear flux and shock formation

To illustrate the mechanism of the creation of discontinuities in nonlinear first-order hyperbolic problems, consider a nonlinear analogy to (1.113), (1.114),

$$u_t(x, t) + [f(u(x, t))]_x = 0 \quad \text{for all } x \in \mathbb{R}, t > 0, \quad (1.127)$$

$$u(x, 0) = u_0(x), \quad (1.128)$$

where the flux function $f : \mathbb{R} \rightarrow \mathbb{R}$ is once continuously differentiable. For demonstration purposes let us pick the function

$$f(u) = \frac{1}{2}u^2.$$

This choice leads to Burgers' equation (1.11),

$$u_t(x, t) + u(x, t)u_x(x, t) = 0. \quad (1.129)$$

The characteristics of equation (1.127) are defined as

$$\begin{aligned}x'(t) &= \frac{df}{du}(u(x(t), t)) = u(x(t), t), \\ x(0) &= x_0.\end{aligned} \quad (1.130)$$

Using (1.130) and (1.129), it is easy to verify that the solution $u(x(t), t)$ along these characteristics is constant,

$$\frac{d}{dt}u(x(t), t) = \frac{\partial u}{\partial x}(x(t), t) \underbrace{u(x(t), t)}_{x'(t)} + \frac{\partial}{\partial t}u(x(t), t) = 0.$$

Since $x'(t) = u(x(t), t)$ is the slope of the characteristics and $u(x(t), t)$ is constant, also in this case the characteristics are straight lines. A characteristic curve passing through $(x_0, 0)$ has the slope $u(x_0, 0) = u_0(x_0)$. When two different characteristic curves, carrying two different values of the solution on them, intersect, a discontinuity (shock) is born. This is illustrated in Figure 1.7.

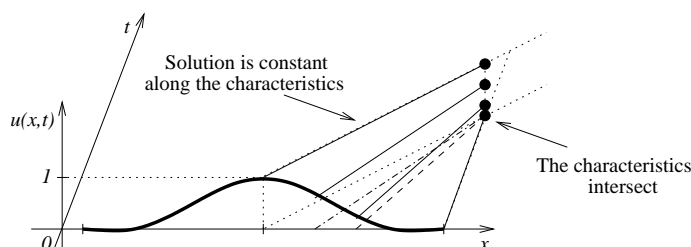


Figure 1.7 Formation of shock in the solution $u(x, t)$ of Burgers' equation.

Nonlinear hyperbolic problems constitute a more or less autonomous field in applied mathematics, and there is a wide class of literature dedicated to both their theoretical and computational aspects. See the literature listed at the end of the previous paragraph and references therein.

1.5.6 Exercises

Exercise 1.22 Under sufficient regularity conditions for the flux \mathbf{f} and the solution \mathbf{u} , show that every solution \mathbf{u} of (1.104) is conserved in time, i.e., it satisfies condition (1.106). *Hint:* Integrate (1.104) over \mathbb{R} , use the fundamental theorem of calculus and decay conditions for functions integrable in \mathbb{R} .

Exercise 1.23 (Exact solution to the wave equation) Calculate the eigenvectors of the matrix \mathbf{A} defined in (1.100). Use the characteristic variables to construct the exact solution (1.102) of the linear first-order hyperbolic system (1.100), (1.101).

Exercise 1.24 Prove a simplified version of the localization theorem: Let $\Omega \subset \mathbb{R}^d$ be an open bounded set. Let $f \in C(\overline{\Omega})$. Let

$$\int_{\sigma} f \, dx = 0$$

be valid for all open bounded sets $\sigma \subset \Omega$. Then f is zero everywhere in Ω .

Exercise 1.25 Consider a linear hyperbolic problem of the form (1.117), (1.118) with the flux function $\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$, where the matrix \mathbf{A} has the form

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Consider a general initial condition $\mathbf{u}(x, 0) = \mathbf{u}_0(x)$ for all $x \in \mathbb{R}$. Write the exact solution to this problem.