

Theory and Numerics of Partial Differential Equations

*A Beginner's Guide to Linear, Nonlinear,
and Stochastic Problems*

Evelyn Sander & Thomas Wanner
Department of Mathematical Sciences
George Mason University
Fairfax, VA 22030, USA

Preliminary version, September 18, 2018

Contents

Contents	iii
Preface	ix
1 Introduction to Differential Equations	1
1.1 Ordinary and Partial Differential Equations	1
1.1.1 Ordinary Differential Equations	1
1.1.2 Existence and Uniqueness of Solutions	4
1.1.3 A First Glimpse of Partial Differential Equations	7
1.1.4 Domains and Their Boundaries	10
1.2 Four Classical Partial Differential Equations	23
1.2.1 The Heat Equation	23
1.2.2 Laplace's Equation	28
1.2.3 The Transport Equation	37
1.2.4 The Wave Equation	43
1.3 Classification of Partial Differential Equations	55
1.3.1 Classification of Linear Second-Order Equations	56
1.3.2 Linear and Nonlinear Equations	60
1.3.3 A Deeper Look: Higher-Order Equations	64
1.4 A Deeper Look: Stochastic Differential Equations	71
1.4.1 Random Effects and Brownian Motion	71
1.4.2 Basic Properties of the Wiener Process	81
1.4.3 Differential Equations Perturbed by Additive Noise	87
1.4.4 A First Glance at Ito's Stochastic Calculus	93
1.5 Review Questions and Problems	102
1.5.1 Review Questions	102
1.5.2 Problems	105
2 Separation of Variables	109
2.1 Generalized Fourier Series Expansions	109
2.1.1 Orthogonal Functions	110
2.1.2 Complete Orthonormal Sets	115
2.1.3 Extensions and Generalizations	126
2.2 Sturm-Liouville Boundary Value Problems	128
2.2.1 Symmetry and Regular Sturm-Liouville Problems	128
2.2.2 Some Concrete Sturm-Liouville Problems	141
2.2.3 A Deeper Look: Comparison and Oscillations	155
2.3 A First Taste of Separation of Variables	170

2.3.1	The Vibrating String of Finite Length	171
2.3.2	Two Elliptic Problems on a Rectangle	182
2.3.3	The Diffusion of Heat in One Dimension	190
2.4	Eigenvalue Problems via Separation of Variables	194
2.4.1	Eigenfunctions and Their Geometry	194
2.4.2	Completeness of the Set of Eigenfunctions	197
2.4.3	Solving Time-Dependent Problems via Eigenfunctions .	202
2.5	Nonhomogeneous Equations and Boundary Conditions	209
2.5.1	Solving Nonhomogeneous Linear Evolution Equations	210
2.5.2	Nonhomogeneous Boundary Conditions	218
2.5.3	A Deeper Look: The Laplace Transform	228
2.6	Separation of Variables in Different Coordinate Systems	236
2.6.1	Polar Coordinates in the Plane	236
2.6.2	Cylindrical Coordinates and Bessel Functions	243
2.6.3	Spherical Coordinates and Spherical Harmonics	245
2.7	Review Questions and Problems	251
2.7.1	Review Questions	251
2.7.2	Problems	253
3	Basic Numerical Techniques	263
3.1	Numerical Computation of Derivatives	263
3.1.1	Finite Difference Approximations of the Derivative . .	265
3.1.2	Numerical Approximation of the Jacobian Matrix . . .	269
3.1.3	Choosing the Optimal Increment	270
3.2	Numerical Solution of Nonlinear Equations	272
3.2.1	Root Finding in One Dimension	274
3.2.2	Newton's Method for Systems of Equations	284
3.2.3	Broyden's Method for Systems	290
3.3	Numerics for Ordinary Differential Equations	294
3.3.1	Three Basic Numerical Methods	295
3.3.2	Stability and Stiff Differential Equations	310
3.3.3	A Deeper Look: Error Propagation & Convergence ..	319
3.4	Differentiation of Periodic Functions	331
3.4.1	The Discrete Fourier Transform	332
3.4.2	Trigonometric Interpolation	336
3.4.3	Derivative Approximation for Periodic Functions . .	342
3.5	Spectral Differentiation for Non-Periodic Functions	353
3.5.1	Dirichlet and Neumann Boundary Conditions	354
3.5.2	Chebyshev Differentiation	362
3.5.3	Two-Dimensional Spectral Methods	377
3.6	A Deeper Look: Numerics for Stochastic Equations	388
3.6.1	Uniform Random Number Generation	388
3.6.2	Gaussian Random Numbers and Brownian Motion . .	395
3.6.3	Stochastic Differential Equation Numerics	401
3.7	Review Questions and Problems	412
3.7.1	Review Questions	412
3.7.2	Problems	413
4	Linear Elliptic Equations	419
4.1	Introduction to Linear Elliptic Problems	419

4.1.1	Notation and Linear Elliptic Test Problems	419
4.1.2	Three Approximation Methods for Derivatives	424
4.1.3	A Deeper Look: Elliptic Existence Theory	427
4.2	Finite Difference Method	449
4.2.1	One-Dimensional Boundary Value Problems	450
4.2.2	Theoretical Aspects of Finite Differences	463
4.2.3	Two-Dimensional Equations	473
4.3	Spectral Methods	483
4.3.1	Limitations of Trigonometric Interpolation	483
4.3.2	Chebyshev Differentiation	488
4.3.3	Domain Transformations and Conformal Mappings . .	503
4.4	Eigenvalue Problems	517
4.4.1	Numerics of Elliptic Eigenvalue Problems	517
4.4.2	A Deeper Look: A Variational Approach	537
4.4.3	A Deeper Look: Spectral Properties	548
4.5	A Look Beyond: Musical Acoustics	563
4.5.1	Theoretical and Numerical Foundations	563
4.5.2	Strings and Wind Instruments	568
4.5.3	Drums and Vibrating Plates	575
4.6	A Deeper Look: Brownian Motion and Elliptic Problems	587
4.6.1	Harmonic Functions via Random Walks	587
4.6.2	Occupancy Times and the Poisson Equation	594
4.6.3	Diffusion Processes and General Elliptic Operators . .	599
4.7	Review Questions and Problems	605
4.7.1	Review Questions	605
4.7.2	Problems	608
5	Nonlinear Elliptic Equations	611
5.1	Finite Difference Method	611
5.1.1	The Stationary Allen-Cahn Equations	611
5.1.2	The Stationary Schnakenberg Model	613
5.1.3	Finite Difference Methods on the Unit Square	615
5.2	Chebyshev Spectral Methods	616
5.2.1	The Basic Method	616
5.2.2	Mixed Boundary Conditions	616
5.2.3	Polygonal Domains via Conformal Maps	616
5.3	Spectral Galerkin Methods	617
5.3.1	The Spectral Galerkin Approach	617
5.3.2	The Stationary Allen-Cahn Equation Revisited	625
5.3.3	Diblock-Copolymer Stationary States	628
5.4	Theoretical Aspects of Nonlinear Elliptic Equations	633
5.4.1	Parameter-Dependent Nonlinear Problems	633
5.4.2	On the Location of Bifurcation Points	636
5.4.3	A Deeper Look: Conditions for Bifurcation	636
5.5	A Look Beyond: Continuation Methods	640
5.5.1	Predictor-Corrector Methods	640
5.5.2	Continuation for One- and Two-Dimensional Problems	644
5.5.3	Bifurcation Detection	647
5.6	A Look Beyond: Equilibria in Materials Science Problems	649
5.6.1	Continuation in Stationary Cahn-Hilliard Equations .	649

5.7	5.6.2 Equilibria for Stationary Diblock-Copolymers	649
	Review Questions and Problems	649
	5.7.1 Review Questions	649
	5.7.2 Problems	650
6	Parabolic Equations	653
6.1	Numerical Methods for Parabolic Problems	653
	6.1.1 Semi-Discretization Methods for Linear Equations	653
	6.1.2 Numerics for Nonlinear Parabolic Equations	660
	6.1.3 A Deeper Look: Stability and Convergence	664
6.2	A Look Beyond: Dynamics of Material Microstructures	664
	6.2.1 The Allen-Cahn Equation	664
	6.2.2 The Cahn-Hilliard Equation	664
	6.2.3 The Diblock-Copolymer Equation	666
6.3	A Look Beyond: Sea Shells and Animal Coats	666
	6.3.1 Turing Patterns and Morphogenesis	666
	6.3.2 Why are Animal Coats Unique?	666
	6.3.3 Formation of Sea Shell Patterns	666
6.4	Long-Term Dynamics and Global Attractors	666
	6.4.1 Equilibrium Solutions and Their Stability	666
	6.4.2 Global Attractors and Their Dimension	668
	6.4.3 A Look Beyond: Slow Motion of Interfaces	668
6.5	A Look Beyond: Parabolic Equations with Additive Noise	668
	6.5.1 Noise Model and Its Numerical Simulation	668
	6.5.2 Phase Separation in the Cahn-Hilliard-Cook Model	668
	6.5.3 Nucleation in Stochastic Systems	668
6.6	A Deeper Look: Stochastic Aspects of Parabolic Problems	668
	6.6.1 Kolmogorov's Backward Equation	668
	6.6.2 Densities and the Fokker-Planck Equation	669
	6.6.3 Invariant Measures for Stochastic Equations	669
6.7	Review Questions and Problems	669
	6.7.1 Review Questions	669
	6.7.2 Problems	669
7	Hyperbolic Equations	671
7.1	Introduction to Hyperbolic Equations	671
	7.1.1 The Viscous Burgers' Equation	671
	7.1.2 First Order Equations	675
	7.1.3 Shock Waves	680
	7.1.4 Rarefaction Solutions	684
	7.1.5 Second Order Equations	685
7.2	Characteristics	686
	7.2.1 The Symbol of a Linear Differential Operator	687
	7.2.2 Definition of Characteristics	688
7.3	Numerical Methods for First Order Problems	693
	7.3.1 First Order Quasi-Linear Equations	693
	7.3.2 First Order Nonlinear Equations	693
	7.3.3 First Order Upwind Methods	693
	7.3.4 Lax-Wendroff and Wendroff Methods	693
7.4	Numerical Methods for Second Order Problems	693

7.4.1	Second Order Quasi-Linear Hyperbolic Equations	693
7.4.2	Wave Equation and Finite Difference Methods	693
7.4.3	Wave Equation and the Method of Lines	693
7.5	A Look Beyond: Traffic Flow	693
7.6	Review Questions and Problems	693
7.6.1	Review Questions	693
7.6.2	Problems	694
Notation		697
Matlab Codes		699
Bibliography		701
Index		705

Chapter 1

Introduction to Differential Equations

This chapter introduces a number of basic concepts which will be used throughout the book. A partial differential equation is an equation involving an unknown function of two or more variables and some of its partial derivatives. On the surface, this appears to be a slight generalization of an ordinary differential equation, in which the unknown function depends only on a single variable. However, extra variables make quite a difference. We begin by contrasting the theory of existence and uniqueness for ordinary differential equations with the lack of a unified theory for partial differential equations. We then introduce some fundamental examples of partial differential equations, show how they arise in applied sciences such as physics, chemistry, and biology, and describe some of their easily accessible properties. In addition, we present classification schemes for partial differential equations which turn out to be extremely useful in later chapters. We end the chapter with the following topic from the study of ordinary differential equations which is usually not covered in a first year course, namely stochastic differential equations.

1.1 • Ordinary and Partial Differential Equations

In this first section we review some basic facts and terminology for differential equations. We discuss the existence and uniqueness of solutions for ordinary differential equations and the lack of similar results for partial differential equations.

1.1.1 • Ordinary Differential Equations

We now briefly review some essential facts about ordinary differential equations. An *ordinary differential equation* is an equation involving an unknown function $x = x(t)$ of a single variable t and its derivatives. In the following set of examples, we describe in the context of ordinary differential equations the concepts and terminology which will then be used throughout the book in the context of partial differential equations.

Example 1.1 (Exponential growth equation). The *exponential growth equation*

$$\frac{dx}{dt} = kx$$

models growth ($k > 0$) or decay ($k < 0$) of a population for which the rate of change of the population is directly proportional to the current population.

The ultimate goal of ordinary differential equations is to gain information about *solutions*. A solution is a function $x(t)$, which satisfies the differential equation. For example, the constant function $x(t) \equiv 0$ for all values of t solves the exponential growth equation. But the constant function 0 is not the only solution for this equation. For any constant $C \in \mathbb{R}$ the function

$$x(t) = Ce^{kt} \quad \text{for all } t \in \mathbb{R}$$

is a solution as well. This can easily be verified by substituting this formula into the differential equation. Since every C gives rise to a different solution, this differential equation has more than one solution. However, we can specify an *initial condition* in the form $x(0) = x_0$. The function $x(t) = x_0 e^{kt}$ is the unique solution to the exponential growth equation under this additional constraint.

We also point out the following terminology: The exponential growth equation is a *linear* equation in that the dependent variable x and its derivative dx/dt appear only linearly. It is a linear *homogeneous* equation since the constant function $x(t) \equiv 0$ is a solution. It is a *first-order equation*, since the equation only involves the first derivative of x . Furthermore, it is *autonomous*, since the equation does not explicitly depend on the independent variable t . ■

Example 1.2 (The logistic equation). The logistic equation is of the form

$$x' = \mu x(1-x) = \mu x - \mu x^2,$$

where $\mu \in \mathbb{R}$ is a fixed constant. This ordinary differential equation is *nonlinear*, since the dependent variable x appears nonlinearly in the term μx^2 . Similar to the previous example, this equation is a *first-order equation* that is *autonomous*. Notice that we use the alternative notation x' for the derivative, rather than dx/dt . ■

Example 1.3 (A nonlinear equation). A second example of a nonlinear equation is given by

$$\frac{d^2x}{dt^2} + 3x \frac{dx}{dt} - 5 = 0.$$

The nonlinearity of the equation occurs in the term $3x dx/dt$, since the dependent variable and its derivative occur in a product of higher degree than linear. ■

Example 1.4 (A boundary value problem). The equation

$$x'' + x = 0 \quad \text{for } 0 < t < 2\pi, \quad \text{with } x(0) = x(2\pi) = 0$$

has solution $x(t) = \sin t$. This equation is the example of a *boundary value problem*, as values of the function are specified at two ends of an interval. ■

Example 1.5 (A second-order equation). The differential equation

$$x'' + \lambda x + 3 = \sin t, \tag{1.1}$$

where $\lambda \in \mathbb{R}$ is a fixed constant, is a *second-order ordinary differential equation* since the highest derivative in the equation is a second derivative. The equation is *linear*, since

the dependent variable x and its derivatives appear only in linear terms. Notice that this equation is linear despite the fact that the independent variable appears in the nonlinear term $\sin t$ — linearity is only a condition on the dependent variable and its derivatives.

The linear differential equation in this example is called *nonhomogeneous* since the constant function $x(t) \equiv 0$ is not a solution. Equivalently, a linear equation is called nonhomogeneous, if it includes additive terms which do not contain the dependent variable. Note that this second concept of nonhomogeneous is only equivalent to the first for linear differential equations; thus in this text, we only use the terms homogeneous and nonhomogeneous when discussing linear differential equations. ■

Example 1.6 (The method of variation of parameters). Consider the linear differential equation

$$x' = ax + b(t), \quad (1.2)$$

where a is a constant and $b(t)$ is a continuous function. This equation is first-order, linear, and nonhomogeneous. *Variation of parameters* is a general method to solve this equation with the initial value $x(t_0) = x_0$. The variation of parameters formula gives the solution:

$$x(t) = x_0 e^{a(t-t_0)} + \int_{t_0}^t e^{a(t-s)} b(s) ds. \quad (1.3)$$

Note that the first term in this solution is a solution to the homogeneous problem with the given initial condition, and the second term in this solution is a particular solution to the nonhomogeneous equation with initial condition zero. That this is a solution can be verified by differentiation using the fundamental theorem of calculus. ■

Example 1.7 (A first-order system of equations). The following system of two equations is a *coupled system of ordinary differential equations*:

$$\begin{aligned} x' &= y, \\ y' &= -\lambda x - 3 + \sin t, \end{aligned} \quad (1.4)$$

where $\lambda \in \mathbb{R}$ is a fixed constant. In this case, there are two unknown functions $x = x(t)$ and $y = y(t)$ of the single independent variable t . The system is called *first-order*, since only the first derivatives of the unknown functions are involved; it is *linear*, since both x and y , as well as their derivatives, only occur in linear terms; and it is *nonhomogeneous*, since the pair of trivial functions $x(t) \equiv 0$ and $y(t) \equiv 0$ taken together do not satisfy the system.

Notice that there is a relationship between the differential equation in this example and the second-order differential equation from Example 1.5. Namely, starting with the previous example, if we create a new function y by setting $x' = y$, the result is the current first-order system. This relationship demonstrates the important general method of order reduction. ■

Order reduction

For the rest of our discussion of ordinary differential equations, we will restrict our attention to *first-order systems of ordinary differential equations*. This can be done without loss

of generality by employing the following simple trick. Assume that a given ordinary differential equation involves the unknown function $x(t)$ and its derivatives up to order m , i.e., assume that we have an equation of the form

$$\frac{d^m x}{dt^m} = f\left(t, x, \frac{dx}{dt}, \dots, \frac{d^{m-1}x}{dt^{m-1}}\right). \quad (1.5)$$

Then one can introduce m functions $x_k(t)$ for $k = 0, \dots, m - 1$, which satisfy

$$x_k(t) = \frac{d^k x}{dt^k}(t) \quad \text{for } k = 0, \dots, m - 1.$$

That is, we introduce one new variable for each lower-order derivative of x occurring in the original equation. Then one can easily see that the original differential equation is equivalent to the first-order system

$$\begin{aligned} x'_0 &= x_1, \\ x'_1 &= x_2, \\ &\vdots \\ x'_{m-2} &= x_{m-1}, \\ x'_{m-1} &= f(t, x_0, x_1, \dots, x_{m-1}), \end{aligned} \quad (1.6)$$

in the sense that the set of functions (x_0, \dots, x_{m-1}) satisfies (1.6) if and only if the function x_0 satisfies (1.5). The first $m - 1$ equations of this system ensure that the functions x_1, \dots, x_{m-1} correspond to the first $m - 1$ derivatives of the function x_0 . The last equation reformulates the given differential equation in terms of these new functions. This method of replacing a higher-order ordinary differential equation by a first-order system is known as *order reduction*. For example, we have already noted that the second-order equation (1.1) can be equivalently rewritten as the two-dimensional first-order system (1.4). In fact, the system is the one obtained by the above method if we let $x_0 = x$ and $x_1 = y$. One can readily see that the same method allows one to reduce a higher-order system of ordinary differential equations to a first-order system of higher dimensionality.

1.1.2 • Existence and Uniqueness of Solutions

In the case of the exponential growth equation, the equation plus a specified initial condition yield a unique solution which exists for all time. It is natural to wonder if existence and uniqueness of solutions are special properties of the particular equation. That is, do all ordinary differential equations with specified initial values always have solutions? Are the solutions unique? Do solutions exist for all time? In this subsection we address these three questions.

Existence of solutions

We first turn our attention to the general existence theory of systems of first-order ordinary differential equations, which by the method of order reduction applies to arbitrary systems. We start with the following formal definition.

Definition 1.8 (Solution to a first-order system on an interval). Assume that $F : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a given function, and consider the associated d -dimensional system of ordinary

differential equations given by

$$x' = F(t, x). \quad (1.7)$$

Then a function $x : I \rightarrow \mathbb{R}^d$ is called a solution of this system on the interval $I \subset \mathbb{R}$, if for all $t \in I$ the derivative $x'(t)$ exists and $x'(t) = F(t, x(t))$.

Using the notation from the above definition, we have the following fundamental result.

Theorem 1.9 (Peano). *In the situation of Definition 1.8, if the right-hand side F is a continuous function, then there exist solutions to the first-order system (1.7).*

In fact, for any choice of $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}^d$ there exists an interval I which contains t_0 in its interior and a differentiable function $x : I \rightarrow \mathbb{R}^d$ such that x is a solution of (1.7) with $x(t_0) = x_0$. In this case, we say that x satisfies the initial value problem

$$x' = F(t, x) \quad \text{with} \quad x(t_0) = x_0.$$

The above result settles the question of *existence of solutions* in a satisfying way, since the assumption of a continuous right-hand side is usually satisfied. Yet the theorem shows more. Namely, it shows that in general, an ordinary differential equation has infinitely many solutions — which one can see by varying the initial condition x_0 in the above theorem.

Uniqueness of solutions

Peano's theorem does not answer the question of *uniqueness of solutions to initial value problems*. In other words, if in the above result we fix t_0 and x_0 , is there a unique solution of (1.7) with $x(t_0) = x_0$? For this, consider the initial value problem

$$x' = \frac{3}{2} x^{1/3} \quad \text{with} \quad x(0) = 0.$$

The right-hand side of the differential equation is continuous, but following two functions are both solutions to the initial value problem

$$x_1(t) = 0 \quad \text{and} \quad x_2(t) = \begin{cases} 0 & \text{for } t < 0 \\ t^{3/2} & \text{for } t \geq 0 \end{cases}.$$

In fact, one can even show that this initial value problem has an infinite number of solutions.

In light of the above example it is natural to ask whether there are additional conditions on the right-hand side F of an ordinary differential equation that do guarantee uniqueness. The answer in the affirmative is provided by the next result.

Theorem 1.10 (Picard-Lindelöf). *Consider the situation of Definition 1.8. Assume in addition that the right-hand side $F : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is continuously differentiable. Then for any $t_0 \in \mathbb{R}$ and $x_0 \in \mathbb{R}^d$ the initial value problem*

$$x' = F(t, x) \quad \text{with} \quad x(t_0) = x_0$$

has a unique solution on some small time interval $(t_0 - \varepsilon, t_0 + \varepsilon)$ around t_0 , where the interval radius $\varepsilon > 0$ generally depends on both t_0 and x_0 .

While the above result suffices for our purposes, we would like to point out that the condition of continuous differentiability of F can actually be weakened somewhat. In fact, it is enough for F to be continuous and Lipschitz continuous¹ with respect to x .

The Picard-Lindelöf theorem settles both the questions of existence and uniqueness of solutions of first-order systems of ordinary differential equations in a most satisfying way. Once the initial condition has been specified, one only has to verify the smoothness of the right-hand side of the differential equation.

Even though the existence and uniqueness result is formulated only for first-order systems, one can easily extract corresponding statements for higher-order ordinary differential equations using the order reduction technique. To illustrate this, consider the following simple problem.

Example 1.11 (Trajectory of an object due to gravity). If we throw a ball vertically into the air starting at an initial height x_0 , how can its trajectory be described as a function of time? Elementary physics states that the ball's acceleration due to gravity is a constant denoted by g . Thus motion of the ball is governed by the second-order differential equation

$$x'' = -g \quad \text{with} \quad x(0) = x_0,$$

where g denotes the gravitational constant. Since this equation can easily be integrated twice, we see that the solution of this initial value problem is not unique. In fact, it is given by the family of solutions

$$x(t) = -\frac{g}{2} t^2 + v_0 t + x_0, \quad \text{for arbitrary } v_0 \in \mathbb{R},$$

which correspond to different initial velocities v_0 — and which at first look seems to indicate that the existence and uniqueness theorem does not hold for higher-order ordinary differential equations. Notice, however, that using order reduction to convert the second-order differential equation gives the following equivalent first-order system

$$\begin{aligned} x' &= v, \\ v' &= -g. \end{aligned}$$

The right-hand side of this system is clearly continuously differentiable, and we therefore expect unique solutions if we add initial conditions of the form

$$x(0) = x_0 \quad \text{and} \quad v(0) = v_0,$$

which is in agreement with the above explicit formula. ■

Beyond explicit formulas

Having discussed the questions of existence and uniqueness of solutions of ordinary differential equations, we now address the question of actually finding *explicit formulas for the solutions*. Even though this aspect of differential equations is usually the prime topic of an undergraduate differential equations course, only for a small number of classes of differential equations can such explicit formulas be obtained. There are many cases where

¹Lipschitz continuity basically means that there exists a constant K such that for all points x and y we have $|f(x) - f(y)| \leq K|x - y|$. It is a weaker assumption than differentiability, but a stronger assumption than continuity.

it is impossible to find an explicit closed form representation of the solution. Two such examples are the second-order equation

$$x'' = t x,$$

and the even simpler first-order equation

$$x' = e^{-t^2}.$$

One can prove that neither of these equations has a nontrivial closed form solution which is expressible in terms of elementary functions. In fact, most differential equations — both ordinary and partial — cannot be solved in closed form. As a result, a general study of differential equations cannot rely on the existence of explicit formulas. Traditionally, one makes a choice to proceed either by developing sophisticated analytical tools for studying the qualitative behavior of solutions, or by relying on numerical computations. In this book we take the point of view that both analytical and numerical methods are extremely valuable tools in the study of differential equations, and neither should be presented without the other. Our goal is to introduce techniques from analysis in the context of numerics, and techniques of numerics in the context of analysis, thus providing a unified method of approach.

A summary of the theory of ordinary differential equations

Before turning our attention to some first examples of partial differential equations, we briefly summarize our discussion of ordinary differential equations. An ordinary differential equation relates a function to some of its derivatives. Basic questions of existence and uniqueness of solutions have been solved in a satisfying way, and there is a well-established theory of ordinary differential equations. However, even in simple cases, it is usually not possible to write down an explicit solution to an ordinary differential equation. Therefore we need to rely on a combination of analysis and numerical techniques to study equations of this type. Finally, we have introduced the following useful terminology:

- **Order of an ordinary differential equation:** The order of the highest derivative.
- **Linear ordinary differential equation:** The dependent variable and derivatives of the dependent variable occur only linearly in the first degree.
- **Homogeneous linear ordinary differential equation:** A linear ordinary differential equation in which the function which is constant zero is a solution. Equivalently, a linear equation in which each term contains the dependent variable or a derivative of the dependent variable. A linear equation which is not homogeneous is called nonhomogeneous.
- **Dimension of an ordinary differential equation:** The number of dependent variables in an ordinary differential equation.

Each of these terms are also used in the study of partial differential equations.

1.1.3 • A First Glimpse of Partial Differential Equations

In this section, we introduce partial differential equations. In addition, we compare partial differential equations to ordinary differential equations, commenting in particular on the questions of existence and uniqueness of solutions and explicit solution formulas.

Definition 1.12 (Partial differential equation). A partial differential equation is an equation involving an unknown function $u = u(x, y, \dots)$ of two or more variables and some of its partial derivatives. A solution $u = u(x, y, \dots)$ of such a partial differential equation is a sufficiently differentiable function that satisfies the equation for all (x, y, \dots) in its domain, and which can be extended continuously to the boundary of the domain.

While in the case of ordinary differential equations we could use the shorthand u' of a derivative, we now have to distinguish partial derivatives such as $\partial u / \partial x$ or $\partial u / \partial y$ for a function $u = u(x, y)$. Since this notation can quickly become cumbersome, one generally uses abbreviations via subscripts. For example, for the function $u = u(x, y, z)$ we usually write

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

$$(u_x)_z = u_{xz} = \frac{\partial^2 u}{\partial z \partial x} = \frac{\partial}{\partial z} \left(\frac{\partial u}{\partial x} \right), \quad \text{etc.}$$

Throughout this book, we will generally use the above subscript notation for partial derivatives. However, there will be exceptions whenever another notation is more appropriate. (One of these exceptions occurs later in this chapter in the form of the multi-index notation.) Notice also that the notion of solution introduced above dictates that as points in the domain approach the boundary of the domain, the function values of the solution have to converge. We will discuss domains and their boundaries in more detail below.

On a formal level, the step from ordinary to partial differential equations seems to be just the step from functions of one variable to functions of several variables. However, as will be demonstrated throughout the book, the differences between the two topics are profound. Before addressing this point more, we present a few standard examples of partial differential equations which have their origin in physical examples. These examples serve as prototypes, in that each exhibits properties which generalize to a whole class of partial differential equations.

1. **The heat equation:** The *heat equation* or *diffusion equation* in one spatial dimension is a partial differential equation for an unknown function $u = u(t, x)$ given by

$$u_t = k u_{xx},$$

where $k > 0$ is a fixed constant. This equation is the prototype of a *parabolic equation*, and describes the temporal evolution of heat as a function of location, as well as simple diffusion processes. Notice that in this partial differential equation, the independent variables are denoted by t and x . This is due to the interpretation of these variables as a time and a spatial location.

2. **Laplace's equation:** The standard example of an *elliptic equation* is the *Laplace equation*, which is given by

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

One can easily see that affine functions of the form $u(x, y) = ax + by + c$ solve the Laplace equation. But there are also more complicated nonlinear solutions such as $u(x, y) = x^2 - y^2$ and even $u(x, y) = \ln \sqrt{x^2 + y^2}$.

3. **The transport equation:** The following equation is a first-order homogeneous linear partial differential equation in two variables

$$u_t + a(t, x)u_x = 0,$$

where $a(t, x)$ denotes is a fixed known function. The equation is also referred to as the *advection equation*. In the case when $a \in \mathbb{R}$ is a constant, this is called the *simple transport equation*. Solutions of this equation are of the form $u = u(t, x)$, i.e., they depend on time t and a spatial location x . One can easily verify that for any differentiable function $w : \mathbb{R} \rightarrow \mathbb{R}$ the function

$$u(t, x) = w(x - at)$$

is a solution of the simple transport equation. The transport equation is a first-order example of a *hyperbolic equation*.

4. **The wave equation:** Finally, the one-dimensional *wave equation*

$$u_{tt} = c^2 u_{xx}$$

for a fixed constant $c \in \mathbb{R}$ and an unknown function $u = u(t, x)$ is a model for a vibrating string and other oscillatory phenomena. It serves as a prototype of a second-order *hyperbolic equation*.

Using the terminology developed for ordinary differential equations, all of these equations are linear since the unknown function u and its derivatives appear only linearly. They are all homogeneous, since the constant function 0 solves each equation. They are one-dimensional, since each is a single equation with a single unknown function. The transport equation is first-order, since only first derivatives appear. The rest of the equations are second-order, since each contains a second-order partial derivative u_{tt} or u_{xx} of the function u . (We note that the mixed derivative u_{xt} is also considered to be a second-order partial derivative.) For example, consider the three following equations for $u(t, x)$

$$\begin{aligned} u_{ttx} - \sin(tx) u_{xx} &= 0, \\ u_{ttx} - \sin(tx) u_{xx} &= t^2 + x, \\ u u_{ttx} - \sin(tx) u_{xx} &= 0. \end{aligned}$$

The first is a third-order linear homogeneous equation; namely u_{ttx} is third-order, and the function u and its derivatives appear only linearly (both x and t are independent variables for the equation). The second equation is a third-order linear nonhomogeneous equation, since $u \equiv 0$ will not solve the equation: If $u \equiv 0$ the left-hand side of the equation is zero whereas the right-hand side is nonzero for all t and x values such that $t^2 + x \neq 0$. The third equation is a third-order nonlinear equation since $u u_{ttx}$ is a nonlinear term with respect to u and its derivatives. These concepts are described more formally in Section 1.3.

It is natural to wonder whether any of the basic questions of existence, uniqueness, and computability of solutions can be addressed in a satisfactory way for partial differential equations — and the answer is unfortunately a resounding “No!” for all three questions. The infinitely-many solutions given for Laplace’s equation and the transport equation demonstrate that without additional constraints, partial differential equations in general have infinitely many solutions. Recall that this was already the case with ordinary differential equations. In the case of ordinary differential equations one could achieve unique solutions through the specification of an initial value. For partial differential equations, since there is more than one variable, it is not surprising that an initial condition is insufficient to guarantee uniqueness. In many cases, we can recover existence and uniqueness results by specifying not only an initial condition but also by specifying *boundary conditions*. However, there are partial differential equations for which uniqueness and even existence are not guaranteed by the specification of such conditions. In fact, in 1957 the

mathematician Hans Lewy gave an example showing that solutions can fail to exist even for a linear partial differential equation [48]. Furthermore, there is no theorem stating the exact criteria for existence or uniqueness of solutions for a partial differential equation. For example, the Navier-Stokes equation is the standard partial differential equation model used to describe fluid flow; settling the question of whether solutions to the Navier-Stokes equation always exist is currently prized at one million dollars, courtesy of the Clay Mathematics Institute [12]. In terms of the computability of solutions — in cases in which unique solutions are known to exist — both the existence of closed form solutions and techniques for numerical approximation of solutions are far more delicate for partial differential equations than for ordinary differential equations.

As a consequence of the above comments, there is no unified theory of existence and uniqueness for partial differential equations like there is for ordinary differential equations. Furthermore, even when partial differential equations have solutions, the behavior and properties of these solutions can be radically different. However, it has been recognized over time that partial differential equations can be roughly divided into three different categories called *elliptic*, *parabolic*, and *hyperbolic equations* such that equations within each of these categories share similar characteristic features and can be treated with similar techniques. Our above examples provide basic examples for each of these categories. In the next section, we derive these equations, and establish basic properties of solutions. We will then give a more detailed description of the classification of partial differential equations. Common to all three categories is the fact that for most examples, there are no closed form solutions. The best one can hope for is to have a set of both analytical and numerical techniques at hand which allow one to work towards understanding a given model.

1.1.4 • Domains and Their Boundaries

In order to specify a unique solution $x(t)$ for an ordinary differential equation, it is necessary to combine the equation with an auxiliary condition, namely the initial value condition $x(t_0) = x_0 \in \mathbb{R}^d$. (See Section 1.1.2.) That is, at a single time, a single point is specified. In contrast, in order to specify a solution for a partial differential equation, it is necessary to append the equation with an entire curve or surface of values for the function or for its derivative.² In particular, these values are generally given along the boundary of the domain. In this subsection, we discuss different types of functional domains in dimension higher than one, we describe what is meant by the boundary of a domain, and we explain what it means to evaluate a multivariable function or its derivative on the boundary of a domain. The discussion requires the use of topological and geometric concepts. A number of the concepts will be familiar, as they are standard topics covered in a multivariable calculus course.

Definition 1.13 (Domain of a function). Let $F : \Omega \rightarrow \mathbb{R}^k$ be a function defined on $\Omega \subset \mathbb{R}^d$. The set Ω is defined as the domain of F . In this text, we usually consider only domains which are finite unions of disjoint open connected sets.

Recall that open sets are defined as follows. If $B_r(p) \subset \mathbb{R}^d$ denotes the open ball of points strictly less than distance r from the point p , then a set $\Omega \subset \mathbb{R}^d$ is open, if for every point $p \in \Omega$ we can find a ball $B_r(p)$ with $r > 0$ which is completely contained in Ω . Intuitively this means that every point $p \in \Omega$ lies in the interior of the set, i.e., there

²This condition is a necessary but in general not sufficient condition. That is, the solution may still not exist or may not be unique.

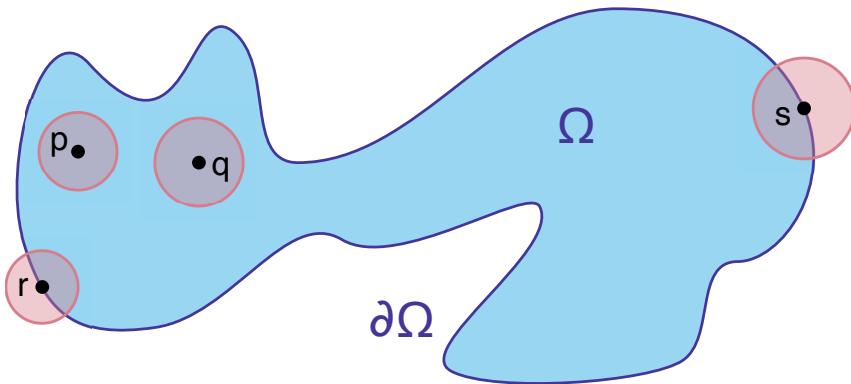


Figure 1.1. Interior and boundary points of a domain. The points p and q lie in the interior of Ω , since there are disks centered at these points which lie completely in Ω . In contrast, the points r and s are boundary points, since every ball centered at either point contains both points in Ω and also points which are not in Ω .

are no points nearby which are not in the set Ω .

We now turn our attention to a discussion of the boundary of domains, and what we generally assume for the behavior of the function F on this boundary. If Ω is an open one-dimensional interval (a, b) , then the boundary of the domain consists of the two endpoints a and b of the interval. We will generally assume that the function F is at least differentiable on (a, b) , and either the function values or the derivative values exist at the boundary points. That is, either $F(a)$ or the right derivative $F'(a)$ exists, and either $F(b)$ or the left derivative $F'(b)$ exists. According to our definition of a solution of a partial differential equation, continuity is assumed at the boundary of the domain, meaning that when $F(a)$ is given, the right limit of $F(x)$ as x approaches a is $F(a)$, and similarly with a left limit at b when $F(b)$ is given. On the other hand, if the derivative value is known, then the function has to be one-sided continuous, but the derivative may not be continuous at that boundary point. This completes the discussion of function evaluation on domain boundaries in one dimension, since in this text we only consider one-dimensional domains Ω which are finite unions of disjoint open intervals.

We now consider domains in \mathbb{R}^d where $d > 1$. The boundary of a domain is an intuitive concept, formalized in the following definition. Note that this definition in fact also formalizes our notion of boundary in one dimension.

Definition 1.14 (Boundary of a domain). Let Ω be a subset of \mathbb{R}^d . A point p is a boundary point of Ω , if for every value of r , the open ball $B_r(p)$ contains both points in Ω and points that are not in Ω . See Figure 1.1. The set of all boundary points of Ω is called the boundary of Ω , written $\partial\Omega$.

Definition 1.15 (Closure of a domain). Let Ω be a subset of \mathbb{R}^d . Then the closure $\overline{\Omega}$ of the domain is defined as the union of Ω with the boundary $\partial\Omega$. It is in fact the smallest closed subset of \mathbb{R}^d which contains Ω .

Recall that a subset $S \subset \mathbb{R}^d$ is closed, if its complement $\mathbb{R}^d \setminus S$ is an open subset of \mathbb{R}^d . To illustrate the definition of the boundary, we present a couple of simple sets and their boundaries, one each in \mathbb{R}^2 and in \mathbb{R}^3 .

Example 1.16 (Boundaries of disks and balls). For any real number $a > 0$, the open disk

$$\Omega_{disk} = \{x \in \mathbb{R}^2 : \|x\| = \sqrt{x_1^2 + x_2^2} < a\}$$

has a circle as its boundary, given by

$$\partial\Omega_{disk} = \{x \in \mathbb{R}^2 : \|x\| = a\}.$$

Correspondingly, in three dimensions the solid open ball

$$\Omega_{ball} = \{x \in \mathbb{R}^3 : \|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2} < a\}$$

has a spherical surface as its boundary, namely

$$\partial\Omega_{ball} = \{x \in \mathbb{R}^3 : \|x\| = a\}.$$

We will return to this example several times to introduce additional concepts related to the boundary, which are essential for our study of partial differential equations. ■

In the above example, the boundary for the two-dimensional region is a one-dimensional set, and the boundary for the region in \mathbb{R}^3 is a two-dimensional set. Furthermore, these boundaries are the graph of a continuously differentiable curve and a continuously differentiable surface, respectively. Such boundaries are called continuously differentiable boundaries. They allow for the definition of outward unit normal vectors. These concepts will now be formalized.

Definition 1.17 (Continuously differentiable surface). A subset $\mathcal{S} \subset \mathbb{R}^d$ is called a continuously differentiable surface, if it can be written as the graph of a continuously differentiable function which expresses one of the coordinates of the points in \mathcal{S} as a function of the remaining $d - 1$ coordinates.

More precisely, the set \mathcal{S} is a continuously differentiable surface, if there exists an integer $k \in \{1, \dots, d\}$ as well as a continuously differentiable function T defined on a subset of \mathbb{R}^{d-1} with nonempty interior, such that

$$x = (x_1, \dots, x_d) \in \mathcal{S} \quad \text{if and only if} \quad x_k = T(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_d).$$

In general, each portion of a given surface \mathcal{S} can be written using several different representations in the above form, i.e., several possible choices of k , but in order to write the entire surface \mathcal{S} , it is necessary to patch a series of different functions together. This is illustrated in the example below. Furthermore, in some cases we need the function T to be more than once continuously differentiable. In these cases we say that $\mathcal{S} \subset \mathbb{R}^d$ is an m -times continuously differentiable surface, for some $m \geq 2$.

Example 1.18. The boundary of the open disk Ω_{disk} from Example 1.16 is continuously differentiable. To illustrate this, let $p = (p_1, p_2)$ be any point in $\partial\Omega_{disk}$ with $p_2 > 0$. Then for $x = (x_1, x_2)$ in a neighborhood of p the boundary is the graph of the function

$$x_2 = T_1(x_1) = \sqrt{a^2 - x_1^2},$$

which is differentiable near p_1 due to $p_2 > 0$. The remaining boundary points can be treated similarly, using the following differentiable functions T in a neighborhood of a

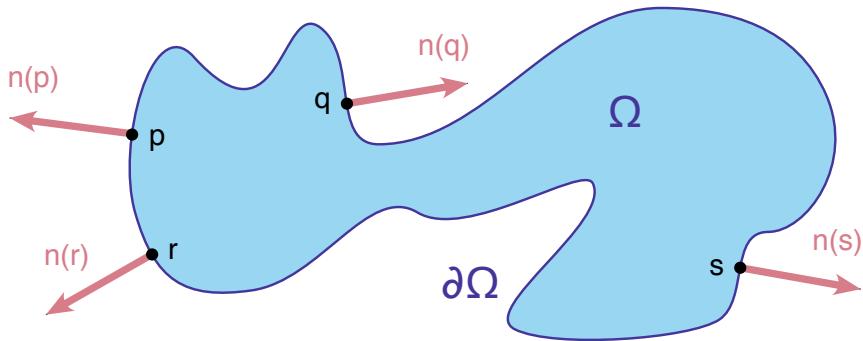


Figure 1.2. Sample outward unit normal vectors. For the four boundary points p, q, r , and s the diagram shows their associated outward unit normal vectors $n(p), n(q), n(r)$, and $n(s)$, all of which are orthogonal to the boundary $\partial\Omega$ at the respective base points and point into the complement of the domain.

boundary point p :

$$\begin{aligned} \text{For } p_2 < 0 &\quad \text{use } x_2 = T_2(x_1) = -\sqrt{a^2 - x_1^2}, \\ \text{for } p_1 > 0 &\quad \text{use } x_1 = T_3(x_2) = +\sqrt{a^2 - x_2^2}, \\ \text{for } p_1 < 0 &\quad \text{use } x_1 = T_4(x_2) = -\sqrt{a^2 - x_2^2}. \end{aligned}$$

These patches taken together cover the entire boundary circle. ■

Definition 1.19 (Outward unit normal vector). Let $\Omega \subset \mathbb{R}^d$ be an open region with boundary $\partial\Omega$. Assume that $p \in \partial\Omega$ is a boundary point such that a neighborhood of p in $\partial\Omega$ is given by a continuously differentiable surface $\mathcal{S} \subset \partial\Omega$ in the sense of Definition 1.17. Then the vector n is called the outward unit normal vector to the boundary $\partial\Omega$ at the point p , if n is a vector of unit length which is orthogonal to the surface \mathcal{S} at p , and which points away from Ω at p . Specifically, for all sufficiently small positive values of δ , the point $p + \delta n$ is not contained in Ω . See Figure 1.2.

Finding the outward unit normal vector is straightforward, if one has an explicit formula for the surface \mathcal{S} around p . For this, assume that \mathcal{S} is given as the graph of a function T as in Definition 1.17, and define the vector

$$v = \left(\frac{\partial T}{\partial x_1}, \dots, \frac{\partial T}{\partial x_{k-1}}, -1, \frac{\partial T}{\partial x_{k+1}}, \dots, \frac{\partial T}{\partial x_d} \right),$$

where all derivatives are evaluated at $(p_1, \dots, p_{k-1}, p_{k+1}, \dots, p_d)$. Then the outward unit normal vector to $\partial\Omega$ at p is given by either $v/\|v\|$ or by $-v/\|v\|$, depending on which of these two vectors points outward from the domain.

Remark 1.20 (Notation for points and vectors). In first-year undergraduate calculus courses, one generally uses different notation to distinguish between points in Euclidean space \mathbb{R}^d and d -dimensional vectors. This distinction is usually removed in higher-level mathematics courses, and the definition $x = (x_1, x_2, x_3)$ can represent either a point or a vector, based on context. Throughout this text, we will adhere to this more advanced notation.

In the context of vectors, however, their orientation is essential for matrix operations. For example, if x denotes a row vector and y a column vector of the same dimension d , then the matrix product xy results in a scalar, while the product yx gives a $d \times d$ -matrix. For this reason, whenever we talk about a vector x , we always implicitly assume that it is a column vector. Clearly, there are instances when a vector should be viewed as a row vector, and we will state this explicitly in these cases.

Unfortunately, we will be encountering many concrete vectors in the remainder of this book, and if every one of these vectors is typeset as a column, the text would become considerably harder to read. We therefore typically declare vectors in the form $x = (x_1, \dots, x_d)$, and implicitly remember that x is actually a column vector. This allows us for example to define the (column) vectors $x = (1, 2, 3)$ and $y = (1, 0, 1)$, and then write matrix multiplication identities such as

$$x^t y = 4 \quad \text{and} \quad xy^t = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 3 & 0 & 3 \end{pmatrix},$$

where the superscript t denotes the transpose of a matrix. In some cases, however, in order to state the orientation clearly, we will make explicit use of the transpose notation.

The concept of outward unit normal vector will be illustrated in the following continuation of Example 1.16.

Example 1.21. The boundary of the open disk Ω_{disk} from Example 1.16 has the property that every point p admits an outward unit normal vector. Let $p = (p_1, p_2)$ be any point in $\partial\Omega_{disk}$ with $p_2 > 0$, and express the boundary using T_1 in Example 1.18. The vector v above is given by

$$v = \left(\frac{\partial T_1}{\partial x_1}(p_1), -1 \right) = \left(\frac{-p_1}{\sqrt{a^2 - p_1^2}}, -1 \right), \quad \text{with length } \|v\| = \frac{a}{\sqrt{a^2 - p_1^2}}.$$

Direct inspection then shows that at the point p the outward unit normal vector is

$$n = -\frac{v}{\|v\|} = -\left(\frac{-p_1}{a}, \frac{-\sqrt{a^2 - p_1^2}}{a} \right) = \left(\frac{p_1}{a}, \frac{p_2}{a} \right).$$

This formula gives the outward unit normal vectors for all points on the open upper semicircle of $\partial\Omega_{disk}$. We leave the cases when the boundary is expressed in terms of T_2 , T_3 , or T_4 as easy exercises for the reader. The calculation is similar for the outward normal vectors for the boundary of the three-dimensional ball. ■

While the above definition of a continuously differentiable surface and its associated outward unit normal vectors suffices in many situations, there are cases in which the boundary cannot easily be expressed as the graph of a function T . In these cases, we need alternative methods for determining the outward unit normal vectors. One such approach is to parametrize the surface, as described in the remark below. This uses the Jacobian derivative matrix, defined as follows.

Definition 1.22 (Jacobian derivative matrix). Let φ be a continuously differentiable function $\varphi : D \subset \mathbb{R}^k \rightarrow \mathbb{R}^d$. For a point $z \in D$ the Jacobian derivative matrix of φ =

$(\varphi_1, \dots, \varphi_d)$ at z is given by

$$D\varphi(z) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(z) & \dots & \frac{\partial \varphi_1}{\partial x_k}(z) \\ \vdots & & \vdots \\ \frac{\partial \varphi_d}{\partial x_1}(z) & \dots & \frac{\partial \varphi_d}{\partial x_k}(z) \end{pmatrix} \in \mathbb{R}^{d \times k}.$$

Remark 1.23 (Parameterized surfaces). Rather than being given in the form presented in Definition 1.17, sometimes it is more convenient to use the following equivalent definition.

A subset $\mathcal{S} \subset \mathbb{R}^d$ is a continuously differentiable surface, if it is the image

$$\mathcal{S} = \varphi(D)$$

of a set $D \subset \mathbb{R}^{d-1}$ under a continuously differentiable function $\varphi : D \rightarrow \mathbb{R}^d$, where for each point $z \in D$ the Jacobian matrix $D\varphi(z) \in \mathbb{R}^{d \times (d-1)}$ has linearly independent column vectors. Also in this definition we assume that D has nonempty interior.

If some part of the boundary of a region Ω is parameterized by the function φ , then at a point $p = \varphi(z) \in \partial\Omega$, the outward unit normal vector is orthogonal to all the column vectors in the Jacobian matrix $D\varphi(z)$.

The following example shows how the outward unit normal vector can be determined for parameterized surfaces in two and three dimensions.

Example 1.24. We continue our discussion of the boundaries of disks and balls, but now in the context of parameterized surfaces. For the case of the planar disk Ω_{disk} of radius a , one can easily see that its boundary is given as the image of the interval $D = [0, 2\pi)$ under the traditional polar parametrization of the circle in the form

$$\varphi(z) = (a \cos z, a \sin z).$$

In this case, the Jacobian matrix of φ at a point z is given by

$$D\varphi(z) = \begin{pmatrix} -a \sin z \\ a \cos z \end{pmatrix} \quad \text{for } z \in [0, 2\pi).$$

Now consider a point $p \in \partial\Omega_{disk}$ of the form $p = \varphi(z) = (a \cos z, a \sin z)$. Then the unit outward normal vector has to be orthogonal to the vector $D\varphi(z)$, i.e., it has to be a multiple of the vector $(a \cos z, a \sin z)$. Normalizing this vector and verifying that the resulting vector points away from Ω_{disk} at p gives

$$n = (\cos z, \sin z),$$

in agreement with the previously discussed approach for finding n .

We now turn our attention to the ball Ω_{ball} of radius a in three dimensions. In this case one can use spherical coordinates to describe the boundary $\partial\Omega_{ball}$ as the image of the two-dimensional domain $D = [0, 2\pi) \times [0, \pi]$ under the parameterization

$$\varphi(z) = (a \cos z_1 \sin z_2, a \sin z_1 \sin z_2, a \cos z_2).$$

Now consider a point $p \in \partial\Omega_{ball}$ of the form $p = \varphi(z)$. The Jacobian of φ at z is of the form

$$D\varphi(z) = \begin{pmatrix} -a \sin z_1 \sin z_2 & a \cos z_1 \cos z_2 \\ a \cos z_1 \sin z_2 & a \sin z_1 \cos z_2 \\ 0 & -a \sin z_2 \end{pmatrix} \in \mathbb{R}^{3 \times 2}.$$

In order to determine the outward unit normal vector, we need to find a vector which is orthogonal to both columns of this Jacobian matrix. The standard method for finding such a vector in three dimensions is to use the cross product. Thus let

$$\begin{aligned} v &= \frac{\partial \varphi}{\partial z_1}(z) \times \frac{\partial \varphi}{\partial z_2}(z) \\ &= (-a^2 \cos z_1 \sin^2 z_2, -a^2 \sin z_1 \sin^2 z_2, -a^2 \cos z_2 \sin z_2), \end{aligned}$$

then normalization furnishes

$$\frac{v}{\|v\|} = (-\cos z_1 \sin z_2, -\sin z_1 \sin z_2, -\cos z_2).$$

Notice that this vector points towards the origin in \mathbb{R}^3 . Therefore, if we define

$$n = -\frac{v}{\|v\|} = (\cos z_1 \sin z_2, \sin z_1 \sin z_2, \cos z_2),$$

then n is a unit vector which is orthogonal to the columns of $D\varphi(z)$ and pointing away from the origin at p — and therefore it is an outward unit normal vector for the boundary of the ball Ω_{ball} . ■

Note that in contrast to the method based on Definition 1.17, the parameterization method allows one to use one single parameterization which works at every point on the boundary, while in the situation of Definition 1.17 we needed four different functions T in the case of the planar disk.

Finally we present a third method for determining the outward unit normal vector to the boundary of a domain Ω . Whenever applicable, this method is the most convenient of the three. It relies on the following definition and results from multivariable calculus.

Definition 1.25 (Gradient of a function). *The gradient of $H : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ at the point $p \in D$ is the row vector*

$$\nabla H(p) = \left(\frac{\partial H}{\partial x_1}(p), \dots, \frac{\partial H}{\partial x_d}(p) \right).$$

Notice that despite our convention discussed in Remark 1.20, the gradient is always considered a row vector, in order to be consistent with Definition 1.22 of the Jacobian matrix.

Theorem 1.26 (A level surface). *A subset $\mathcal{S} \subset \mathbb{R}^d$ is a continuously differentiable surface whenever there exists a continuously differentiable function $H : D \rightarrow \mathbb{R}$ which satisfies*

$$\nabla H(x) \neq 0 \quad \text{for all } x \in D,$$

and for some constant $c \in \mathbb{R}$, \mathcal{S} satisfies the level set equation

$$\mathcal{S} = \{x \in D : H(x) = c\}.$$

In this definition we assume that $D \subset \mathbb{R}^d$ has nonempty interior. Such a surface is called a level surface.

Theorem 1.27 (Gradient and level set). Let $H : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function. Let $c \in \mathbb{R}$ and $p \in D$ be such that $H(p) = c$. Then at the point p , the vector $\nabla H(p)$ is orthogonal to the level set $\{x \in D : H(x) = c\}$.

Combining these two results gives a method of calculating the outward normal vector n to $\partial\Omega$ whenever $\partial\Omega$ is level surface: For each $p \in \partial\Omega$, the outward normal vector $n(p)$ is a scalar multiple of $\nabla H(p)$. The following example demonstrates this method.

Example 1.28. In the situation of Example 1.16, we can use the previous remark if we write the circular boundary of the disk in the form

$$\partial\Omega_{disk} = \{(x_1, x_2) \in \mathbb{R}^2 : H(x_1, x_2) = x_1^2 + x_2^2 = a^2\}.$$

At a point $p \in \partial\Omega_{disk}$, the vector $\nabla H(p) = (2p_1, 2p_2)$ is orthogonal to the level set $\partial\Omega_{disk} = \{x : H(x) = a^2\}$. One can then easily verify (for example by graphing) that this vector is pointing outward from the disk rather than inward. Normalizing, this furnishes the outward unit normal vector

$$n = \frac{(p_1, p_2)}{\sqrt{p_1^2 + p_2^2}} = \left(\frac{p_1}{a}, \frac{p_2}{a} \right),$$

which agrees with our previous two approaches. The same method can easily be applied to the three-dimensional ball Ω_{ball} , since

$$\partial\Omega_{ball} = \{x \in \mathbb{R}^3 : H(x) = x_1^2 + x_2^2 + x_3^2 = a^2\}.$$

Therefore at a point p on the $\partial\Omega_{ball}$, a normal vector is in the direction $\nabla H(p) = (2p_1, 2p_2, 2p_3)$, and again we verify separately that this vector is pointing outward from the solid ball. Normalizing finally gives

$$n = \frac{p}{\sqrt{p_1^2 + p_2^2 + p_3^2}} = \left(\frac{p_1}{a}, \frac{p_2}{a}, \frac{p_3}{a} \right),$$

as before. ■

So far our example domains had outward unit normal vectors at all points on the boundary. This does not always have to be the case. The next example describes some boundaries which are not continuously differentiable, but are almost so.

Example 1.29 (Boundaries of rectilinear regions). Let a, b, c, d be four arbitrary real numbers with $a < b$ and $c < d$, and consider the open two-dimensional rectilinear set

$$\Omega_R = \{(x_1, x_2) \in \mathbb{R}^2 : a < x_1 < b, c < x_2 < d\}.$$

This set has a rectangular curve as its boundary. Namely, $\partial\Omega_R$ is the rectangle connecting the corners (a, c) , (a, d) , (b, d) , and (b, c) . This boundary is the graph of four continuously differentiable curves each connecting a pair of corners. Two such curves meet with non-zero angle at each corner. However, the fact that the boundary has corners means that it cannot be written as the graph of a single continuously differentiable curve.

At each point along each of the four boundary edges, there is a well-defined outward unit normal. In fact, since each fixed edge is a line, every point along that edge has the same outward normal vector. Specifically, for the vertical edges, the outward normal vector is a horizontal vector pointing away from the domain. For the horizontal edges, the outward normal vector is a vertical vector pointing away from the domain. The unit outward normal vector is not defined at the four corners of the boundary rectangle.

This example can easily be generalized to higher dimension. For example, in three dimensions, the open rectilinear set

$$\Omega_B = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : a < x_1 < b, c < x_2 < d, e < x_3 < f\}$$

has a rectilinear surface as its boundary; namely $\partial\Omega_B$ is the rectilinear surface with the eight corners $(a, c, e), (b, c, e), (a, d, e), (b, d, e)$, as well as $(a, c, f), (b, c, f), (a, d, f)$, and (b, d, f) . This boundary is the graph of six continuously differentiable rectilinear surfaces. At each edge two such surfaces meet at non-zero angle, and at each corner three such surfaces meet at non-zero angle. However, the fact that there are corners and edges means that the boundary cannot be written as the graph of a single continuously differentiable surface.

For each fixed face of the rectilinear boundary, since the face is flat, every point on the face has the same outward unit normal. For example, every point on the top face has a vertical outward unit normal vector pointing upwards, whereas every point on the bottom face has a vertical unit outward normal pointing downwards. The unit outward normal is undefined along the twelve edges and at the eight vertices of the boundary. ■

The above examples are special cases of piecewise differentiable boundaries, which will now be introduced.

Definition 1.30 (Piecewise differentiable boundaries). Let $\Omega \subset \mathbb{R}^d$ be an open connected region with boundary $\partial\Omega$. We say that Ω has a piecewise differentiable boundary, if there exists a subset $\mathcal{E} \subset \partial\Omega$ which is the union of finitely many $(d-2)$ -dimensional surfaces in $\partial\Omega$, such that at every point $p \in \partial\Omega \setminus \mathcal{E}$ an outward unit normal vector to the boundary $\partial\Omega$ exists in the sense of Definition 1.19.

More precisely, if $d = 2$ we assume that \mathcal{E} is a finite collection of points, and for $d = 3$ we assume that \mathcal{E} is a finite collection of continuously differentiable curves together with their intersection points.

Thus far all examples are bounded sets. However, the discussion also applies to unbounded regions. The following examples are unbounded regions with piecewise differentiable boundaries.

Example 1.31 (Regions of infinite extent). The following are a few examples of boundaries for regions which extend infinitely. In order to get the reader comfortable with the different notations for vectors that will be used throughout the book, for the purposes of the present example, the letter x always denotes a point in \mathbb{R}^d , for various values of d , and is given in components as $x = (x_1, \dots, x_d)$.

- For every d the open region \mathbb{R}^d has an empty boundary.
- The open region $\{x \in \mathbb{R}^2 : a < x_2 < b\}$ has as its boundary the two horizontal lines $\{x_2 = a\}$ and $\{x_2 = b\}$. The boundary is not connected, but it is continuously differentiable, since the boundary consists of two smooth curves that never meet. For each boundary point on the bottom line, the unit outward normal vector is

a downward pointing vertical vector. On the top line, the unit outward normal vector is an upward pointing vertical vector.

- The open region $\{x \in \mathbb{R}^3 : a < x_2 < b, c < x_3 < d\}$ has a boundary consisting of the four surfaces $\{x_2 = a \text{ and } c \leq x_3 \leq d\}$, $\{x_2 = b \text{ and } c \leq x_3 \leq d\}$, as well as $\{x_3 = c \text{ and } a \leq x_2 \leq b\}$, and $\{x_3 = d \text{ and } a \leq x_2 \leq b\}$. This boundary is piecewise differentiable. Pairs of the four flat horizontal strips meet at edges. Within each of the strips, there is a well-defined unit outward normal vector, but the unit outward normal vector is undefined on the edges joining the strips.
- To add further complexity to the second example, we now consider the unbounded open region $\{x \in \mathbb{R}^2 : x_1 > 0, a < x_2 < b\}$. It has a boundary consisting of three curves: The two horizontal rays $\{x_2 = a \text{ and } x_1 \geq 0\}$ and $\{x_2 = b \text{ and } x_1 \geq 0\}$ which are unbounded, and the bounded vertical line segment $\{x_1 = 0 \text{ and } a \leq x_2 \leq b\}$. Also this boundary is piecewise differentiable. Along each ray and the line segment, there is a well-defined unit outward normal vector, whereas the vector is undefined at the two corners where rays meet the line segment.
- Adding complexity to the third example, we finally consider the unbounded open region $\{x \in \mathbb{R}^3 : x_1 > 0, a < x_2 < b, c < x_3 < d\}$. This region has a boundary consisting of five surfaces. These are given by the four unbounded horizontal strips

$$\begin{aligned} &\{x_1 \geq 0 \text{ and } x_2 = a \text{ and } c \leq x_3 \leq d\}, \\ &\{x_1 \geq 0 \text{ and } x_2 = b \text{ and } c \leq x_3 \leq d\}, \\ &\{x_1 \geq 0 \text{ and } x_3 = c \text{ and } a \leq x_2 \leq b\}, \text{ and} \\ &\{x_1 \geq 0 \text{ and } x_3 = d \text{ and } a \leq x_2 \leq b\}, \end{aligned}$$

along with the rectilinear surface $\{x_1 = 0, a \leq x_2 \leq b \text{ and } c \leq x_3 \leq d\}$. The boundary is again piecewise differentiable, with a well-defined unit outward normal vector only on each of the five surfaces, but not on the edges or vertices where the surfaces meet.

We would like to point out that especially the last two examples will be important for our discussion of the heat and the wave equation. ■

We are now able to continue our discussion of functions defined on a domain with boundary. Assume that F is a continuously differentiable function defined on an open connected set Ω with a piecewise differentiable boundary. For F to be a solution of a partial differential equation, we will only require it to solve the equation on Ω . However, at points of the boundary of the domain we will require that the function values are prescribed, that the value of the directional derivative $\partial F / \partial n$ (given below) is prescribed, or that a combination of the function and derivative values is prescribed.

We now show how to compute the directional derivative $\partial F / \partial n$ in the direction of the outward unit normal. This is used in the next section in the computation of certain boundary conditions for higher-dimensional equations. Let $F : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function. The partial derivative with respect to the variable x_k of F at a point p is the rate of change of F in the direction x_k . In fact, it is possible to measure the rate of change of F at p in any direction. Recall that the dot product of two vectors $v = (v_1, \dots, v_d)$ and $w = (w_1, \dots, w_d) \in \mathbb{R}^d$ is defined as $v \cdot w = v_1 w_1 + \dots + v_d w_d$. Using the dot product, the directional derivative is defined as follows.

Definition 1.32 (Directional derivative). Let $F : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function, and let $v \in \mathbb{R}^d$ be a vector of unit length. Then the directional derivative $\partial F / \partial v(p)$ of F at the point $p \in D$ in the direction v is defined as

$$\frac{\partial F}{\partial v}(p) = \lim_{h \rightarrow 0} \frac{F(p + hv) - F(p)}{h} = \nabla F(p) \cdot v,$$

where h is a real number, and the symbol “.” denotes the dot product of the two vectors.

For the study of partial differential equations, we are particularly interested in the special case where $v = n$ is the unit outward normal vector to the boundary of a domain, as formulated in the following definition.

Definition 1.33 (Directional derivative $\partial F / \partial n$). Let Ω be a domain, and let p be a point in the boundary of the domain. Let $n(p)$ be the outward unit normal vector at a point p . Then using the above definition of a directional derivative,

$$\frac{\partial F}{\partial n}(p) = \nabla F(p) \cdot n(p).$$

We can combine this definition with the level set method. In particular, assume that the boundary of a domain $\Omega \subset \mathbb{R}^d$ is given in the level set form, i.e., as the solution set of the equation $H(x) = c$. Then at a boundary point p for the domain Ω , the gradient $\nabla H(p)$ is a normal vector to $\partial\Omega$. If this vector is pointing outwards, then we have

$$\frac{\partial F}{\partial n}(p) = \frac{\nabla F(p) \cdot \nabla H(p)}{\|\nabla H(p)\|}.$$

If on the other hand the vector $\nabla H(p)$ is pointing into the domain Ω at p , then the directional derivative $\partial F / \partial n(p)$ is the negative of this quantity.

Example 1.34 (Computing $\partial F / \partial n$ for a disk in the plane). Let $\Omega_{disk} \subset \mathbb{R}^2$ be the disk of radius a introduced in Example 1.16, but this time for the special case $a = 1$. Let $F : \Omega \rightarrow \mathbb{R}$ be the function

$$F(x_1, x_2) = x_1^2 - 2x_1x_2 + 4x_2.$$

The goal is to find the value of the directional derivative $\partial F / \partial n(p)$ where n is the outward unit normal, and $p = (1/\sqrt{2}, 1/\sqrt{2})$. Since $\nabla F(p) = (2x_1 - 2x_2, -2x_1 + 4)$, evaluating the gradient of F at the point p gives

$$\nabla F(p) = \left(0, -\frac{2}{\sqrt{2}} + 4 \right).$$

We now assume that the boundary is given in parametric form using polar coordinates as before. Then the point p corresponds to setting $\theta_0 = \pi/4$, i.e., we have $r(\pi/4) = (1/\sqrt{2}, 1/\sqrt{2})$. It was shown that in this case the outward unit normal vector is

$$n(p) = \left(\cos \frac{\pi}{4}, \sin \frac{\pi}{4} \right) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right).$$

Combining these two formulas one then obtains

$$\frac{\partial F}{\partial n}(p) = \left(0, -\frac{2}{\sqrt{2}} + 4\right) \cdot \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{0}{\sqrt{2}} + \frac{-2/\sqrt{2} + 4}{\sqrt{2}} = -1 + \frac{4}{\sqrt{2}} > 0,$$

which shows that the function values of F increase in the direction of the outward normal vector. ■

We end this section with statements of some of the important concepts and results from multivariable calculus. We rely heavily on these ideas in the next section in the derivation of classical partial differential equations when the spatial dimension is greater than one. In all of these definitions and results, a *smooth function* is a function which is sufficiently many times differentiable, i.e., we assume that all occurring derivatives exist.

Definition 1.35 (Divergence). Assume that $G : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a smooth function. The divergence of G , written as $\operatorname{div} G$ or $\nabla \cdot G$, is defined by

$$\operatorname{div} G = \frac{\partial G_1}{\partial x_1} + \frac{\partial G_2}{\partial x_2} + \cdots + \frac{\partial G_d}{\partial x_d}.$$

Definition 1.36 (Laplacian). Assume that $G : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ is a smooth function. The Laplacian of G , written as ΔG , is defined by

$$\Delta G = \frac{\partial^2 G}{\partial x_1^2} + \frac{\partial^2 G}{\partial x_2^2} + \cdots + \frac{\partial^2 G}{\partial x_d^2}.$$

The precise meaning of the Laplacian Δ depends on the number of real arguments of the function G , i.e., on the dimension d of the space. If $d = 1$ and $G = G(x)$ then we have $\Delta G = G_{xx}$, in the case $d = 2$ and $G = G(x_1, x_2)$ we have $\Delta G = G_{x_1 x_1} + G_{x_2 x_2}$, and for $d = 3$ and $G = G(x_1, x_2, x_3)$ one defines $\Delta G = G_{x_1 x_1} + G_{x_2 x_2} + G_{x_3 x_3}$.

Theorem 1.37 (The divergence of the gradient is the Laplacian). Let $u : \Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}$ be a smooth function. Then

$$\operatorname{div}(\nabla u) = \Delta u.$$

The verification of this formula in two and three dimensions is left to the exercises.

Theorem 1.38 (Divergence theorem). Assume that $\Omega \subset \mathbb{R}^d$ is a domain with a piecewise-differentiable boundary and that $G : \overline{\Omega} \rightarrow \mathbb{R}^d$ is a smooth function. Denote by $n(s)$ the outward unit normal at any point $s \in \partial\Omega$. Then

$$\int_{\Omega} \operatorname{div} G(x) dx = \int_{\partial\Omega} G(s) \cdot n(s) ds.$$

Arguably the most important consequence of the divergence theorem is the following result, which generalizes the integration by parts formula to higher-dimensional domains.

Theorem 1.39 (Integration by parts formula). *Let $\Omega \subset \mathbb{R}^d$ be a domain with a piecewise differentiable boundary and let $p, u, v : \bar{\Omega} \rightarrow \mathbb{R}$ be smooth functions. As before, $n(s)$ denotes the unit outward normal vector at $s \in \partial\Omega$. Then*

$$\int_{\Omega} p (\nabla u \cdot \nabla v) dx = \int_{\partial\Omega} p v (\nabla u \cdot n) ds - \int_{\Omega} v \operatorname{div}(p \nabla u) dx.$$

Proof. Let $G : \bar{\Omega} \rightarrow \mathbb{R}^d$ be defined by $G(x) = p(x)v(x)\nabla u(x)$. Then using the product rule for differentiation, we get that for $1 \leq k \leq d$ the identity

$$\begin{aligned} \frac{\partial G_k}{\partial x_k} &= \frac{\partial}{\partial x_k} \left(p v \frac{\partial u}{\partial x_k} \right) \\ &= \frac{\partial p}{\partial x_k} v \frac{\partial u}{\partial x_k} + p \frac{\partial v}{\partial x_k} \frac{\partial u}{\partial x_k} + p v \frac{\partial}{\partial x_k} \left(\frac{\partial u}{\partial x_k} \right) \\ &= p \left(\frac{\partial u}{\partial x_k} \frac{\partial v}{\partial x_k} \right) + v \frac{\partial}{\partial x_k} \left(p \frac{\partial u}{\partial x_k} \right) \end{aligned}$$

holds. Therefore,

$$\operatorname{div} G = p(\nabla u \cdot \nabla v) + v \operatorname{div}(p \nabla u).$$

The result now follows directly from the divergence theorem. \square

The special case $p(x) \equiv 1$ is important enough to deserve its own statement.

Theorem 1.40 (Another integration by parts formula). *Assume that $\Omega \subset \mathbb{R}^d$ is a domain with a piecewise differentiable boundary, and let $u, v : \bar{\Omega} \rightarrow \mathbb{R}$ be smooth functions. As before, let $n(s)$ denote the unit outward normal at $s \in \partial\Omega$. Then*

$$\int_{\Omega} \nabla u \cdot \nabla v dx = \int_{\partial\Omega} v \nabla u \cdot n ds - \int_{\Omega} \Delta u v dx.$$

We write the above theorem in one dimension to show that this gives the standard integration by parts formula. For $\Omega = (a, b) \subset \mathbb{R}$, if we let $f = \nabla u = u'$, and $g = v$, then we obtain

$$\int_{\Omega} f(x)g'(x) dx = \int_{\partial\Omega} f(s)g(s)n(s) ds - \int_{\Omega} f'(x)g(x) dx.$$

Since the boundary of an interval is two points, in this case $\partial\Omega = \{a, b\}$, and the unit outward normal vectors are given by $n(a) = -1$ and $n(b) = 1$, we have that

$$\int_{\partial\Omega} f(s)g(s)n(s) ds = f(b)g(b) - f(a)g(a).$$

Therefore the one-dimensional version of the integration by parts formula is the traditional integration by parts formula introduced in single variable calculus.

1.2 ■ Four Classical Partial Differential Equations

We now proceed with the discussion of the four classical partial differential equations: The heat equation, Laplace's equation, the transport equation, and the wave equation. These equations are ubiquitous throughout the theory of partial differential equations for the simple reason that they are the simplest possible partial differential equations which are of low order, linear, homogeneous, and have no mixed partial derivatives. Studying these simple cases can be quite informative before looking at a more general case. In addition to their simplicity, they serve as good models not just for the situations described here, but also for a number of other scientifically relevant situations.

1.2.1 ■ The Heat Equation

The heat equation governs the transfer of heat in a body of matter. The same equation describes the motion of particles via diffusion, cell motility due to chemotaxis, and many other physical, chemical, and biological processes. We derive it here in the context of diffusion of dye in a tube. Note that the heat equation is also referred to as the diffusion equation depending on the context. For consistency, we will always refer to it as the heat equation even when it is used in other contexts.

The derivation of the heat equation in one space dimension

A narrow straight tube is filled with liquid. A small quantity of dye is released into the liquid. The goal is to describe the spatially-varying concentration of the dye in the tube over time when starting with a prescribed initial spatial distribution of dye. We assume that the motion is governed by simple diffusion. According to Fick's law of diffusion, dye moves from areas of high concentration to regions of lower concentration with a rate proportional to the negative concentration gradient. See Figure 1.3.

Assume that the liquid-filled tube is sufficiently narrow that the concentration of dye is unchanging at each cross section. Thus we are able to restrict our model to one spatial dimension, leading to the one-dimensional heat equation. Let $u(t, x)$ denote the concentration of dye at cross section x in the tube at time t . For $x_1 < x_2$, let $M(t)$ denote the mass of the dye in the region $x_1 < x < x_2$. Then

$$M(t) = \int_{x_1}^{x_2} u(t, x) dx .$$

The rate of change of dye mass in this region is given by

$$\frac{dM}{dt}(t) = \int_{x_1}^{x_2} u_t(t, x) dx .$$

The rate of change of the dye mass in the region is also equal to the net flow of dye into the region. By the one-dimensional version of Fick's law of diffusion, the rightward flow of the dye is proportional to $-u_x(t, x)$. That is, the flow is equal to $-k u_x(t, x)$, where the constant $k > 0$ is the same constant for every cross section in the tube. Therefore the net flow into the region is equal to the sum of the inward flows evaluated at the two endpoints of the region. The inward flow at x_1 is the rightward flow $-k u_x(t, x_1)$, whereas the inward flow at x_2 is the leftward flow $+k u_x(t, x_2)$, giving

$$\frac{dM}{dt}(t) = k u_x(t, x_2) - k u_x(t, x_1) .$$

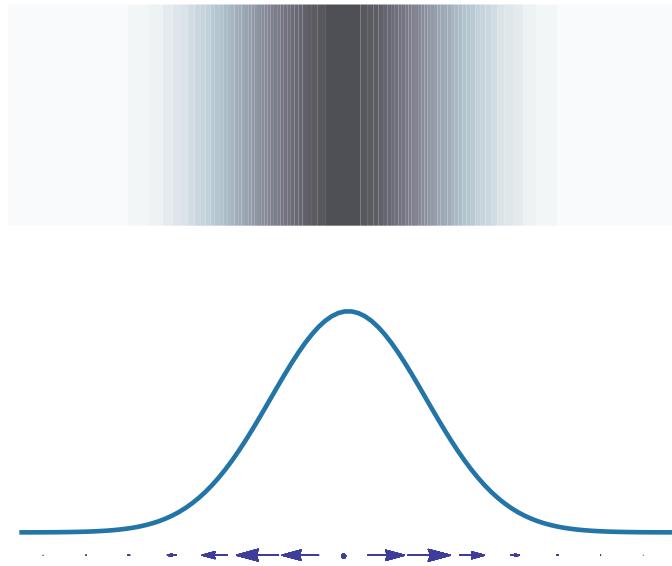


Figure 1.3. This figure shows the physical setup in the derivation of the heat equation. From top to bottom, the images show the following: (a) A small quantity of dye is placed in a narrow straight tube filled with fluid, resulting in a non-uniform distribution of fluid such as the one depicted here. (b) At a fixed time t , the function $u(t, x)$ is the concentration of dye at each point of the fluid. (c) The dye moves from area of high concentration to areas of low concentration. The vectors indicate the direction of flow of the dye in (a) and (b).

Combining the two equations for dM/dt , followed by using the fundamental theorem of calculus, we get

$$\int_{x_1}^{x_2} u_t(t, x) dx = k u_x(t, x_2) - k u_x(t, x_1) = \int_{x_1}^{x_2} k u_{xx}(t, x) dx .$$

Differentiating the left and right sides of this equation with respect to x_2 gives

$$u_t(t, x_2) = k u_{xx}(t, x_2) .$$

This is the one-dimensional heat equation.

The derivation of the heat equation in higher space dimensions

In dimensions $d = 2, 3$, we let $\Omega \subset \mathbb{R}^d$ denote the domain which is filled with liquid. In order to derive the higher-dimensional versions of the heat equation, let $u(t, x)$ be the concentration of dye at point $x = (x_1, \dots, x_d) \in \Omega \subset \mathbb{R}^d$ and time $t \geq 0$.

As in the one-dimensional case, we write the rate of change of dye mass in a subset D of the domain Ω . The mass of dye in D is given by

$$M(t) = \int_D u(t, x) dx ,$$

and the rate of change of dye mass in D is given by

$$\frac{dM}{dt}(t) = \int_D u_t(t, x) dx .$$

We formalize Fick's law of diffusion in higher dimensions as a physical law describing the point rate of change per unit time (or flow rate) of mass per unit area. This quantity is known as mass flux density.

Definition 1.41 (Fick's law of diffusion). *In a system governed by diffusion, the mass flux density is proportional to the negative concentration gradient $-\nabla u$, of mass concentration u where the gradient only includes derivatives with respect to the spatial variables and not with respect to time t .*

As in the above definition, it is standard notation to apply the gradient, divergence, and Laplacian operators only to spatial variables and not to the time variable. We implicitly use this notation throughout the book. For example, for a function $u(t, x)$ with spatial argument $x \in \mathbb{R}^3$ we have $\nabla u = (u_{x_1}, u_{x_2}, u_{x_3})$ and $\Delta u = u_{x_1 x_1} + u_{x_2 x_2} + u_{x_3 x_3}$.

By Fick's law of diffusion, at a point p on the boundary ∂D the flow of dye outward from D is proportional to the directional derivative of the negative concentration u in the direction of the outward unit normal n , i.e., to the scalar projection of the negative gradient $-\nabla u$ onto the outward unit normal. In other words, the outward component of mass flux density at $p \in \partial D$ is

$$-k \frac{\partial u}{\partial n}(p) = -k \nabla u \cdot n,$$

where $k > 0$ is the same proportionality constant at each point. The constant k is known as the diffusion coefficient and is determined by the particular fluid and dye used. The *inward* component of the mass flux density is the negative of the above expression. Thus at fixed time t , the net inflow of dye through the entire boundary ∂D is given by the integral

$$\int_{\partial D} k \nabla u \cdot n \, ds.$$

Furthermore, the net inflow of dye on the boundary of D is equal to the rate of change of dye mass dM/dt . Combining our two expressions for dM/dt gives

$$\int_D u_t(t, x) \, dx = \int_{\partial D} k \nabla u(t, s) \cdot n(s) \, ds.$$

Applying the Divergence Theorem 1.38 to the right-hand side with $G = \nabla u$, and then applying Theorem 1.37, we get

$$\int_D u_t(t, x) \, dx = \int_{\partial D} k \nabla u(t, s) \cdot n(s) \, ds = \int_D k \operatorname{div}(\nabla u) \, dx = \int_D k \Delta u(t, x) \, dx.$$

Since this statement is true for every domain D contained in Ω , we can conclude that the integrands in the leftmost and rightmost expressions have to be equal. That is, the equation

$$u_t = k \Delta u$$

has to be satisfied, which is the higher-dimensional heat equation.

We have derived this equation in the case of diffusion of dye in a fluid. In a different context, consider a body of matter of varying temperature. Fourier's law of conduction states that the rate of change of temperature at a point — also known as heat flux density — is proportional to the negative gradient of temperature. The proportionality constant k

is a property of the material known as thermal conductivity. Through an identical derivation, the temperature profile $u(t, x)$ is governed by the heat equation. There are similar flux laws in other situations as well, but our subsequent discussion focuses on these two cases.

Initial and boundary conditions

If one just considers the heat equation without any further constraints, the equation has infinitely many solutions. In order to specify a unique solution, we have to specify both initial and boundary conditions. The initial conditions describe the starting state of the system, such as the initial concentration of dye in a fluid, or the initial temperature profile of a body of matter. This is given by

- *Initial conditions:* Let $u_0 : \Omega \rightarrow \mathbb{R}$ be given. Then we require

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

In this formulation, the domain $\Omega \subset \mathbb{R}^d$ is assumed to be open. However, it has become standard notation in partial differential equations to assume that the function u , and in some cases its derivatives, also has values on the boundary $\partial\Omega$ of the domain, where the values are determined in terms of limits.

Boundary conditions prescribe values along the boundary of the domain either for u or for its derivative u . In particular, there are several types of commonly used boundary conditions for a domain Ω with boundary $\partial\Omega$:

- *Dirichlet boundary conditions:* Let $f : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ be given, and require

$$u(t, x) = f(t, x) \quad \text{on } \partial\Omega, \quad \text{for all } t > 0.$$

If u describes temperature, and the boundary of the domain is kept at a certain temperature profile $f(t, x)$, then Dirichlet boundary conditions are the appropriate choice.

- *Neumann boundary conditions:* Let $g : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ be given, and require

$$\frac{\partial u}{\partial n}(t, x) = g(t, x) \quad \text{on } \partial\Omega, \quad \text{for all } t > 0,$$

where $\partial u / \partial n$ is defined in Definition 1.33. If u describes a temperature, and the boundary is kept perfectly insulated so that no heat leaks out the boundary, then Neumann boundary conditions with $g \equiv 0$ are the appropriate boundary conditions to choose.

- *Robin boundary conditions:* Robin boundary conditions are a linear combination of Dirichlet and Neumann conditions. Let $h : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ be given, and let $a, b \in \mathbb{R}$. Then we require

$$au(t, x) + b\frac{\partial u}{\partial n}(t, x) = h(t, x) \quad \text{on } \partial\Omega, \quad \text{for all } t > 0.$$

- *Periodic boundary conditions:* For periodic boundary conditions, the domain has to be rectilinear, i.e., it has to be in the form $\Omega = (a_1, b_1) \times \cdots \times (a_d, b_d)$, and we require

$$\begin{aligned} u|_{x_k=a_k} &= u|_{x_k=b_k} \quad \text{and} \\ u_{x_k}|_{x_k=a_k} &= u_{x_k}|_{x_k=b_k} \quad \text{for } k = 1, \dots, d. \end{aligned}$$

Notice that this gives a condition for all $t > 0$ and all points on the boundary of Ω . In the case $d = 1$, this amounts to considering an interval domain with identified endpoints, and therefore one can think of the heat equation being considered on a circular base domain. In the case $d = 2$, identifying the top and bottom edges of the domain, as well as the left and right vertical edges, the base domain can be thought of as a torus.

- *Homogeneous boundary conditions:* In the case of either Dirichlet boundary conditions with $f \equiv 0$ or Neumann boundary conditions with $g \equiv 0$, or in the case of Robin boundary conditions with $h \equiv 0$, the respective boundary conditions are called *homogeneous boundary conditions*.

Once the heat equation is subjected to both an initial condition and a boundary condition of one of the types above, we usually expect the equation to have a unique solution. Note that although this discussion separates the specification of an initial condition from the specification of a boundary condition, the combination of these conditions gives a specification on the boundary of the full time-space domain $\mathbb{R}^+ \times \Omega$.

A solution for a particular heat equation problem

In order to give an idea of the types of solutions occurring for the heat equation, we state a solution to the one-dimensional heat equation, verify that it is indeed a solution, and make observations about the behavior of this solution. In the next chapter, we will develop the method of separation of variables for finding solutions for partial differential equations such as the heat equation. The solution given below is derived once separation of variables has been introduced.

Consider the heat equation on the domain $\Omega = (0, 1)$ with specified initial conditions and homogeneous Dirichlet boundary conditions. Specifically, let n be an integer, and assume

$$\begin{aligned} u_t &= u_{xx} \quad \text{for } x \in \Omega \text{ and } t > 0, \\ &\text{with } u(t, 0) = u(t, 1) = 0, \\ &\text{and } u(0, x) = \sin(n\pi x). \end{aligned}$$

Then a solution to the equation is given by

$$u(t, x) = e^{-n^2\pi^2 t} \sin(n\pi x), \quad \text{where } n \in \mathbb{N}, \quad (1.8)$$

see also Figure 1.4. In order to verify that this is indeed a solution, we plug the given function into the equation, boundary, and initial conditions to obtain

$$\begin{aligned} u_t &= -n^2\pi^2 u, \\ u_{xx} &= -n^2\pi^2 u, \\ u(0, x) &= \sin(n\pi x), \\ u(t, 0) &= 0, \quad u(t, 1) = 0. \end{aligned}$$

Notice that for each n , this solution exponentially decays or *dissipates* to $u \equiv 0$. In fact dissipation to a homogeneous solution is typical for solutions to the heat equation. In contrast, for the wave equation we will soon show that solutions continue to either translate or oscillate with fixed amplitude without any dissipation.

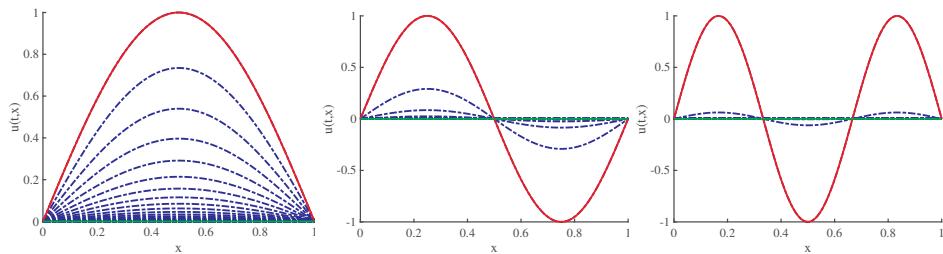


Figure 1.4. Solutions of the heat equation in one dimension given in (1.8) at a series of increasing times. From left to right, the images correspond to the parameter values $n = 1, 2, 3$, respectively. In each image, one can see that the solutions exhibit exponential decay with increasing time; red curves are for time zero, green curves for time 2, and blue curves correspond to equidistant time steps between 0 and 2.

1.2.2 • Laplace's Equation

Our second classical partial differential equation is Laplace's equation. While we do not present a detailed derivation of the equation, we note that it occurs in problems of gravitation, elastic membranes, stationary heat distribution, and electricity and magnetism. In addition, it is the equation for time-independent solutions for both the heat equation and the wave equation. Let $\Omega \subset \mathbb{R}^d$ denote a domain, where for now we assume $d \leq 3$. Then Laplace's equation is given by

$$\Delta u = 0 \quad \text{in } \Omega, \quad (1.9)$$

where the notation Δu is defined in Definition 1.36.

In physical examples, the Laplace equation does not depend on time, and the independent variables are all viewed as spatial variables. Despite this lack of any time dependence, Laplace's equation has infinitely many solutions without any further constraints. Solutions of this equation are historically called *harmonic functions*. In the one-dimensional situation the solutions can be completely described, since $\Delta u = 0$ translates into the condition that the second derivative of the function $u = u(x)$ vanishes, and this means that $u(x) = ax + b$ on every component of Ω , for suitable $a, b \in \mathbb{R}$. In other words, harmonic functions in one dimension are affine functions on each disjoint interval of Ω , and therefore not particularly interesting.

The above reasoning can easily be extended to the case of higher-dimensional domains. One can readily verify that any linear function of the form

$$u(x_1, x_2) = ax_1 + bx_2 + c, \quad \text{for arbitrary } a, b, c \in \mathbb{R},$$

is a solution for $d = 2$, and similarly for $d = 3$. However, unlike in the one-dimensional case, there are many more solutions to Laplace's equation in dimensions two and higher. Some of these will be discussed below.

In order to pin down a unique solution, one again generally needs to specify boundary conditions, which usually are of one of the following forms. As for the heat equation, our base domain $\Omega \subset \mathbb{R}^d$ is defined as an open set, and we implicitly assume that a solution u of Laplace's equation has well-defined function values, and in some cases derivative values, on the boundary $\partial\Omega$ of Ω , which arise via limits of values inside the domain.

- *Dirichlet boundary conditions:* If f is a given real-valued function defined on the boundary $\partial\Omega$, we require

$$u = f \quad \text{on } \partial\Omega.$$

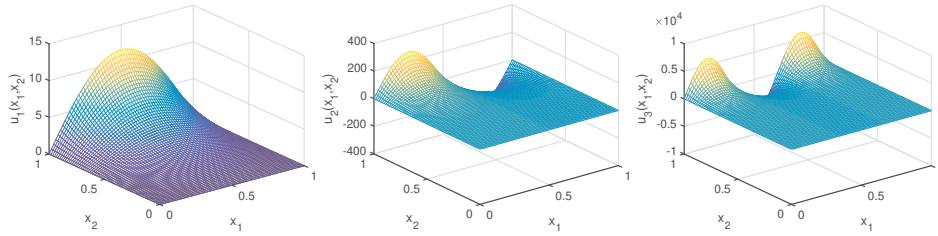


Figure 1.5. Solutions for Laplace's equation for the Dirichlet boundary value problem described in the text. This problem depends on the integer parameter n , and the three images show the solutions for $n = 1, 2, 3$.

For example in the two-dimensional case $u = u(x_1, x_2)$ this has to be interpreted as the equality $u(x_1, x_2) = f(x_1, x_2)$ for all $(x_1, x_2) \in \partial\Omega$.

- **Neumann boundary conditions:** If g is a given real-valued function defined on the boundary $\partial\Omega$, we require

$$\frac{\partial u}{\partial n} = g \quad \text{on} \quad \partial\Omega,$$

where $\partial u / \partial n$ is defined in Definition 1.33.

- **Robin boundary conditions:** If a and b are real constants and if h is a given real-valued function defined on the boundary $\partial\Omega$, we require

$$a u + b \frac{\partial u}{\partial n} = h \quad \text{on} \quad \partial\Omega.$$

- **Periodic boundary conditions:** For periodic boundary conditions, the domain must be rectilinear in the form $\Omega = (a_1, b_1) \times \dots \times (a_d, b_d)$, and for $u = u(x_1, \dots, x_d)$ we require

$$\begin{aligned} u|_{x_k=a_k} &= u|_{x_k=b_k} \quad \text{and} \\ u_{x_k}|_{x_k=a_k} &= u_{x_k}|_{x_k=b_k} \quad \text{for } k = 1, \dots, d. \end{aligned}$$

- **Homogeneous boundary conditions:** As for the heat equation, homogeneous boundary conditions correspond to the case where the functions f , g , or h in the first three cases above are identically zero.

We would like to specifically point out that since Laplace's equation is time independent, there is no notion of an initial condition!

As an example consider Laplace's equation on the unit square in \mathbb{R}^2 with specific boundary conditions. In particular, assume that

$$\Delta u = u_{x_1 x_1} + u_{x_2 x_2} = 0 \quad \text{on} \quad \Omega = (0, 1)^2.$$

The boundary $\partial\Omega$ is given by the four sides of the square. In terms of boundary conditions, suppose that u has to solve Dirichlet boundary conditions, with $u = 0$ on the lower, left, and right sides of the boundary $\partial\Omega$. On the top edge $x_2 = 1$, we assume that

$$u = \frac{\sin(n\pi x_1)(e^{n\pi} - e^{-n\pi})}{2},$$

where $n \in \mathbb{N}$. Then one can easily verify that the function

$$u(x_1, x_2) = \sin(n\pi x_1) \sinh(n\pi x_2)$$

solves the above boundary value problem associated with Laplace's equation. One just has to check that the equation and its boundary conditions are satisfied. In Figure 1.5, these solutions are shown for $n = 1, 2, 3$. In fact, we will see below that for every $n \in \mathbb{N}$ the above solution is the unique solution to the corresponding boundary value problem.

Connection with complex analytic functions

It was mentioned before that solutions of Laplace's equation are also referred to as harmonic functions. As is often the case in mathematics, if the same objects have several different names, this usually indicates that they were discovered separately and independently in different fields. In the context of harmonic functions in two variables, there is a surprising connection with complex analytic functions.

For the benefit of readers who have not yet been exposed to complex analysis, we start with notation and results for complex numbers. This should also serve as a review for a reader who has already seen complex analysis. The quantity i , called the imaginary unit, is a number satisfying

$$i^2 = -1. \quad (1.10)$$

The field of complex numbers is given by

$$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}.$$

Computations in \mathbb{C} obey the same rules as computations in \mathbb{R} , as long as one makes use of the fundamental identity (1.10). For a given complex number $z = x + iy$, the real number x is called its real part, denoted by $x = \operatorname{Re} z$, and the real number y is called its imaginary part, denoted by $y = \operatorname{Im} z$. With every complex number $z = x + iy$ we associate its complex conjugate \bar{z} , which is given by $\bar{z} = x - iy$.

The complex numbers \mathbb{C} contain the real numbers as a special case, one just has to consider complex numbers with zero imaginary part. Thus, one can view complex numbers as an extension of the real number line to a plane. In fact, there is an identification of the complex plane \mathbb{C} and the Euclidean plane \mathbb{R}^2 via

$$\mathbb{C} \ni z = x + iy \leftrightarrow (x, y) \in \mathbb{R}^2. \quad (1.11)$$

Using this identification, the complex modulus of $z \in \mathbb{C}$

$$|z| = \sqrt{z\bar{z}} = \sqrt{x^2 + y^2} \quad (1.12)$$

is the Euclidean distance from the point z to the origin. Thus complex numbers can be visualized in the plane \mathbb{R}^2 , by using their real and imaginary parts as coordinates. Under this identification, the real numbers lie on the horizontal coordinate axis.

Let $f : \Omega \rightarrow \mathbb{C}$ denote a function from a subset $\Omega \subset \mathbb{C}$ of the complex plane into the complex plane. Such a function is called an *analytic function*, if for every point $z_0 \in \Omega$ the derivative

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. Note that even for functions like $f(z) = z^n$, this is a much stronger condition than differentiability in \mathbb{R} . In \mathbb{R} sequences approaching z_0 can only approach from the left or

from the right, while in \mathbb{C} all possible approaches within the complex plane have to be considered, whether they lie on straight lines through z_0 or on complicated curves.

Many functions that can be defined for complex numbers turn out to be analytic, such as for example polynomials. In addition, if a power series

$$f(z) = \sum_{k=0}^{\infty} c_k z^k \quad \text{with} \quad c_k \in \mathbb{C} \quad \text{for} \quad k \in \mathbb{N}_0$$

has radius of convergence $R > 0$, and if $\Omega \subset \mathbb{C}$ denotes the interior of the open disk of radius R centered at the origin $0 \in \mathbb{C}$, then $f : \Omega \rightarrow \mathbb{C}$ is guaranteed to be analytic. This shows for example that $f(z) = e^z$, as well as trigonometric functions like sine and cosine are analytic on $\Omega = \mathbb{C}$. In fact, relating the power series of e^z , sine, and cosine leads to the following celebrated Euler's identity, which in later chapters will be extremely important in our numerical methods.

$$e^{i\theta} = \cos \theta + i \sin \theta, \quad \text{for all } \theta \in \mathbb{R} \quad (\text{Euler's identity}). \quad (1.13)$$

We now return to the connection between analytic functions and solutions to Laplace's equation. This connection uses identification (1.11) between the complex plane \mathbb{C} and the Euclidean plane \mathbb{R}^2 . In particular, if we are given an analytic function $f : \Omega \rightarrow \mathbb{C}$, then we can define two real-valued functions $u, v : \Omega \rightarrow \mathbb{R}$ by considering real and imaginary parts in the form

$$f(x + iy) = u(x, y) + i v(x, y).$$

Then the following holds:

- If $f : \Omega \rightarrow \mathbb{C}$ is analytic, then the two functions $u, v : \Omega \rightarrow \mathbb{R}$, the real and imaginary parts of f , are both harmonic.

For example, consider the exponential function $f(z) = e^z$. Due to Euler's identity (1.13) we have

$$f(x + iy) = e^{x+iy} = e^x e^{iy} = e^x \cos y + i e^x \sin y,$$

i.e., the functions $u(x, y) = e^x \cos y$ and $v(x, y) = e^x \sin y$ are both harmonic and therefore solve Laplace's equation — which of course can easily be verified directly. Slightly less intuitive are the functions that result from the monomials $f(z) = z^n$. If we assume that the integer $n \in \mathbb{N}$ is even, then the binomial formula implies

$$\begin{aligned} f(x + iy) &= (x + iy)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k i^k \\ &= \sum_{\ell=0}^{n/2} \binom{n}{2\ell} x^{n-2\ell} y^{2\ell} i^{2\ell} + \sum_{\ell=0}^{n/2-1} \binom{n}{2\ell+1} x^{n-2\ell-1} y^{2\ell+1} i^{2\ell+1} \\ &= \underbrace{\sum_{\ell=0}^{n/2} \binom{n}{2\ell} (-1)^\ell x^{n-2\ell} y^{2\ell}}_{=u(x,y)} + i \underbrace{\sum_{\ell=0}^{n/2-1} \binom{n}{2\ell+1} (-1)^\ell x^{n-2\ell-1} y^{2\ell+1}}_{=v(x,y)}. \end{aligned}$$

In other words, for any even integer $n \in \mathbb{N}$ both functions

$$\sum_{\ell=0}^{n/2} \binom{n}{2\ell} (-1)^\ell x^{n-2\ell} y^{2\ell} \quad \text{and} \quad \sum_{\ell=0}^{n/2-1} \binom{n}{2\ell+1} (-1)^\ell x^{n-2\ell-1} y^{2\ell+1} \quad (1.14)$$

solve Laplace's equation. We leave it as a simple exercise to treat the remaining case of monomials $f(z) = z^n$ for odd integers $n \in \mathbb{N}$. Even these few examples already demonstrate that although Laplace's equation in one space dimension is trivial, solutions on higher-dimensional domains can be very complicated, and certainly do not fall into a simple formula category.

Mean-value properties

In contrast to the examples that we have seen so far, most solutions of Laplace's equation in higher dimensions cannot be expressed in simple formulas — but they still exhibit remarkable properties. One of the most important of these properties is the mean-value property, which is the subject of the following result. We only state the result for harmonic functions in two variables, but the same results hold in higher dimensions as well.

Theorem 1.42 (Mean-value property). *Let $\Omega \subset \mathbb{R}^2$ denote an open domain, and suppose that the continuous function $u : \bar{\Omega} \rightarrow \mathbb{R}$ is harmonic, i.e., assume that u solves Laplace's equation $\Delta u = 0$ in Ω . Furthermore, let $(x_0, y_0) \in \Omega$ be any interior point of Ω , and let $R > 0$ be such that the circular disk $B_R(x_0, y_0)$ of radius R and with center (x_0, y_0) is completely contained in $\bar{\Omega}$. Then the function value $u(x_0, y_0)$ satisfies*

$$u(x_0, y_0) = \frac{1}{2\pi R} \int_{\partial B_R(x_0, y_0)} u(x, y) ds, \quad (1.15)$$

as well as

$$u(x_0, y_0) = \frac{1}{\pi R^2} \int_{B_R(x_0, y_0)} u(x, y) d(x, y). \quad (1.16)$$

In other words, the value $u(x_0, y_0)$ equals the average of u over the disk $B_R(x_0, y_0)$, as well as the average of u over the boundary of this disk, regardless of its radius.

Proof. For $0 < r \leq R$ we define the function $v(r)$ as the average

$$v(r) = \frac{1}{2\pi r} \int_{\partial B_r(x_0, y_0)} u ds$$

of u over the boundary of the disk $B_r(x_0, y_0)$. In order to evaluate the line integral in this formula, we consider the parameterization of $\partial B_r(x_0, y_0)$ given by

$$(x(\theta), y(\theta)) = (x_0 + r \cos \theta, y_0 + r \sin \theta) \quad \text{for } 0 \leq \theta \leq 2\pi.$$

Then one obtains

$$v(r) = \frac{1}{2\pi r} \int_0^{2\pi} u(x(\theta), y(\theta)) r d\theta = \frac{1}{2\pi} \int_0^{2\pi} u(x_0 + r \cos \theta, y_0 + r \sin \theta) d\theta,$$

and differentiating with respect to r , in combination with the divergence theorem, fur-

nishes

$$\begin{aligned}
v'(r) &= \frac{1}{2\pi} \int_0^{2\pi} (u_x(x(\theta), y(\theta)) \cos \theta + u_y(x(\theta), y(\theta)) \sin \theta) d\theta \\
&= \frac{1}{2\pi r} \int_0^{2\pi} (u_x(x(\theta), y(\theta)) \cos \theta + u_y(x(\theta), y(\theta)) \sin \theta) r d\theta \\
&= \frac{1}{2\pi r} \int_{\partial B_r(x_0, y_0)} \nabla u(x, y) \cdot n(x, y) ds \\
&= \frac{1}{2\pi r} \int_{B_r(x_0, y_0)} \underbrace{\operatorname{div} \nabla u(x, y)}_{=\Delta u(x, y)} d(x, y) = 0.
\end{aligned}$$

Thus, the function v is differentiable on $(0, R)$ with derivative zero. One can also show that v is continuous on $[0, R]$, and therefore, it has to be constant on $[0, R]$. Finally, in the limit $r \rightarrow 0$ the average in the definition of $v(r)$ has to converge to $u(x_0, y_0)$, and so (1.15) follows. In order to prove the second identity (1.16), we use polar coordinates, together with the above definition of $v(r)$ and obtain

$$\begin{aligned}
\int_{B_R(x_0, y_0)} u(x, y) d(x, y) &= \int_0^R \int_0^{2\pi} u(x_0 + r \cos \theta, y_0 + r \sin \theta) r d\theta dr \\
&= \int_0^R 2\pi v(r) r dr = 2\pi u(x_0, y_0) \int_0^R r dr \\
&= \pi R^2 u(x_0, y_0),
\end{aligned}$$

since we have just shown that $v(r) = u(x_0, y_0)$ for all $0 \leq r \leq R$. This completes the proof of the theorem. \square

Upon closer inspection, the proof of the above theorem even provides information on functions which are not harmonic, but only satisfy an inequality of the form $\Delta u \geq 0$ or $\Delta u \leq 0$ in Ω . Since this will be of importance later on, we formulate the following easy corollary.

Corollary 1.43 (Mean-value inequalities). *Let $\Omega \subset \mathbb{R}^2$ denote an open domain, and suppose that $u \in C(\overline{\Omega}) \cap C^2(\Omega)$, i.e., the function u is continuous on $\overline{\Omega}$ and twice continuously differentiable in Ω . Furthermore, let $(x_0, y_0) \in \Omega$ be an interior point of Ω , and let $R > 0$ be such that the circular disk $B_R(x_0, y_0)$ is completely contained in $\overline{\Omega}$. If the function u satisfies the inequality $\Delta u \geq 0$ on Ω , then we have both*

$$u(x_0, y_0) \leq \frac{1}{2\pi R} \int_{\partial B_R(x_0, y_0)} u(x, y) ds \quad \text{and} \quad u(x_0, y_0) \leq \frac{1}{\pi R^2} \int_{B_R(x_0, y_0)} u(x, y) d(x, y).$$

Similarly, if the function u satisfies $\Delta u \leq 0$ in Ω , then the integrals in the above two inequalities provide lower bounds for $u(x_0, y_0)$.

Proof. We only consider the case $\Delta u \geq 0$ on Ω and assume the notation of the last proof. Then one can easily see that $\Delta u \geq 0$ implies $v'(r) \geq 0$ for all $r \in (0, R)$. This in turn furnishes the estimate $u(x_0, y_0) = v(0) \leq v(R)$, which yields the first inequality. The second inequality follows from the fact that $v(r) \geq u(x_0, y_0)$ for all $r \in [0, R]$. \square

We have only stated and proved the mean-value property of harmonic functions and the mean-value inequalities for the case of functions of two real variables. Showing that these results remain true in arbitrary dimensions uses an almost identical proof. One just has to replace the formulas for the perimeter and area of a circular disk by the $(d-1)$ -dimensional volume of the boundary of a d -dimensional ball, and the d -dimensional volume of the ball itself. Furthermore, one can easily show that the mean-value property of Theorem 1.42 characterizes harmonic functions: If u is a function such that the mean-value property holds at all points in the domain and for all feasible radii, then u has to be harmonic.

The maximum principle

While at first sight the mean-value property seems to be just an intriguing geometric peculiarity of harmonic functions, it has extremely important consequences. One of these is related to the concept of maxima and minima of real-valued functions. As one learns in multivariable calculus, if $\Omega \subset \mathbb{R}^d$ is a bounded open domain, and if $u : \overline{\Omega} \rightarrow \mathbb{R}$ is a continuous function, then u attains both a maximum and a minimum on $\overline{\Omega}$. More precisely, there exist points p_{\min} and p_{\max} in $\overline{\Omega}$ such that

$$u(p_{\min}) \leq u(p) \leq u(p_{\max}) \quad \text{for all } p \in \overline{\Omega}.$$

This fact is a reflection of the interaction of continuity of a function and the compactness of the domain $\overline{\Omega}$. Yet even though this result is extremely powerful and general, it does not provide any insight into how many points p_{\min} and p_{\max} with the above property exist, or where in the domain the points can be expected. For the special case of harmonic functions, the possible locations can be narrowed down dramatically — and this is the subject of the maximum principle for solutions of Laplace's equations.

In the following, we provide a complete statement for the maximum principle for harmonic functions in general dimensions. However, since we established the mean-value property only for the two-dimensional case, we will give the proof of the maximum principle only in the two-dimensional situation. But the reader will quickly see that the same method can be adapted to the higher-dimensional case as well.

Theorem 1.44 (Maximum principle for harmonic functions). *Let $\Omega \subset \mathbb{R}^d$ be a bounded connected open domain. Assume further that $u : \overline{\Omega} \rightarrow \mathbb{R}$ is continuous, twice differentiable in Ω , and that $\Delta u = 0$ in Ω . In other words, assume that u is harmonic in Ω and can be continuously extended to the closure $\overline{\Omega}$. Then the following hold:*

- (a) *If u attains its maximum at an interior point $p_{\max} \in \Omega$, then the function u is constant on $\overline{\Omega}$. Similarly, if u attains its minimum at a point $p_{\min} \in \Omega$, then u is constant on $\overline{\Omega}$.*
- (b) *If u is not a constant function, then u achieves both its maximum and its minimum at the boundary $\partial\Omega$.*

Proof. It is clear that (b) is just a reformulation of (a). In order to prove (a), we only consider the two-dimensional case $d = 2$. For this, let $p_{\max} = (x_{\max}, y_{\max})$ be a point in Ω at which u attains its maximum, and let $R > 0$ be any radius such that $B_R(p_{\max}) \subset \Omega$. Since for all points $p = (x, y) \in \Omega$ we have $u(x, y) \leq u(x_{\max}, y_{\max})$, the mean-value property (1.16)

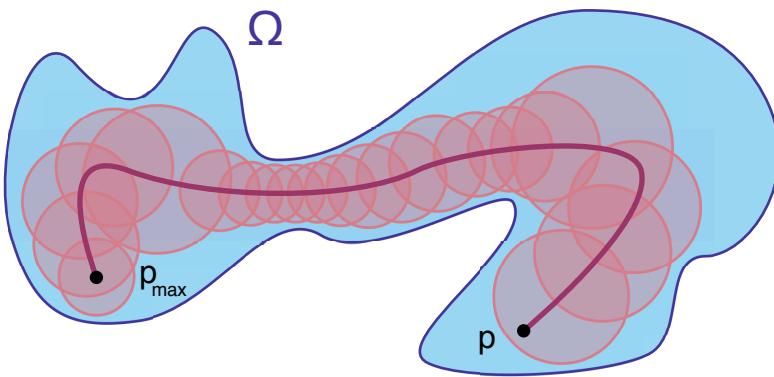


Figure 1.6. Proof of the maximum principle. To show that u is constant on Ω , one needs to connect the two points p_{\max} and p via a sequences of disks in Ω , such that the center of the next disk is always contained in the previous disk. Notice that while p_{\max} has to be the center of the first disk, the point p can lie anywhere in the last one.

implies the inequality

$$\begin{aligned} u(p_{\max}) &= \frac{1}{\pi R^2} \int_{B_R(p_{\max})} u(x, y) d(x, y) \\ &\leq \frac{1}{\pi R^2} \int_{B_R(p_{\max})} u(x_{\max}, y_{\max}) d(x, y) = u(p_{\max}), \end{aligned} \quad (1.17)$$

and therefore one has to have

$$\int_{B_R(p_{\max})} u(x, y) d(x, y) = \int_{B_R(p_{\max})} u(x_{\max}, y_{\max}) d(x, y). \quad (1.18)$$

Now assume that there exists a point p in $B_R(p_{\max})$ at which u has a function value strictly smaller than $u(p_{\max})$. Then there exists a neighborhood $U \subset B_R(p_{\max})$ of p on which the function values are strictly smaller than the maximal function value of u . This, however, implies the strict inequality

$$\int_{B_R(p_{\max})} u(x, y) d(x, y) < \int_{B_R(p_{\max})} u(x_{\max}, y_{\max}) d(x, y),$$

which contradicts (1.18). Thus, our above assumption on $u(p)$ has to be wrong, and therefore the function u has to be constant on the whole ball $B_R(p_{\max})$.

In the first part of the proof we have just shown that if u attains its maximal value at $p_{\max} \in \Omega$ and if $R > 0$ is such that $B_R(p_{\max}) \subset \Omega$, then u attains its maximum value at every point in $B_R(p_{\max})$, in other words, the function u is constant on $B_R(p_{\max})$. Now let $p \in \Omega$ be arbitrary. Since u is open, we can connect p_{\max} and p by a smooth path which lies in Ω , and along this path, we can find disks $B_{R_k}(p_k) \subset \Omega$ for $k = 1, \dots, M$ such that $p_1 = p_{\max}$, the point $p \in B_{R_M}(p_M)$, as well as $p_{k+1} \in B_{R_k}(p_k)$ for $k = 1, \dots, M-1$, see also Figure 1.6. Due to the first part of the proof, u is constant on the first disk $B_{R_1}(p_1)$. Thus, the function u attains its maximum at $p_2 \in B_{R_1}(p_1)$, and therefore is constant on the second disk as well. Continuing, we see that the harmonic function u is constant on

all M disks, and therefore $u(p) = u(p_{\max})$. Since $p \in \Omega$ was arbitrary, we obtain that u is constant on Ω , and due to the continuity of u therefore also on $\bar{\Omega}$.

The case where u attains its global minimum in Ω can be treated similarly, and is left as an exercise for the reader. \square

The above result and its proof are quite surprising. The maximum principle follows from an inherently local argument with very strict underlying geometry, namely spheres in the domain, but gives us global information on the solution of the Laplace equation, regardless of the actual geometry of the large domain. In fact, one can show that the maximum principle holds in general for elliptic partial differential equations, which will be described in more detail later in the book — even for equations whose underlying differential operators do not have the mean-value property.

While the main focus of the present section are properties of solutions to Laplace's equation, the maximum principle can also be formulated for certain functions which are not harmonic. In fact, this more general result is hiding in plain sight in the above proof! This leads to the *strong maximum principle*.

Theorem 1.45 (Strong maximum principle). *Let $\Omega \subset \mathbb{R}^d$ be a bounded connected open domain. Assume further that $u \in C(\bar{\Omega}) \cap C^2(\Omega)$ satisfies $\Delta u \geq 0$ in Ω . Then we have:*

- (a) *If u attains its maximum at an interior point $p_{\max} \in \Omega$, then the function u is constant on the closed domain $\bar{\Omega}$.*
- (b) *If u is not a constant function, then u achieves its maximum at the boundary $\partial\Omega$.*

If on the other hand we have $\Delta u \leq 0$ in Ω , then both (a) and (b) remain valid if “maximum” is replaced by “minimum.”

Proof. As before, we only have to establish (a). Suppose therefore that $\Delta u \geq 0$. Then we can proceed as in the proof of Theorem 1.44 up to (1.17), where the first equality is no longer satisfied. In the previous result, this equality was the result of the mean-value property for harmonic functions. Note, however, that in our current situation, we can use Corollary 1.43 to replace the equality sign by \leq , and the reader can easily verify that this change leaves the remaining arguments of the proof of Theorem 1.44(a) intact. This completes the proof of the strong maximum principle, since the case $\Delta u \leq 0$ can be reduced to the case $\Delta u \geq 0$ by considering $-u$ instead of u . \square

The strong maximum principle will prove to be extremely useful later on. For now, however, we stick to the case of harmonic functions and provide two easy consequences of the maximum principle, which shed further light on the solutions of Laplace's equation, in particular in the context of boundary value problems with Dirichlet boundary conditions. The first result allows us to transfer inequalities on the boundary of a domain to inequalities which hold in the interior of the domain.

Corollary 1.46 (Comparison principle for Laplace's equation). *Let $\Omega \subset \mathbb{R}^d$ be a bounded open domain. Assume further that $u, v \in C(\bar{\Omega}) \cap C^2(\Omega)$ both satisfy Laplace's equation, i.e., we have $\Delta u = 0$ and $\Delta v = 0$ in Ω , and both u and v can be continuously extended to the closure $\bar{\Omega}$. Finally suppose that on the boundary of Ω , the function u is at least as large as v , i.e., assume that*

$$u(p) \geq v(p) \quad \text{for all } p \in \partial\Omega.$$

Then we have

$$u(p) \geq v(p) \quad \text{for all } p \in \Omega,$$

i.e., the inequality on the boundary transfers into the interior of the domain. Similarly, if the inequality $u(p) \leq v(p)$ is satisfied for all $p \in \partial\Omega$, then $u(p) \leq v(p)$ for all points $p \in \Omega$.

Proof. We only consider the case $u(p) \geq v(p)$ for all $p \in \partial\Omega$. If we let $w = u - v$, then the function w is harmonic due to $\Delta w = \Delta u - \Delta v = 0$. Furthermore, for all $p \in \partial\Omega$ we have the inequality $w(p) \geq 0$. We now distinguish two cases. If the function w is constant, then clearly $w(p) \geq 0$ for all $p \in \overline{\Omega}$, since the inequality holds on the boundary. On the other hand, if the function w is not constant, then according to Theorem 1.44(b) it achieves its minimum value at a point $p_{\min} \in \partial\Omega$. This furnishes for all $p \in \Omega$ the inequality $w(p) \geq w(p_{\min}) \geq 0$. Taken together, the two cases imply $w \geq 0$ on $\overline{\Omega}$, and the result follows. \square

With the comparison principle at hand, we can now provide a preliminary answer to the question of uniqueness of solutions to Laplace's equation. We have already seen that without imposing any additional constraints, the equation usually has infinitely many solutions on a domain $\Omega \subset \mathbb{R}^d$. However, if we impose Dirichlet boundary conditions, then any solution has to be unique, provided it exists. In fact, we can even consider an extension of Laplace's equation, where the zero right-hand side is replaced by an arbitrary function. More precisely, we have the following result.

Corollary 1.47 (Uniqueness of solutions). *Let $\Omega \subset \mathbb{R}^d$ be a bounded open domain, and let $f : \Omega \rightarrow \mathbb{R}$ and $g : \partial\Omega \rightarrow \mathbb{R}$ denote two given continuous functions. Assume further that the function $u : \overline{\Omega} \rightarrow \mathbb{R}$ is continuous, as well as twice differentiable in Ω , and that u solves the following boundary value problem:*

$$\begin{aligned} \Delta u &= f && \text{in } \Omega, \\ u &= g && \text{on } \partial\Omega. \end{aligned}$$

Then the solution u is uniquely determined. In other words, if v is any function with $\Delta v = f$ in Ω and $v = g$ on $\partial\Omega$, then $u = v$ on $\overline{\Omega}$.

Proof. Let v be another solution of the boundary value problem, and define $w = u - v$. Then w is a harmonic function due to $\Delta w = \Delta u - \Delta v = f - f = 0$. Moreover, on the boundary of Ω we have $w = g - g = 0$. If we now apply the comparison principle from Corollary 1.46 twice, this implies both $w \geq 0$ and $w \leq 0$ on $\overline{\Omega}$, i.e., we obtain $w = 0$ as desired. \square

The above result shows for example that the solutions depicted in Figure 1.5 are the unique solutions of their respective boundary value problems.

1.2.3 • The Transport Equation

The classical partial differential equations discussed so far have one aspect in common: Both are second-order equations that involve the Laplace operator with respect to the spatial variables. For our third example, we leave the realm of second-order problems and take a closer look at a specific first-order partial differential equation, called the transport equation.

Derivation of the transport equation

The transport equation describes the concentration $u(t, x)$ of a substance being transported in a moving fluid. For this introduction, we only consider the one-dimensional situation, and so both arguments t and x are real independent variables. To derive the transport equation in one dimension, consider a fluid flowing through a narrow tube with constant velocity a , containing a concentration $u(t, x)$ of dye at time t and cross section of the tube at x . Since the fluid flows with velocity a , the flow of dye in the rightward direction at the point x is given by $au(t, x)$. Thus the rate of change of mass of dye in a region from $x = x_1$ to $x = x_2$ is given by the difference between the flow in on the left and the flow out on the right. Therefore the rate of change of mass is given by

$$au(t, x_1) - au(t, x_2) \quad \text{for all } t \geq 0 \quad \text{and} \quad x_1 < x_2.$$

By the fundamental theorem of calculus,

$$au(t, x_1) - au(t, x_2) = -(au(t, x_2) - au(t, x_1)) = - \int_{x_1}^{x_2} au_x(t, x) dx. \quad (1.19)$$

The mass of dye between x_1 and x_2 is given by the integral of the concentration

$$M(t) = \int_{x_1}^{x_2} u(t, x) dx.$$

Differentiating this equation, the rate of change of the mass of dye in this region satisfies

$$\frac{dM}{dt}(t) = \int_{x_1}^{x_2} u_t(t, x) dx, \quad (1.20)$$

and equating (1.19) and (1.20) one obtains

$$\int_{x_1}^{x_2} u_t(t, x) dx = - \int_{x_1}^{x_2} au_x(t, x) dx.$$

If we differentiate both sides of this equation with respect to the upper limit of integration x_2 , then the fundamental theorem of calculus implies

$$u_t(t, x_2) = -au_x(t, x_2) \quad \text{for all } t \geq 0 \quad \text{and} \quad x_2 \in \mathbb{R},$$

and we therefore get the transport equation

$$u_t + au_x = 0 \quad \text{for all } t \geq 0 \quad \text{and} \quad x \in \mathbb{R}, \quad (1.21)$$

where the initial concentration of the substance is given by

- *Initial condition for the transport equation:* $u(0, x) = f(x)$.

The transport equation in the above form is probably the easiest partial differential equation, and we will have no problem finding its general solution. A slightly more complicated but natural extension occurs if velocity a in the equation is not constant, but actually depends on time t and/or the location x . This leads to the general transport equation

$$u_t + a(t, x)u_x = 0 \quad \text{for all } t \geq 0 \quad \text{and} \quad x \in \mathbb{R}, \quad (1.22)$$

where $a : \mathbb{R}_0^+ \times \mathbb{R} \rightarrow \mathbb{R}$ is a given function. We will now show that both of these partial differential equations are easily solvable.

Solution of the simple transport equation

To begin with, consider the simple transport equation (1.21). We need to find a solution $u(t, x)$ such that $u_t + au_x = 0$. If we denote the gradient of the function u with respect to both variables t and x as $\nabla u(t, x) = (u_t(t, x), u_x(t, x))$, then the partial differential equation can be rewritten in the form

$$(1, a) \cdot \nabla u(t, x) = 0 \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \mathbb{R}.$$

This condition can be interpreted geometrically — it says that the gradient vector of our unknown solution u is always orthogonal to the fixed vector $(1, a) \in \mathbb{R}^2$. If one now recalls that the gradient vector of a smooth function is always orthogonal to the level curves of the function, we obtain that

- At every point $(t, x) \in \mathbb{R}_0^+ \times \mathbb{R}$, the level curve of the function u through this point is parallel to the vector $(1, a)$.

Since the vector $(1, a)$ is fixed, this fact indicates that the level curves of our unknown solution u have to be lines with slope a , which implies that the solution u has to be constant along such lines.

Being constant along lines is a condition that can readily be checked. For this, assume that $u(t, x)$ is a differentiable solution of the transport equation, and consider lines of the form

$$x = x_0 + at \quad \text{for } t \geq 0,$$

where $x_0 \in \mathbb{R}$ is arbitrary. Plugging this expression into the function u furnishes

$$v(t) = u(t, x_0 + at)$$

as a function of t . Note that this function is differentiable, and its derivative can be computed as

$$v'(t) = u_t(t, x_0 + at) + au_x(t, x_0 + at) = 0,$$

since u was assumed to be a solution of the transport equation. This indeed implies that the function v is constant, i.e., we have

$$v(t) = u(t, x_0 + at) = v(0) = u(0, x_0) \quad \text{for all } t \geq 0,$$

and with $x = x_0 + at$ this shows that

$$u(t, x) = u(0, x - at) \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \mathbb{R}.$$

In other words, the solution is of the form $u(t, x) = f(x - at)$ for $f(x) = u(0, x)$. Thus, we have proved the following result.

Lemma 1.48 (General solution of the simple transport equation). Consider the simple transport equation

$$u_t + au_x = 0 \quad \text{for all } t \geq 0 \quad \text{and } x \in \mathbb{R},$$

where the velocity $a \in \mathbb{R}$ is constant. Furthermore, suppose we are given a differentiable initial condition $f : \mathbb{R} \rightarrow \mathbb{R}$. Then the unique solution of the simple transport equation which satisfies

$$u(0, x) = f(x)$$

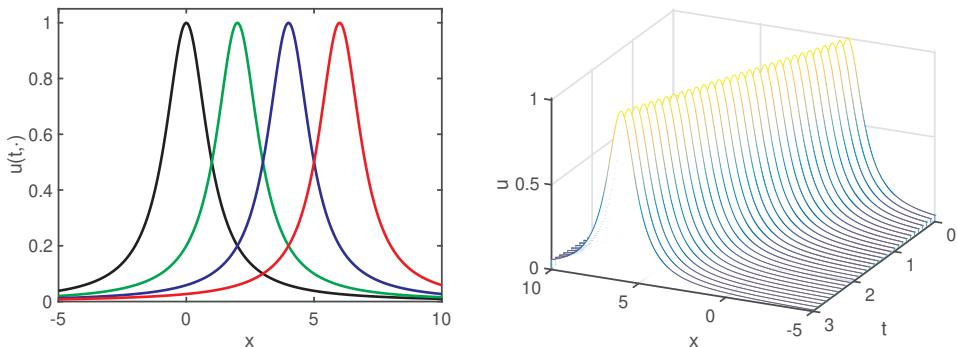


Figure 1.7. Solution of the simple transport equation $u_t + 2u_x = 0$ with initial condition $f(x) = 1/(1+x^2)$. The left image shows the solution snapshots $u(t, \cdot)$ for $t = 0, 1, 2, 3$ (resp. black, green, blue, red), the right image depicts the evolution of the solution u .

for all $x \in \mathbb{R}$ is given by

$$u(t, x) = f(x - at) \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \mathbb{R}.$$

This solution in fact also makes sense for negative values of t .

Notice that the key to establishing this result was the observation that the solution u has to be constant along lines of slope a . Such lines are called *characteristic lines*, and their generalizations will play an important role in the last chapter of this book. The above result is illustrated for the case $a = 2$ in Figure 1.7, where we use the specific initial condition $f(x) = 1/(1+x^2)$.

Solution of the general transport equation

We now turn our attention to the general transport equation $u_t + a(t, x)u_x = 0$ in (1.22), where the velocity a is actually a function of both time t and location x . As the basic form of the equation has not changed, the geometric argument of the last subsection is still valid, and it leads to the following conclusion:

- At every point $(t, x) \in \mathbb{R}_0^+ \times \mathbb{R}$, the level curve of the function u through this point is tangential to the vector $(1, a(t, x))$, i.e., if the level curve can be parameterized with respect to time t , it has slope $a(t, x)$ at the point (t, x) .

It is clear from this statement that we can no longer expect the level curves of u to be lines. Instead, they have to line up with the direction field provided by the function $a(t, x)$ — and this in turn means that the level curves would have to be solutions of the ordinary differential equation $\dot{x} = a(t, x)$. We can formalize this observation in the following result.

Lemma 1.49 (Solution of the general transport equation). Consider the general transport equation

$$u_t + a(t, x)u_x = 0 \quad \text{for all } t \geq 0 \quad \text{and} \quad x \in \mathbb{R}, \quad (1.23)$$

where the variable velocity $a : \mathbb{R}_0^+ \times \mathbb{R} \rightarrow \mathbb{R}$ is smooth. Furthermore, let $\varphi : I \rightarrow \mathbb{R}$ denote a solution of the associated ordinary differential equation

$$\dot{x} = a(t, x), \quad (1.24)$$

which is defined on some nontrivial interval $I \subset \mathbb{R}_0^+$. Then any differentiable solution u of the general transport equation (1.23) has to be constant on the graph of the function φ .

Proof. Consider the function $v : I \rightarrow \mathbb{R}$ defined by evaluating the solution u on the graph of φ , i.e., define

$$v(t) = u(t, \varphi(t)) \quad \text{for all } t \in I.$$

Then v is differentiable, and we obtain

$$v'(t) = u_t(t, \varphi(t)) + u_x(t, \varphi(t))\varphi'(t) = u_t(t, \varphi(t)) + u_x(t, \varphi(t))a(t, \varphi(t)) = 0,$$

where the first equality comes from the chain rule, the second since as a solution of (1.24) the function φ satisfies $\varphi'(t) = a(t, \varphi(t))$ on I , and the third from the fact that u was assumed to be a solution of (1.23). \square

Here is a brief example showing how the above result can be used. The example illustrates the so-called *method of characteristics* in the context of a transport equation with a velocity function a which is both time and space dependent.

Example 1.50. Consider the transport equation

$$u_t + (5 - x + e^{-2t})u_x = 0 \quad \text{for all } t \geq 0 \quad \text{and } x \in \mathbb{R}, \quad (1.25)$$

i.e., let $a(t, x) = 5 - x + e^{-2t}$ in Lemma 1.49. Then the associated initial value problem is given by

$$\dot{x} = 5 - x + e^{-2t}, \quad \text{with } x(t_0) = x_0.$$

This ordinary differential equation is linear, and therefore all solutions exist everywhere on \mathbb{R} . Sample solution curves are shown in Figure 1.8. Since this equation is nonhomogeneous, we find the general solution using the variation of parameters formula (1.3), which leads to

$$\varphi(t) = x_0 e^{-(t-t_0)} + \int_{t_0}^t e^{-(t-s)} (5 + e^{-2s}) ds.$$

For $t = 0$ this furnishes

$$\varphi(0) = x_0 e^{t_0} + \int_{t_0}^0 (5e^s + e^{-s}) ds = x_0 e^{t_0} + 4 - 5e^{t_0} + e^{-t_0}.$$

This shows that if u solves the transport equation (1.25) with $u(0, x) = f(x)$, then u is given by

$$u(t, x) = f(xe^t + 4 - 5e^t + e^{-t}) \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \mathbb{R}.$$

Notice that this solution is actually also valid for negative values of t . For the specific initial condition $f(x) = 1/(1+x^2)$ the unique solution

$$u(t, x) = \frac{1}{1 + (xe^t + 4 - 5e^t + e^{-t})^2}$$

is shown in Figure 1.9. \blacksquare

More examples can be found in the exercises accompanying this chapter. As in the example above, there are many situations in which studying the ordinary differential equation $\dot{x} = a(t, x)$ given in (1.24) makes it possible to derive explicit formulas for solutions

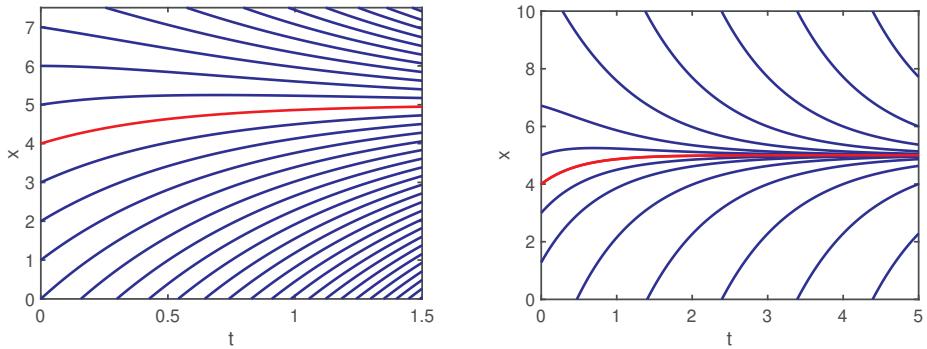


Figure 1.8. Sample solution curves of the ordinary differential equation associated to the general transport equation (1.25). In this case, the ordinary differential equation is given by $\dot{x} = 5 - x + e^{-2t}$, and the solutions are shown for two different (t, x) -regions. As $t \rightarrow \infty$, all solutions of this equation approach the particular solution $5 - e^{-2t}$, which is shown in red.

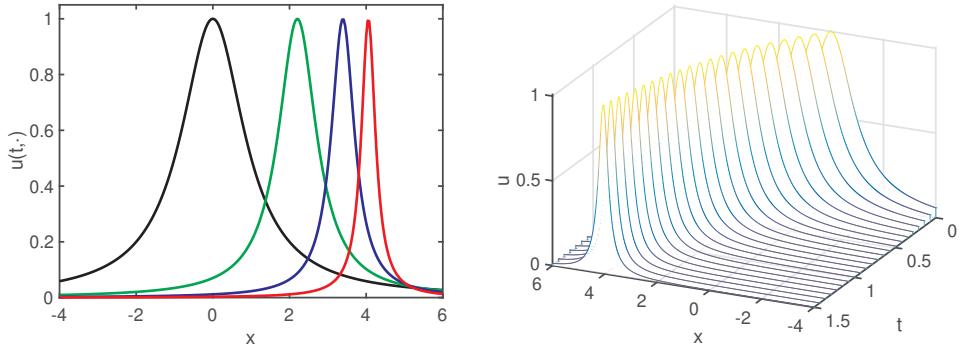


Figure 1.9. Solution of the transport equation $u_t + (5 - x + e^{-2t})u_x = 0$ with initial condition $f(x) = 1/(1 + x^2)$. The left image shows the snapshots $u(t, \cdot)$ for $t = 0, 1/2, 1, 3/2$ (resp. black, green, blue, red), and the right image depicts the evolution of the solution u .

of transport equations — and even if this is not possible one can usually derive qualitative properties of the solution behavior. In principle, the general procedure is as follows. Suppose we try to find a differentiable solution u of the transport equation

$$u_t + a(t, x)u_x = 0 \text{ on } \mathbb{R}_0^+ \times \mathbb{R}, \quad \text{with} \quad u(0, x) = f(x) \text{ on } \mathbb{R}.$$

Let $\varphi : I \rightarrow \mathbb{R}$ denote the maximal solution to the initial value problem

$$\dot{x} = a(t, x) \quad \text{with} \quad x(t_0) = x_0,$$

where $(t_0, x_0) \in \mathbb{R}^+ \times \mathbb{R}$ is arbitrary, but for the moment fixed. If 0 is contained in the existence interval I , then we have

$$u(t_0, x_0) = f(\varphi(0)).$$

Note that this procedure is not guaranteed to work. It is certainly possible that the maximal existence interval I of the solution φ does not contain zero, i.e., the solution φ could blow up in backward time. In this case, the function value $u(t_0, x_0)$ cannot be determined from the initial condition f .

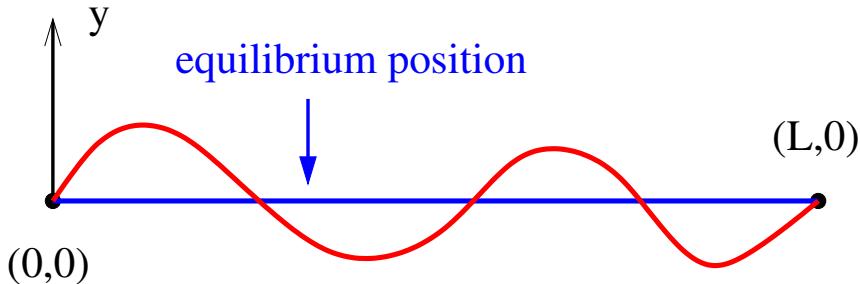


Figure 1.10. Coordinate system for the vibrating string.

Also note that above we consider a solution starting at (t_0, x_0) , for $t_0 > 0$, and follow the solution backward to see if it reaches $t = 0$. One could easily imagine starting instead at the point $(0, x_*)$ and following the solution forward to see if it blows up in finite time. However, forward blowup in finite time does not give information about existence of solutions. If a solution starting at $(0, x_*)$ blows up at time $t_* > 0$, one can only conclude that the initial value at point x_* does not influence the solutions at or beyond time t_* . Further discussion of this and other aspects of the method of characteristics appears in Chapter 7.

1.2.4 • The Wave Equation

With our last classical partial differential equation we return again to the case of second-order equations. In contrast to the case of the heat equation, in this case the temporal variable t appears in a second derivative.

Derivation of the wave equation

We now derive the wave equation in one dimension, in the context of the vibrating string problem. Specifically, a taut string is stretched between two pegs distance L apart, and given an initial displacement and velocity. The goal is to model its position as a function of space and time. A particular example of this is a plucked string of a guitar.

The string is at rest when it lies along the line segment between the two pegs. We now introduce a coordinate system such that the line segment between the two pegs is a horizontal line segment, and one of the pegs is located at the point $(0, 0)$. Since the pegs are distance L apart, this implies that the second peg is located at the point $(L, 0)$. Thus the rest position of the string is given by the set of all points (x, y) with $y = 0$ and $0 < x < L$. When not at rest, the position of the string is given by a set of points (x, y) in the plane. The position of the string is time-dependent, but the left end of the string always remains fixed at the left peg $(x, y) = (0, 0)$, and the right end is fixed at the right peg $(x, y) = (L, 0)$. See Figure 1.10.

We assume that for the purposes of the string's position and velocity, we can use the *continuum hypothesis*, which says that the number of particles composing the string is sufficiently large that we can consider them as acting together as a continuous function. Thus for each t and s , where s is a distance along the string from the left end and t is the time, we have a cross section of the string given by the single point $(x(t, s), y(t, s))$, where x and y are both continuous functions.

The force along the length of the string due to stretching it from its length at its rest

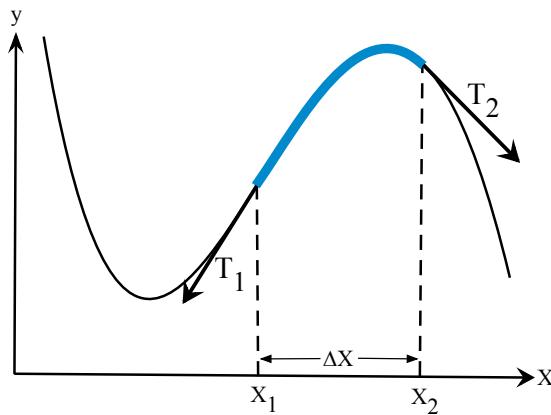


Figure 1.11. We consider a small piece of string between x_1 and x_2 . The net force on this piece of string due to tension is the sum of the tension vectors T_1 on the left and T_2 on the right.

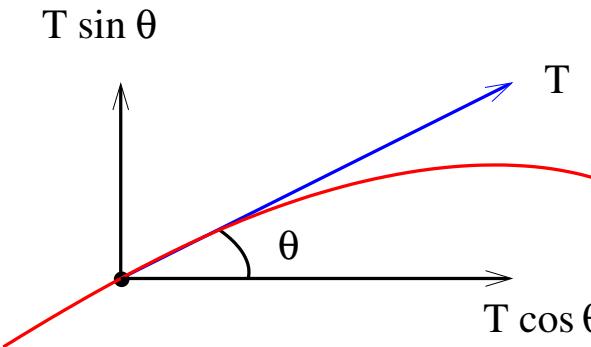


Figure 1.12. Tension of the string.

position is called the *tension*. According to Hooke's law, if a string of length L and constant cross section σ is stretched to the length $L + \Delta L$, then the magnitude of the tension is proportional to both the cross section σ and the ratio $\Delta L/L$. That is,

$$|T| = \frac{E\sigma\Delta L}{L}.$$

The proportionality constant, called the *Young's modulus* and denoted here by E , is a property of the material from which the string is made. For example, the Young's modulus of steel is up to 100 times the Young's modulus of nylon.

We now consider the force due to tension on a small piece of string S between x_1 and $x_2 = x_1 + \Delta x$. See Figure 1.11. The net tension on the piece of string is equal to the sum of the tension vectors T_1 and T_2 on the two ends of S . Note that the tensions on S at the two endpoints are always pointing away from S and are thus acting in opposite directions. Let θ denote the rightward angle between the string and the horizontal direction at a point along the string. That is,

$$\tan \theta = \frac{\partial y}{\partial x}.$$

The force on S due to tension is a force along the string pointing away from S . Thus as depicted in Figure 1.12, the horizontal and vertical components of the tension can be written in terms of θ . On the left-hand end of S the vector $T_1 = -|T_1|(\cos \theta(x_1), \sin \theta(x_1))$, and on the right-hand end of S the vector T_2 is given by $|T_2|(\cos \theta(x_2), \sin \theta(x_2))$.

As long as θ is small, we can use a linear approximation for all trigonometric functions. In particular, the linear approximations are

$$\cos \theta \approx 1, \quad \sin \theta \approx \theta, \quad \text{and} \quad \frac{\partial y}{\partial x} = \tan \theta \approx \theta.$$

Furthermore, if the displacement of the string is sufficiently small, we can approximate the amount of tension at each point in the string by a uniform constant, meaning that $|T_1| = |T_2|$ is a constant scalar value which we call $|T|$. Applying these approximations, the net vertical component of tension on S is given by

$$T_{\text{vertical}} = |T| \left(\frac{\partial y}{\partial x}(x_2) - \frac{\partial y}{\partial x}(x_1) \right).$$

The net horizontal force on S due to tension is

$$T_{\text{horizontal}} = |T| - |T| = 0.$$

That is, there is no net tension in the horizontal direction. This implies that there is no motion in the horizontal direction. Thus, we can drop the variable s , and y can be written as a function of t and x .

We now find the mass of the piece of string S . We assume that the string has a constant density ρ and constant cross-sectional area σ . By these assumptions, the mass of the string between x_1 and $x_2 = x_1 + \Delta x$ is equal to

$$m = \rho \sigma \Delta x.$$

The vertical component of the acceleration of the string is given by

$$a(t, x) = \frac{\partial^2 y}{\partial t^2}(t, x).$$

Using the physical law $F = ma$ on the vertical displacement, we get

$$F = |T| \left(\frac{\partial y}{\partial x}(t, x_2) - \frac{\partial y}{\partial x}(t, x_1) \right) = ma = \rho \sigma \Delta x \frac{\partial^2 y}{\partial t^2}.$$

Letting

$$c^2 = \frac{|T|}{\rho \sigma},$$

rearranging terms, and taking the limit as $\Delta x \rightarrow 0$ we get

$$\frac{\partial^2 y}{\partial t^2} = c^2 \lim_{\Delta x \rightarrow 0} \frac{\frac{\partial y}{\partial x}(t, x_2) - \frac{\partial y}{\partial x}(t, x_1)}{\Delta x} = c^2 \frac{\partial^2 y}{\partial x^2}.$$

Thus, the vertical displacement of a string due to the force of tension is governed by the wave equation:

$$\frac{\partial^2 y}{\partial t^2}(t, x) = c^2 \frac{\partial^2 y}{\partial x^2}(t, x), \quad \text{for } 0 < x < L,$$

along with the boundary conditions

$$y(t, 0) = y(t, L) = 0.$$

Using the same terminology as given for the heat equation, these boundary conditions are called *homogeneous Dirichlet boundary conditions*. Finally, even though in this section the way we used the variables y and x was naturally caused by their interpretation as coordinates in the plane, from now on we will replace the dependent variable y by u , as we did for both the heat and the Laplace equation.

Initial and boundary conditions

While in the previous section we only derived the wave equation in one spatial dimension, there is also a higher-dimensional analogue which models oscillation and vibration phenomena. To present this more general situation, let $\Omega \subset \mathbb{R}^d$ denote again a bounded open domain, where for now the reader can certainly think of the cases $d = 1, 2, 3$. In this subsection, the variable $x \in \mathbb{R}^d$ will therefore generally denote a vector of the form $x = (x_1, \dots, x_d)$, while the temporal variable t is still real. In this general setting, the wave equation can be written as

$$u_{tt} = c^2 \Delta u,$$

where the Laplacian Δ only acts on the spatial variables, i.e., we have

$$\Delta u = u_{x_1 x_1} + u_{x_2 x_2} + \dots + u_{x_d x_d}.$$

As in the previous examples, the wave equation needs to be supplemented by further conditions, in order to pin down a unique solution. The most commonly used boundary conditions are defined as follows.

- *Dirichlet boundary conditions:* Assume that $f = f(t, x)$ is a function which is defined on the set $\mathbb{R}^+ \times \partial\Omega$. Then we require that

$$u = f \quad \text{on} \quad \mathbb{R}^+ \times \partial\Omega.$$

More precisely, we require that $u(t, x) = f(t, x)$ for all points $(t, x) \in \mathbb{R}^+ \times \partial\Omega$.

- *Neumann boundary conditions:* Assume that a function $g : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ is given. Then we require

$$\frac{\partial u}{\partial n} = g \quad \text{on} \quad \mathbb{R}^+ \times \partial\Omega,$$

where $\partial u / \partial n$ is defined in Definition 1.33. Notice that as in the case of the Dirichlet boundary conditions, the function g can depend on time t .

- *Robin boundary conditions:* The above two types of boundary conditions can in some sense be combined. For this, let $h : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ be a given function, and let $a, b \in \mathbb{R}$ be real numbers. Then we require

$$au + b \frac{\partial u}{\partial n} = h \quad \text{on} \quad \mathbb{R}^+ \times \partial\Omega.$$

- *Periodic boundary conditions:* As in the cases of the heat equation and the Laplace equation, for periodic boundary conditions the domain has to be rectilinear in the form $\Omega = (a_1, b_1) \times \dots \times (a_d, b_d)$, and we require

$$\begin{aligned} u|_{x_k=a_k} &= u|_{x_k=b_k} \quad \text{and} \\ u_{x_k}|_{x_k=a_k} &= u_{x_k}|_{x_k=b_k} \quad \text{for } k = 1, \dots, d. \end{aligned}$$

Note that also in this formulation, we obtain conditions for all $t > 0$ and all points on the boundary of Ω .

- *Homogeneous boundary conditions:* For the first three types of boundary conditions above, if we have $f = 0$, $g = 0$, or $h = 0$, then the boundary conditions are called homogeneous Dirichlet, homogeneous Neumann, or homogeneous Robin boundary conditions, respectively.

We need to specify the state of the system at the initial time $t = 0$. However, since the wave equation has a second derivative with respect to time t , the initial condition must now specify both the initial displacement and the initial velocity. More precisely, we have the following.

- *Initial conditions:* The initial conditions for the wave equation are specified in the form

$$u(0, x) = u_0(x) \quad \text{and} \quad u_t(0, x) = v_0(x) \quad \text{for all } x \in \Omega,$$

where u_0 and v_0 are two given real-valued functions which are defined on Ω .

In general, supplying the wave equation with these initial and suitable boundary conditions will lead to a well-posed problem, i.e., we will usually obtain a uniquely determined solution.

Fundamental frequency and harmonics

Before moving on, we would like to present a few specific solutions in the one-dimensional situation. While the next chapter contains analytical ways to find certain closed form solutions to the wave equation using separation of variables, for now we state these solutions without worrying about how they were found. Consider the one-dimensional domain $\Omega = (0, L)$. Then for $n \in \mathbb{N}$ the functions

$$u(t, x) = \sin \frac{n\pi x}{L} \cos \frac{cn\pi t}{L}$$

are solutions to the problem of the vertical displacement of a vibrating string with pinned ends $u(t, 0) = u(t, L) = 0$, i.e., to the one-dimensional wave equation with homogeneous Dirichlet boundary conditions. To verify this, let n be fixed and let u be defined as above. Clearly we have both $u(t, 0) = 0$ and $u(t, L) = 0$. We also see that $u_{tt} = -c^2(n\pi/L)^2 u$, and $u_{xx} = -(n\pi/L)^2 u$. Therefore $u_{tt} = c^2 u_{xx}$, and our verification is complete. Notice that we did not impose any initial conditions, but one can compute directly that $u(0, x) = \sin(n\pi x/L)$ and $u_t(0, x) = 0$.

We now look more carefully at these solutions. Setting $n = 1$ in the above formula for u gives a periodic solution with period

$$\frac{2L}{c}$$

and frequency

$$\nu = \frac{c}{2L}.$$

The frequency ν is the lowest frequency possible for any n , and is therefore called the *fundamental frequency*. For $n = 2$, the solution has frequency

$$\frac{c}{L}.$$

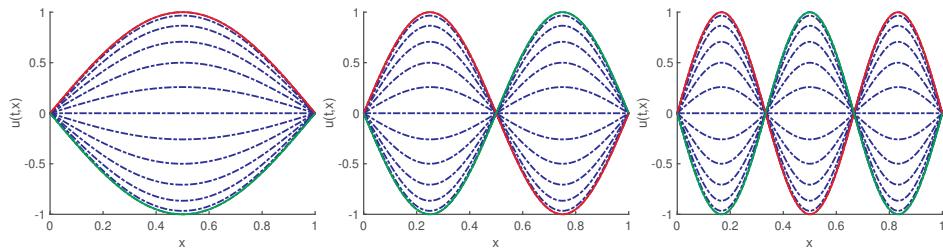


Figure 1.13. The vibrating string for $n = 1, 2, 3$. The red curve shows the solution at time zero. The green curve shows the solution after a half period.

Therefore it oscillates at twice the speed of the $n = 1$ solution. In music, if a pitch is double the frequency of another pitch, then they are said to differ by one *octave*. Thus the $n = 2$ solution has a frequency one octave above the fundamental. It is the second to lowest frequency possible, and is referred to as the *first harmonic*. For $n = 3$, the frequency is

$$\frac{3c}{2L}.$$

This is 1.5 octaves above the fundamental. It is called the second harmonic. This progression for all $n \in \mathbb{N}$ gives the full sequence of harmonics. See Figure 1.13.

On a guitar, the sequence of harmonics is the full set of notes that can be played on an open string without fingering. It may seem counterintuitive to be able to play a frequency other than the fundamental without fingering. Playing harmonics on an open guitar string would involve simultaneously displacing the string at multiple points. As a more intuitive example of harmonics, the same principle applies to wind instruments, where it is quite common to play multiple harmonics without changing the fingering. For example, a trumpet player routinely plays seven different pitches (the fundamental and first six harmonics) for each fixed fingering.

The principle of superposition

The wave equation obeys an important property that is shared by all linear differential equations subject to suitable boundary conditions. Rather than formulating the property for the most general situation, we only consider the one-dimensional wave equation on a bounded domain, and with homogeneous Dirichlet boundary conditions.

Lemma 1.51 (Superposition of solutions to the wave equation). *Assume that both $u(t,x)$ and $v(t,x)$ are solutions for the wave equation on the one-dimensional domain $\Omega = (0,L)$, subject to homogeneous Dirichlet boundary conditions. Then for arbitrary real numbers a and b the linear combination*

$$w(t,x) = au(t,x) + bv(t,x)$$

is also a solution for the wave equation with homogeneous Dirichlet boundary conditions.

Proof. In order to prove the result, both the boundary conditions and the validity of the partial differential equation have to be verified. As for the boundary conditions, we only consider the point $x = 0$ as the calculation is identical for $x = L$. This furnishes

$$w(t,0) = au(t,0) + bv(t,0) = a \cdot 0 + b \cdot 0 = 0.$$

As for the partial differential equation, let (t, x) be arbitrary. Then we have

$$\begin{aligned} w_{tt}(t, x) &= au_{tt}(t, x) + bv_{tt}(t, x) \\ &= ac^2u_{xx}(t, x) + bc^2v_{xx}(t, x) \\ &= c^2(au_{xx}(t, x) + bv_{xx}(t, x)) \\ &= c^2w_{xx}(t, x), \end{aligned}$$

which completes the proof. \square

The principle of superposition applies to any homogeneous linear partial differential equation with any homogeneous or periodic boundary conditions. The general form of this principle is stated in detail in Lemma 1.63.

In the context of the vibrating string, the principle of superposition implies that any mixture of the harmonics is a solution to the vibrating string problem. For example, the following is a solution

$$u(t, x) = 5 \sin \frac{\pi x}{L} \cos \frac{c\pi t}{L} + 3 \sin \frac{2\pi x}{L} \cos \frac{2c\pi t}{L} + \sin \frac{3\pi x}{L} \cos \frac{3c\pi t}{L},$$

as is any other linear combination of the harmonics. Such solutions will be studied more systematically when we introduce separation of variables.

Energy conservation for the wave equation

Assume that $u(t, x)$ is a solution to the vibrating string problem $u_{tt} = c^2 u_{xx}$ with homogeneous Dirichlet boundary conditions $u(t, 0) = u(t, L) = 0$. Define the function

$$E(t) = \int_0^L (u_t(t, x)^2 + c^2 u_x(t, x)^2) dx.$$

This function is known as the energy for the wave equation. The first term in the integral accounts for the kinetic energy of the string, while the second term represents potential energy. As it turns out, this energy has a fundamental property: The function $E(t)$ is constant with respect to time t for every fixed solution u . In order to show that this is the case, we let u denote an arbitrary solution and compute the derivative dE/dt . If this derivative vanishes, we are done. Taking the formula for the energy and differentiating with respect to t furnishes

$$\begin{aligned} \frac{dE}{dt}(t) &= \frac{d}{dt} \int_0^L (u_t(t, x)^2 + c^2 u_x(t, x)^2) dx \\ &= \int_0^L (2u_t(t, x)u_{tt}(t, x) + 2c^2 u_x(t, x)u_{xt}(t, x)) dx \\ &= 2 \int_0^L u_t(t, x)u_{tt}(t, x) dx + 2c^2 \int_0^L u_x(t, x)u_{xt}(t, x) dx. \end{aligned}$$

Using integration by parts, we see that

$$\begin{aligned} \int_0^L u_x(t, x)u_{xt}(t, x) dx &= u_x(t, L)u_t(t, L) - u_x(t, 0)u_t(t, 0) \\ &\quad - \int_0^L u_{xx}(t, x)u_t(t, x) dx. \end{aligned}$$

Note that since $u(t,0) = u(t,L) = 0$ is independent of t , their derivatives with respect to t satisfy $u_t(t,0) = u_t(t,L) = 0$ — and the boundary terms in the above formula drop out of the equation. This implies

$$\begin{aligned}\frac{dE}{dt}(t) &= 2 \int_0^L (u_t(t,x)u_{tt}(t,x) - c^2 u_t(t,x)u_{xx}(t,x)) dx \\ &= 2 \int_0^L u_t(t,x)(u_{tt}(t,x) - c^2 u_{xx}(t,x)) dx,\end{aligned}$$

but since u satisfies the wave equation, the expression in parentheses is zero. Therefore

$$\frac{dE}{dt}(t) = 0,$$

and $E(t)$ is a fixed constant independent of t . In other words, solutions to the wave equation conserve the energy E . This does not mean that the kinetic energy stays constant. Rather, there is a constant exchange between kinetic and potential energy, but without any energy loss.

Traveling wave solutions and finite propagation speed on the whole real line

So far, we have concentrated on the wave equation restricted to bounded domains Ω . In this subsection, we consider the unbounded case $\Omega = \mathbb{R}$. Consider the function

$$u(t,x) = f(x - ct),$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is twice differentiable. Then

$$u_{tt} = c^2 f''(x - ct) \quad \text{and} \quad u_{xx} = f''(x - ct),$$

and therefore one has

$$u_{tt} = c^2 u_{xx}.$$

The function u satisfies the initial condition

$$u(0,x) = f(x), \quad \text{and} \quad \frac{\partial u}{\partial t}(0,x) = -cf'(x).$$

This function $u(t,x)$ is called a *traveling wave solution* of the wave equation since it has the property that the solution is displaced rigidly (without a change of shape) rightwards with velocity c . See Figure 1.14.

The solution $u(t,x)$ constructed in the last paragraph is a solution on the whole real line, and for a general initial condition for the displacement $u(0,\cdot)$. But it does require a very strange initial condition for the initial velocity, which is related to the initial displacement f . Note, however, that by the same reasoning, the function $f(x + ct)$ solves the wave equation, and this gives a wave traveling to the left. Since the wave equation satisfies the superposition principle, also the function

$$u(t,x) = \frac{f(x - ct) + f(x + ct)}{2}$$

satisfies the wave equation, this time for the initial data

$$u(0,x) = f(x), \quad \text{and} \quad \frac{\partial u}{\partial t}(0,x) = 0.$$

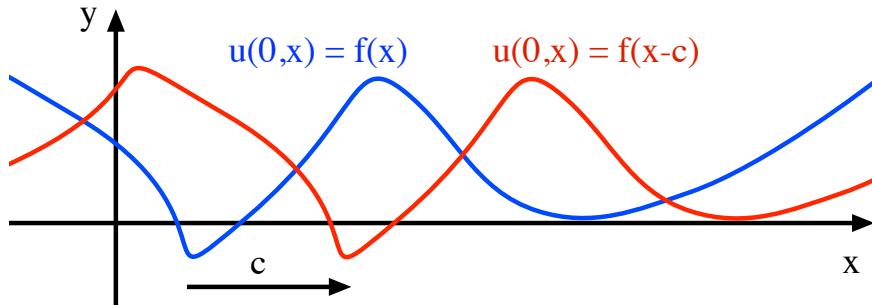


Figure 1.14. Traveling wave solutions. The function $u(t, x) = f(x - ct)$ is a traveling wave solution to the wave equation, since over time, the solution is rigidly displaced rightwards with constant velocity c .

In other words, we have found a solution to the wave equation on the real line which realizes an arbitrary initial displacement, and which starts with initial velocity identical to zero.

Is it possible to reverse the role of initial displacement and velocity, and find a solution v of the wave equation with initial velocity given by a smooth function $g(x)$, but with zero initial displacement? For this, consider the function

$$v(t, x) = \frac{1}{2c} \int_{x-ct}^{x+ct} g(\xi) d\xi.$$

Then we obviously have the identity $v(0, x) = 0$, and the partial derivatives of v can be computed as

$$\begin{aligned} \frac{\partial v}{\partial t}(t, x) &= \frac{g(x+ct) \cdot c - g(x-ct) \cdot (-c)}{2c} = c \cdot \frac{g(x+ct) + g(x-ct)}{2c} \quad \text{and} \\ \frac{\partial^2 v}{\partial t^2}(t, x) &= c \cdot \frac{g'(x+ct) \cdot c + g'(x-ct)(-c)}{2c} = c^2 \cdot \frac{g'(x+ct) - g'(x-ct)}{2c}, \end{aligned}$$

as well as

$$\begin{aligned} \frac{\partial v}{\partial x}(t, x) &= \frac{g(x+ct) - g(x-ct)}{2c} \quad \text{and} \\ \frac{\partial^2 v}{\partial x^2}(t, x) &= \frac{g'(x+ct) - g'(x-ct)}{2c}. \end{aligned}$$

These formulas immediately imply that v solves the wave equation, and in addition we have $v_t(0, x) = g(x)$.

As a final step, recall that the wave equation satisfies the superposition principle, i.e., the sum $u + v$ is also a solution — in fact a solution whose initial displacement is given by f , and its initial velocity by g . Thus, we have established the existence part of the following result.

Theorem 1.52 (d'Alembert's formula). Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a twice differentiable function, and let $g : \mathbb{R} \rightarrow \mathbb{R}$ be differentiable. Then there exists a unique solution to the wave equation

$$u_{tt} = c^2 u_{xx}$$

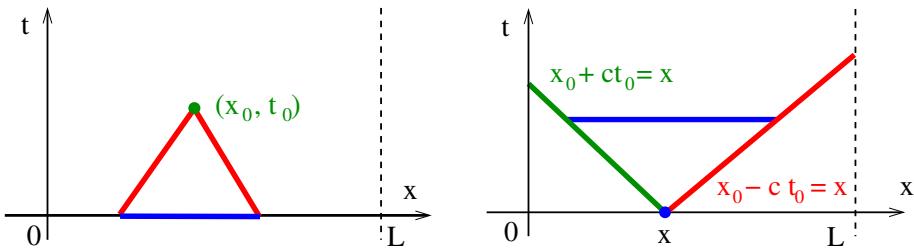


Figure 1.15. Characteristics and domain of dependence.

on the whole real line with initial conditions

$$u(0, x) = f(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(0, x) = g(x).$$

This solution is given by d'Alembert's formula

$$u(t, x) = \frac{f(x - ct) + f(x + ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} g(\xi) d\xi. \quad (1.26)$$

D'Alembert's formula allows us to derive one of the central properties of the wave equation, namely that of *finite speed of information propagation*. That is, consider a point $(t_0, x_0) \in \mathbb{R}_0^+ \times \mathbb{R}$. Our goal is to show that only specific portions of the initial conditions f and g are responsible for the value $u(t_0, x_0)$. According to (1.26), the function values $f(x_0 \pm ct_0)$ affect one part of the solution. In fact, no other part of f contributes to the function value $u(t_0, x_0)$ in the sense that changing $f(x)$ for $x \neq x_0 \pm ct_0$ would leave $u(t_0, x_0)$ unchanged. The left image in Figure 1.15 shows the corresponding x -values as the intersections of the two depicted red rays emanating from (t_0, x_0) with the x -axis. We now turn our attention to the influence of the second part of the initial data, the function g . According to (1.26), exactly the values $g(x)$ for $x_0 - ct_0 \leq x \leq x_0 + ct_0$ contribute to $u(t_0, x_0)$, i.e., exactly all x -values which are shown in blue in the left image of Figure 1.15. This blue region is called the *initial domain of dependence*.

We now reverse the above point of view. Given a value $x \in \mathbb{R}$, our new goal is to determine the values $u(t_0, x_0)$ affected by $f(x)$ or $g(x)$. Based on the discussion of the previous paragraph, the affected points are all points (t_0, x_0) which satisfy

$$x - ct_0 \leq x_0 \leq x + ct_0.$$

These points lie on a horizontal line segment between two rays emanating from $(0, x)$. The line segment is shown in blue, and the two rays are shown in green and red, respectively, in the right image of Figure 1.15. In other words, information from the initial conditions f and g propagates at a maximal speed c in the wave equation, and this speed is called the *propagation speed*. Furthermore, the green and red lines in the right image of Figure 1.15 are called *characteristics* for the wave equation. These objects will return later in our study of hyperbolic problems.

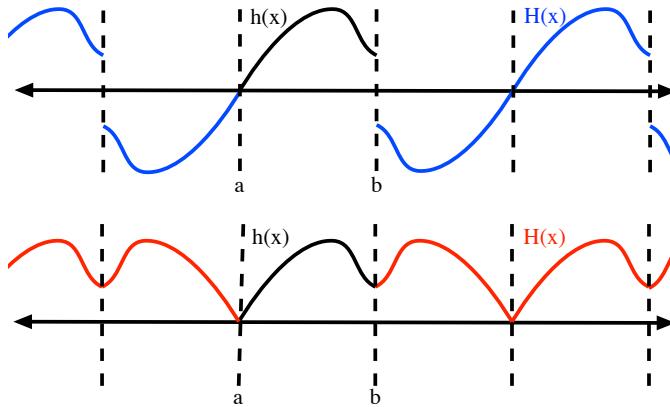


Figure 1.16. The odd extension and the even extension of function $h(x)$.

Propagation of information on finite domains

So far, we have only considered information propagation for solutions of the wave equation which are defined for all $x \in \mathbb{R}$. What about solutions to the finite domain problem

$$\begin{aligned} u_{tt} = c^2 u_{xx} \quad \text{in} \quad \Omega = (0, L), \quad \text{subject to} \quad u(t, 0) = 0, \quad u(t, L) = 0, \\ u(0, x) = f(x), \quad u_t(0, x) = g(x), \end{aligned} \quad (1.27)$$

where now the initial data consists of two functions $f, g : (0, L) \rightarrow \mathbb{R}$. It turns out, that in this special situation we can still construct the solution u using d'Alembert's formula. For this, we need the following definition, which will also prove to be useful on other occasions.

Definition 1.53 (Even and odd extensions). Consider a function $h : [a, b] \rightarrow \mathbb{R}$ which is given on a bounded interval $[a, b]$.

(a) A function $H : \mathbb{R} \rightarrow \mathbb{R}$ is called the odd extension of h onto \mathbb{R} , if H satisfies

$$H(a+x) = h(a+x) \text{ for } 0 \leq x \leq b-a, \quad (1.28)$$

$$H(b+x) = -h(b-x) \text{ for } 0 \leq x \leq b-a, \quad (1.29)$$

as well as

$$H(x + 2(b-a)) = H(x) \quad \text{for all } x \in \mathbb{R}. \quad (1.30)$$

(b) A function $H : \mathbb{R} \rightarrow \mathbb{R}$ is called the even extension of h onto \mathbb{R} , if H satisfies

$$H(a+x) = h(a+x) \text{ for } 0 \leq x \leq b-a,$$

$$H(b+x) = h(b-x) \text{ for } 0 \leq x \leq b-a,$$

as well as

$$H(x + 2(b-a)) = H(x) \quad \text{for all } x \in \mathbb{R}.$$

While the above definition might seem daunting at first glance, the underlying geometry is very simple. The top image in Figure 1.16 shows a function $h(x)$ defined on $[a, b]$

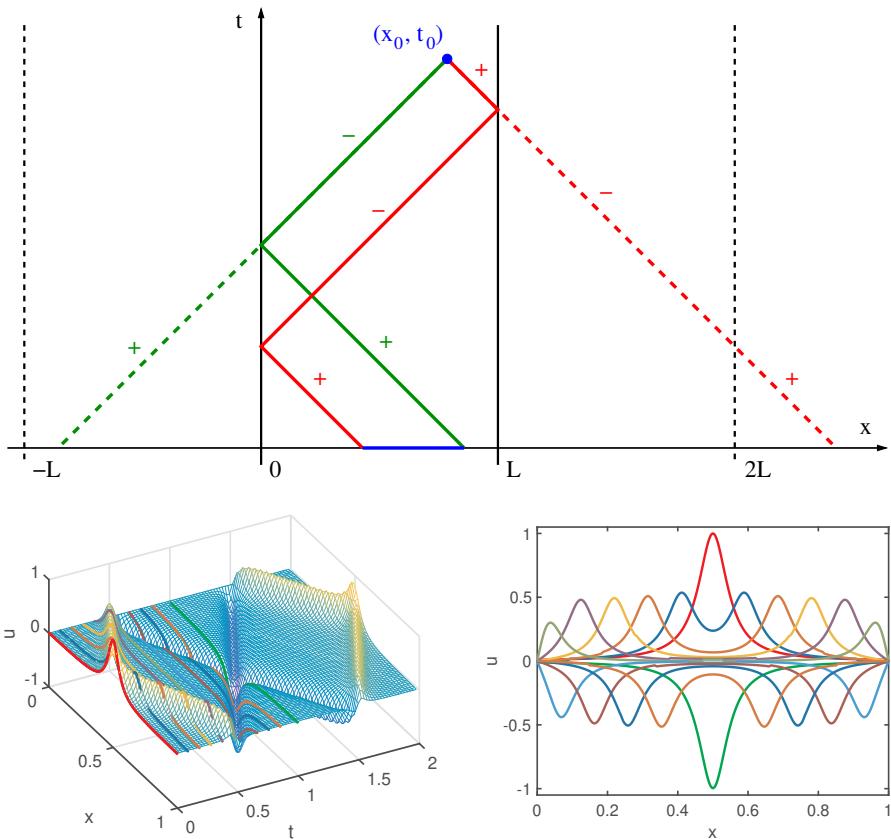


Figure 1.17. The top image shows that characteristics for the wave equation on a bounded domain bounce off the edge and change sign. The bottom left image depicts a solution to the wave equation with an initial condition having a single peak in the middle of the domain, with zero initial velocity. On the lower left, the solution is shown as a function of time and space, with fixed time solution curves superimposed for $t = 0$ (red), $t = 1$ (green), and nine evenly spaced times in between. The lower right image depicts these fixed time solution curves with the same color scheme as on left. This solution has a sign change along characteristics as it bounces off the edge, as depicted schematically in the upper image.

along with its odd extension. The bottom image of the figure shows the even extension of the same function $h(x)$. To explain this, we only consider part (a) for now, but part (b) will be of interest in later chapters. How does the extension H arise from h ? According to (1.28), the functions H and h agree on the interval $[a, b]$ — hence the notion of extension. The identity in (1.29) defines the function H on the interval $[b, b + (b - a)]$, i.e., it extends the original function to an interval of double length, by reflecting the graph of h about the point $(b, 0)$. Thus, the formulas (1.28) and (1.29) uniquely extend h to the interval $[a, a + 2(b - a)]$, which has length $2(b - a)$. Finally, identity (1.30) implies that H is periodic on all of \mathbb{R} with period $2(b - a)$. Altogether, the function H is obtained by an odd reflection, followed by a periodic extension. For example, if we apply this procedure to $h(x) = \sin x$ on $[a, b] = [0, \pi]$, then $H(x) = \sin x$ on \mathbb{R} . Similarly one can show that Definition 1.53(b) extends h to \mathbb{R} by an initial even reflection, followed by periodic extension.

We now return to our wave equation (1.27). Let F and G denote the odd extensions of f and g respectively, from $[0, L]$ to all of \mathbb{R} . Note also that due to the boundary conditions, we do assume that both f and g vanish at the points $x = 0$ and $x = L$. Furthermore, define

$$u(t, x) = \frac{F(x - ct) + F(x + ct)}{2} + \frac{1}{2c} \int_{x-ct}^{x+ct} G(\xi) d\xi, \quad \text{for } t \geq 0, \quad x \in \mathbb{R}. \quad (1.31)$$

According to D'Alembert's formula (1.26), the function defined in this way is a solution to the wave equation for all $(t, x) \in \mathbb{R}_0^+ \times \mathbb{R}$. Furthermore, for $t = 0$ we have

$$u(0, x) = F(x) \quad \text{and} \quad u_t(0, x) = G(x) \quad \text{for all } x \in \mathbb{R}.$$

In addition, for all $t \geq 0$ we have

$$u(t, 0) = \frac{1}{2} (F(-ct) + F(ct)) + \frac{1}{2c} \int_{-ct}^{+ct} G(\xi) d\xi = 0,$$

as well as

$$u(t, L) = \frac{1}{2} (F(L - ct) + F(L + ct)) + \frac{1}{2c} \int_{-ct}^{+ct} G(L + \xi) d\xi = 0,$$

since both F and G are odd with respect to $x = 0$ and $x = L$ according to Definition 1.53. In other words, the formula (1.31) provides an explicit representation of the solution of the boundary value problem (1.27).

The solution representation (1.31) enables us to directly apply the discussion of the characteristics from before, see Figure 1.17. For a given point (t_0, x_0) the characteristic lines can be traced backwards to the initial conditions at time $t = 0$. However, since the initial data f and g have been extended to all of \mathbb{R} by odd reflections, we can equivalently consider the reflections of the characteristic lines at the boundaries $x = 0$ and $x = L$ — as long as we make note of the fact that every reflection reverses the sign of f or g . In this way, it is possible to determine the domain of dependence as a subset of $[0, L]$. Again, see Figure 1.17.

1.3 • Classification of Partial Differential Equations

The examples from the previous section have clearly demonstrated that partial differential equations can exhibit a wide variety of very different behaviors. On one level, this variety is to be expected, as partial differential equations arise as models for very different physical processes. But on another level, these differences pose a fundamental problem. Unlike ordinary differential equations which can basically be studied using a unified set of tools, the tools required for the study of partial differential equations usually depend very much on the type of the underlying equation. Thus, when confronted with a particular equation, one needs to be able to recognize which set of tools is most likely applicable. In order to facilitate this, several classification schemes have been devised over time, and in this section we present two of them. The first classification scheme divides partial differential equations into three basic classes — known as elliptic, hyperbolic, and parabolic — based on the general properties of their solutions. These classes are generalizations of the examples from the previous section. The second, much coarser classification is between linear and nonlinear problems.

Classification of partial differential equations can seem daunting at first, in particular for the case of higher-order equations. We therefore tread slowly and begin by considering linear second-order problems. Only after this do we formalize the classification method by introducing a new notation for partial derivatives, called the multi-index notation. This will allow us to distinguish different types of nonlinear partial differential equations, and to extend our classification to nonlinear second-order partial differential equations and to higher-order partial differential equations.

1.3.1 • Classification of Linear Second-Order Equations

In this section we give a method for dividing linear second-order partial differential equations into three basic classes — known as elliptic, hyperbolic, and parabolic. We have already discussed equations of each of these types in the previous section, but what if we are confronted with a new linear second-order model? We will see that assigning one of the above three labels to such an equation basically amounts to a problem in linear algebra, i.e., to finding the eigenvalues of a certain associated matrix. Since the notation involved for d -dimensional domains might be somewhat intimidating at first, we give a couple of simple examples to illustrate the method in an informal way.

Example 1.54. Consider the homogeneous linear second-order equation

$$u_{xx} = c^2 u_{yy} \quad \text{with} \quad c \neq 0,$$

which is the wave equation in one space dimension in slightly disguised form. The classification method consists of three basic steps:

- First of all, we bring all the terms of the differential equation to one side. For the wave equation, this becomes

$$u_{xx} - c^2 u_{yy} = 0.$$

- Secondly, create a matrix based only on the coefficients of the second-order partial derivatives in the following way. If c_{xx} is the coefficient of u_{xx} , c_{xy} the coefficient of u_{xy} , and c_{yy} the coefficient of u_{yy} , then this matrix is given by

$$\mathcal{A} = \begin{pmatrix} c_{xx} & c_{xy}/2 \\ c_{xy}/2 & c_{yy} \end{pmatrix}.$$

For the case of the wave equation this gives

$$\mathcal{A} = \begin{pmatrix} 1 & 0 \\ 0 & -c^2 \end{pmatrix}.$$

- In the last step, we compute the eigenvalues of the matrix \mathcal{A} . Recall that the number λ is an eigenvalue and the nonzero vector $v \neq 0$ is an eigenvector for the matrix \mathcal{A} as long as

$$\mathcal{A} - \lambda I \quad \text{is noninvertible, and} \quad \mathcal{A}v = \lambda v. \quad (1.32)$$

In general, eigenvalues can be either real or complex numbers. Note, however, that the matrix \mathcal{A} is symmetric, and for any symmetric matrix with real-valued entries, all of its eigenvalues are real. This fact is used in the following classification. If all the

eigenvalues of \mathcal{A} are nonzero and have the same sign, then the differential equation is called *elliptic*. If the eigenvalues are all nonzero but there exist eigenvalues with opposite signs, then the differential equation is called *hyperbolic*. Finally, if there is an eigenvalue of zero, then the differential equation is called *parabolic*. For the wave equation, the eigenvalues are the diagonal entries of \mathcal{A} , i.e., we have the two eigenvalues

$$\lambda = 1 \quad \text{and} \quad \lambda = -c^2 < 0.$$

Since these numbers are of opposite sign, we conclude that the wave equation is hyperbolic.

We would like to point out that this classification does not depend on the particular form of the rewritten equation. For example, if we would have rewritten the wave equation in the form $c^2 u_{yy} - u_{xx} = 0$, we would have still obtained eigenvalues of opposite signs, even though the matrix \mathcal{A} would have been different. ■

Example 1.55. Following the same steps as in the previous example, consider the linear partial differential equation

$$5u_{xx} + 20u_{yy} = 12u_x + 8u_{xy}.$$

If we bring all terms of the equation to one side we obtain

$$5u_{xx} + 20u_{yy} - 12u_x - 8u_{xy} = 0.$$

In this case, the matrix \mathcal{A} is given by

$$\mathcal{A} = \begin{pmatrix} 5 & -4 \\ -4 & 20 \end{pmatrix},$$

since we only involve the second-order derivatives in the formation of the matrix. The characteristic equation for the eigenvalues of \mathcal{A} is $\det(\lambda I - \mathcal{A}) = 0$. This gives

$$\lambda^2 - 25\lambda + 84 = (\lambda - 4)(\lambda - 21) = 0,$$

implying that the two eigenvalues are $\lambda = 4$ and $\lambda = 21$. Thus, both eigenvalues have the same sign, and the equation is elliptic. ■

Similarly one can show that the heat equation is an example of a parabolic partial differential equation.

With these simple examples of linear second-order partial differential equations in two variables in mind, we can now turn our attention to the case of higher-dimensional domains. Consider a base domain $\Omega \subset \mathbb{R}^d$. From now on, we denote points in the domain again by $x = (x_1, \dots, x_d)$. Then a linear second-order differential operator L on Ω can be applied to functions $u : \Omega \rightarrow \mathbb{R}$, and in its most general form it is given by

$$Lu = \sum_{1 \leq i \leq j \leq d} A_{i,j}(x) u_{x_i x_j} + \sum_{1 \leq i \leq d} B_i(x) u_{x_i} + C(x) u, \quad (1.33)$$

where $A_{i,j}$, B_i , and C are functions from Ω into \mathbb{R} , for all $1 \leq i \leq j \leq d$. This differential operator is linear, since the partial derivatives of u only occur linearly. It is second-order,

since only derivatives of u up to order two are present. The condition $i \leq j$ is made without loss of generality since we are assuming that u is twice continuously differentiable, meaning that for every choice of i, j we have $u_{x_i x_j} = u_{x_j x_i}$ ³. Notice, however, that the coefficient functions $A_{i,j}$, B_i , and C can be arbitrary functions of the vector $x \in \Omega$.

In order to classify partial differential equations which are generated by L , we associate an x -dependent matrix $\mathcal{A}(x)$ with the second-order derivatives of L by defining

$$\mathcal{A}(x) = \frac{1}{2} \begin{pmatrix} 2A_{1,1}(x) & A_{1,2}(x) & A_{1,3}(x) & \cdots & \cdots & A_{1,d}(x) \\ A_{1,2}(x) & 2A_{2,2}(x) & A_{2,3}(x) & & & \vdots \\ A_{1,3}(x) & A_{2,3}(x) & 2A_{3,3}(x) & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 2A_{d-1,d-1}(x) & A_{d-1,d}(x) \\ A_{1,d}(x) & \cdots & \cdots & \cdots & A_{d-1,d}(x) & 2A_{d,d}(x) \end{pmatrix}. \quad (1.34)$$

As described in the beginning of this section, the classification of second-order linear partial differential equations will depend only on the eigenvalues of this matrix. Note that due to its definition, the matrix $\mathcal{A}(x)$ is symmetric. Any symmetric matrix with real-valued entries has a full set of d eigenvectors and corresponding real eigenvalues, such that the eigenvectors are distinct, but the list of eigenvalues might contain repeat values. With this, we can now describe our basic classification trichotomy.

Definition 1.56 (Classification of second-order linear equations). Consider an open domain $\Omega \subset \mathbb{R}^d$, as well as the linear second-order partial differential operator L over Ω from (1.33), and let $\mathcal{A}(x)$ denote the $d \times d$ -matrix defined in (1.34). Finally, let

$$Lu = f \quad \text{for some } f : \Omega \rightarrow \mathbb{R} \quad (1.35)$$

denote the linear partial differential equation associated with L .

- The partial differential equation (1.35) is called elliptic at $x \in \Omega$, if all eigenvalues of the matrix $\mathcal{A}(x)$ are nonzero and have the same sign.
- The partial differential equation (1.35) is called parabolic at $x \in \Omega$, if at least one eigenvalue of the matrix $\mathcal{A}(x)$ is zero.
- The partial differential equation (1.35) is called hyperbolic at $x \in \Omega$, if all eigenvalues of the matrix $\mathcal{A}(x)$ are nonzero and there are both positive and negative eigenvalues.

If in either of the three cases above the stated property is satisfied for all $x \in \Omega$, then (1.35) is simply called elliptic, parabolic, or hyperbolic, respectively.

For the special case of linear second-order partial differential equations on domains $\Omega \subset \mathbb{R}^d$ the above definition gives a complete answer to the problem of grouping equations which share similar properties. In fact, the following rough description of the essential features of each class can be observed throughout the remainder of this book.

- Elliptic equations correspond to *stationary processes* in nature, and therefore, in general they do not contain an independent variable that is interpreted as time. Solutions of elliptic equations will generally be smooth, provided the coefficient functions of the differential operator are smooth.

³This result is sometimes called Clairaut's Theorem.

- Parabolic equations correspond to *diffusive processes*, i.e., processes that try to establish some sort of equilibrium over time. As such, one of the independent variables in a parabolic equation is usually interpreted as time. Examples include both heat diffusion and the diffusion of pollutants, and in general such equations have a pronounced smoothing effect. Even if the initial distribution of say a pollutant is not continuous, the discontinuities will disappear due to the averaging effect of diffusion.
- Hyperbolic equations correspond to *oscillatory processes*, such as wave motion or wave propagation. Also these equations usually have a distinguished independent variable which is interpreted as time. In contrast to the parabolic case, however, solutions of hyperbolic equations generally do not gain smoothness, in fact, they can develop singularities over time.

Several of the above properties are intimately tied to the notion of *characteristic surfaces* associated with a given partial differential equation, which will be introduced later in the context of hyperbolic equations. For now, however, we restrict ourselves to presenting a few examples.

Example 1.57. Recall from Example 1.54 that the wave equation $u_{tt} = c^2 u_{xx}$ in one space dimension, where $c \neq 0$, can be rewritten as $-u_{tt} + c^2 u_{xx} = 0$, and therefore we have

$$\mathcal{A}(t, x) = \begin{pmatrix} -1 & 0 \\ 0 & c^2 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Since the eigenvalues of $\mathcal{A}(t, x)$ are $-1 < 0$ and $c^2 > 0$, the wave equation is hyperbolic. In fact, it is hyperbolic for all (t, x) , since the matrix $\mathcal{A}(t, x)$ does not explicitly depend on (t, x) . ■

Example 1.58. Rewriting the heat equation $u_t = u_{xx}$ in one space dimension in the form $-u_t + u_{xx} = 0$ gives

$$\mathcal{A}(t, x) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Since the eigenvalues of $\mathcal{A}(t, x)$ are 0 and 1, the heat equation is parabolic. ■

Example 1.59. The above two examples can easily be generalized to higher-dimensional domains. If $\Omega \subset \mathbb{R}^d$, and if we write $x = (x_1, \dots, x_d)$ for points $x \in \Omega$, then the Laplace equation

$$\Delta u = \sum_{k=1}^d u_{x_k x_k} = 0$$

is elliptic, since the matrix $\mathcal{A}(x)$ is the identity matrix. Similarly, one can show that the higher-dimensional heat equation

$$-u_t + \sum_{k=1}^d u_{x_k x_k} = 0$$

is always parabolic, and the wave equation

$$-u_{tt} + \sum_{k=1}^d u_{x_k x_k} = 0$$

is always hyperbolic. In the latter two cases, the matrix \mathcal{A} depends on the $d + 1$ arguments (t, x_1, \dots, x_d) , and we have $\mathcal{A}(t, x_1, \dots, x_d) \in \mathbb{R}^{(d+1) \times (d+1)}$. In fact, in both cases the matrix is a diagonal matrix in which all but one diagonal entry is 1. The first entry is 0 for the heat equation (since the t term is first order), and -1 for the wave equation (since the t term is second order and has the opposite sign from the x_k terms). ■

Example 1.60. For the so-called Tricomi equation $x_2 u_{x_1 x_1} + u_{x_2 x_2} = 0$ we have

$$\mathcal{A}(x_1, x_2) = \begin{pmatrix} x_2 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

This equation changes its character based on the specific location in space. It is elliptic as long as $x_2 > 0$, parabolic for $x_2 = 0$, and hyperbolic for $x_2 < 0$. ■

1.3.2 • Linear and Nonlinear Equations

The classification of the last section into elliptic, parabolic, and hyperbolic equations is complete — as long as one considers only linear second-order partial differential equations. However, as we will see later in this book, many important models are given by higher-order equations. Moreover, the assumption of linearity is in most situations only a significant oversimplification of the actual true underlying model.

In this section we therefore turn to a second meaningful way in which partial differential equations can be classified — one which parallels similar classifications in many areas of mathematics, for example in the study of systems of algebraic equations. On the one hand, if algebraic systems exhibit the fundamental property of being linear, then one can easily classify how many solutions occur, and even develop algorithms such as Gaussian elimination which will always allow one to find these solutions. On the other hand, nonlinear equations in several variables are generally harder to study. There are no general statements about the existence and/or multiplicity of solutions, and in many cases one has to be content with finding approximate solutions. Likewise, in the study of differential equations, linear equations can be dealt with much more easily than nonlinear ones.

Linear equations

In the current section we introduce the necessary notation and definitions to distinguish linear from nonlinear equations. In order to present this classification in a concise way, and at the same time general enough to cover all of the types of partial differential equations which will be covered in this book, we first need to introduce a new notation for writing partial derivatives, which is called the *multi-index notation*.

Definition 1.61 (Multi-index notation). Let $d \in \mathbb{N}$ denote an arbitrary positive integer, and for each $k = 1, \dots, d$, let $\alpha_k \in \mathbb{N}_0$ denote an arbitrary nonnegative integer. Then a multi-index is a vector

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{N}_0^d$$

consisting of d non-negative integers $\alpha_1, \dots, \alpha_d$. For any such multi-index α we further define its absolute value $|\alpha|$ and its factorial $\alpha!$ via

$$|\alpha| = |\alpha_1| + |\alpha_2| + \dots + |\alpha_d| \quad \text{and} \quad \alpha! = \alpha_1! \alpha_2! \cdots \alpha_d!.$$

Now let $u : \Omega \rightarrow \mathbb{R}$ denote a sufficiently differentiable real-valued function which is defined on a subset $\Omega \subset \mathbb{R}^d$, and denote its argument vector by $x = (x_1, x_2, \dots, x_d)$. Then for any multi-index $\alpha \in \mathbb{N}_0^d$ we define an associated partial derivative of u via

$$D^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_d^{\alpha_d}}.$$

Finally, for any multi-index $\alpha \in \mathbb{N}_0^d$ and any vector $x \in \mathbb{R}^d$ we define

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}.$$

This last notation is needed for our upcoming discussion of classification of higher-order partial differential equations. Throughout this definition, an exponent of zero implies that the corresponding term x_i^0 equals one.

We would like to point out that in the above definition we always assume that the function u is at least $|\alpha|$ -times continuously differentiable, since only then are we assured that the specific order of the sequence of partial derivatives does not matter.

To illustrate the above definition, consider a smooth function $u : \mathbb{R}^3 \rightarrow \mathbb{R}$. Then we have for example

$$\begin{aligned} D^{(2,2,1)} u &= \frac{\partial^5 u}{\partial x_1^2 \partial x_2^2 \partial x_3} = u_{x_1 x_1 x_2 x_2 x_3}, & D^{(0,0,4)} u &= \frac{\partial^4 u}{\partial x_3^4} = u_{x_3 x_3 x_3 x_3}, \\ D^{(0,2,1)} u &= \frac{\partial^3 u}{\partial x_2^2 \partial x_3} = u_{x_2 x_2 x_3}, & D^{(1,0,1)} u &= \frac{\partial^2 u}{\partial x_1 \partial x_3} = u_{x_1 x_3}, \\ D^{(1,0,0)} u &= \frac{\partial u}{\partial x_1} = u_{x_1}, & D^{(0,0,0)} u &= u. \end{aligned}$$

While this notation might seem somewhat clumsy at first sight, it will prove to be very useful in a number of cases. The first such case is the following central definition.

Definition 1.62 (Linear partial differential equation). A partial differential equation for an unknown function u depending on d independent variables x_1, \dots, x_d is called a linear partial differential equation if it can be written in the form

$$Lu = f, \quad \text{where} \quad Lu = L(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u, \quad (1.36)$$

for coefficient functions $a_\alpha(x)$ and a right-hand side $f(x)$, where $x = (x_1, \dots, x_d)$. In this case, we call $L = L(x, D)$ the differential operator associated with the above linear partial differential equation.

The partial differential equation (1.36) is called homogeneous if we have $f \equiv 0$, otherwise it is called nonhomogeneous. Furthermore, the order of the differential equation is the largest value $|\alpha|$ for which $a_\alpha \neq 0$.

Even though the definition is formulated for functions of the form $u = u(x_1, \dots, x_d)$, one can always rename the independent variables. Thus, the above definition also applies to functions $u = u(t, x)$, etc., in the obvious way. Notice also that in the definition we introduced two abbreviations L and $L(x, D)$ for the differential operator on the left-hand

side of the linear equation. The second of these will prove to be useful for our discussion of the symbol associated with a differential operator.

Looking back at the previous section we see that all of the partial differential equations that we have encountered so far are linear. For example, the one-dimensional heat equation for a function $u = u(t, x)$ can be rewritten as the second-order linear partial differential equation

$$Lu = u_t - u_{xx} = D^{(1,0)}u - D^{(0,2)}u = 0.$$

More generally, we can also recover linear second-order partial differential equations as described in the last section. For this, consider a base domain $\Omega \subset \mathbb{R}^d$. Then according to Definition 1.62, a second-order linear differential operator can be written as

$$Lu = L(x, D)u = \sum_{|\alpha| \leq 2} a_\alpha(x) D^\alpha u, \quad u : \Omega \rightarrow \mathbb{R}, \quad (1.37)$$

where $a_\alpha : \Omega \rightarrow \mathbb{R}$. At first glance, this form seems to differ from the one given in (1.33). To reunite both representations, one can proceed as follows. For any integer $i = 1, \dots, d$, let $e_i \in \mathbb{R}^d$ denote the i -th canonical basis vector, i.e., we have $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ with the entry 1 exactly in the i -th component. Furthermore, for any $1 \leq i \leq j \leq d$ define

$$A_{i,j}(x) = a_{e_i + e_j}(x), \quad B_i(x) = a_{e_i}(x), \quad C(x) = a_0(x). \quad (1.38)$$

Then the differential operator in (1.37) can be rewritten as

$$Lu = L(x, D)u = \sum_{1 \leq i \leq j \leq d} A_{i,j}(x) u_{x_i x_j} + \sum_{1 \leq i \leq d} B_i(x) u_{x_i} + C(x) u,$$

which is exactly (1.33).

As we will see throughout this book, there are a number of analytical methods for linear differential equations, which in many cases provide explicit solution representations. Even more importantly, the set of all solutions of a given linear equation is closed under the formation of linear combinations. We had already introduced this fact in the context of the wave equation, but now we can formulate and prove it for the general case.

Lemma 1.63 (Superposition principle). *Consider a linear differential operator L as in Definition 1.62, and let $f_1 = f_1(x)$ and $f_2 = f_2(x)$ denote two right-hand sides. Furthermore, assume that u_1 and u_2 satisfy the linear differential equations*

$$Lu_1 = f_1 \quad \text{and} \quad Lu_2 = f_2.$$

Then for any real numbers $c_1, c_2 \in \mathbb{R}$ the linear combination

$$u = c_1 u_1 + c_2 u_2 \quad \text{solves the differential equation} \quad Lu = c_1 f_1 + c_2 f_2.$$

In particular, if u_1 and u_2 satisfy the homogeneous linear differential equation

$$Lu = 0,$$

then so does any linear combination of u_1 and u_2 .

Proof. This result follows immediately from the fact that differentiation is a linear operation. More precisely, we have

$$L(c_1 u_1 + c_2 u_2) = c_1 \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u_1 + c_2 \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u_2 = c_1 Lu_1 + c_2 Lu_2,$$

which together with $Lu_1 = f_1$ and $Lu_2 = f_2$ proves the result. \square

The superposition principle shows that the set of all solutions of a homogeneous linear partial differential equation is a vector space, and as such one can hope to find a suitable basis of solutions, from which all solutions can be easily generated. This can in fact be done, and we will be addressing this issue in considerable detail later on.

Nonlinear equations

We now turn our attention to nonlinear partial differential equations, i.e., partial differential equations which cannot be written in the form of Definition 1.62. For equations of this type, no general statements can be made about the structure of the solution space, and therefore we will usually study these equations in later chapters using numerical methods only. Nevertheless, not all nonlinear partial differential equations are created equal, and we subdivide them in the following way.

Definition 1.64 (Nonlinear partial differential equation). A nonlinear partial differential equation for an unknown function u which depends on d independent variables x_1, \dots, x_d is a partial differential equation which cannot be written in the form of Definition 1.62. Nonlinear partial differential equations are further divided into three subclasses.

1. A nonlinear partial differential equation is called semilinear, if it is of the form

$$\sum_{|\alpha|=m} a_\alpha(x) D^\alpha u + a_0(D^{m-1}u, D^{m-2}u, \dots, D^1u, u, x) = 0,$$

where we use the notation $x = (x_1, \dots, x_d)$, and we define

$$D^k u = \{D^\alpha u : |\alpha| = k\}$$

as the collection of all partial derivatives of u of order k . In the above form, we assume that at least one coefficient function a_α is not identically zero, for some $|\alpha| = m$.

2. A nonlinear partial differential equation is called quasilinear, if it is of the form

$$\begin{aligned} & \sum_{|\alpha|=m} a_\alpha(D^{m-1}u, D^{m-2}u, \dots, D^1u, u, x) D^\alpha u \\ & + a_0(D^{m-1}u, D^{m-2}u, \dots, D^1u, u, x) = 0, \end{aligned}$$

where again we assume that at least one coefficient function a_α is not identically zero for some $|\alpha| = m$.

3. A nonlinear partial differential equation is called fully nonlinear, if it is neither semilinear nor quasilinear.

Notice that the above classification of nonlinear equations is based completely on the behavior of the highest-order derivatives. A partial differential equation of order m is called semilinear, if the terms involving the partial derivatives of order m form a linear differential operator in the sense of Definition 1.62. This means in particular that the coefficient functions of this linear differential operator only depend on x . If instead the coefficient functions depend also on lower-order partial derivatives of u , then the equation is quasilinear. For example, the second-order *Allen-Cahn equation*

$$u_t = \varepsilon^2 u_{xx} + u - u^3, \quad \text{where } \varepsilon > 0 \text{ is a parameter,}$$

is a semilinear partial differential equation for the one-dimensional spatial variable x , as is the fourth-order *Cahn-Hilliard equation*

$$u_t = -(\varepsilon^2 u_{xx} + u - u^3)_{xx} = -\varepsilon^2 u_{xxxx} - u_{xx} + 3u^2 u_{xx} + 6u u_x^2.$$

On the other hand, the one-dimensional *Burgers' equation*

$$u_t + uu_x = 0$$

is quasilinear, while the d -dimensional *Eikonal equation*

$$u_{x_1}^2 + u_{x_2}^2 + \dots + u_{x_d}^2 = 1$$

is fully nonlinear. In general, the difficulty of a nonlinear partial differential equation increases in the order presented in Definition 1.64, i.e., semilinear equations are in general easier than quasilinear ones, and fully nonlinear equations are the most involved nonlinear problems. In particular, we now state a classification result for all second-order semilinear partial differential equations, and the next section will discuss higher-order linear and semilinear partial differential equations, but it is beyond the scope of this book to classify even second-order quasilinear or fully nonlinear partial differential equations.

Definition 1.65 (Classification of second-order semilinear equations). Consider a domain $\Omega \subset \mathbb{R}^d$ and a semilinear partial differential equation as in Definition 1.64(1). Define

$$A_{i,j}(x) = a_{e_i+e_j}(x)$$

as in (1.38), and create a $d \times d$ matrix $\mathcal{A}(x)$ as in (1.34). Then the semilinear equation is elliptic if the eigenvalues are nonzero and have the same sign. It is parabolic if at least one eigenvalue is zero. It is hyperbolic if the eigenvalues are nonzero and there are both positive and negative eigenvalues. That is, once the matrix $\mathcal{A}(x)$ is formed from the coefficients on the second-order terms, the classification follows exactly as in the linear case in Definition 1.56.

Using this definition, we see that the Allen-Cahn equation $u_t = \varepsilon^2 u_{xx} + u - u^3$ is parabolic, with the same matrix \mathcal{A} as the heat equation. The equation

$$\varepsilon^2(u_{x_1 x_1} + u_{x_2 x_2}) + u - u^3 = 0, \text{ where } \varepsilon > 0 \text{ is a parameter,}$$

known as the stationary Allen-Cahn equation, is elliptic with the same matrix \mathcal{A} as Laplace's equation. Finally, for any constants the equation

$$u_{tt} = k^2 u_{xx} + au + bu^2, \text{ where } a, b, \text{ and } k > 0 \text{ are parameters,}$$

known as the Klein-Gordon equation, is hyperbolic with the same matrix \mathcal{A} as the wave equation.

1.3.3 • A Deeper Look: Higher-Order Equations

In contrast to second-order linear and semilinear partial differential equations, classifying higher-order equations is considerably more difficult — and we will therefore only classify certain types of higher-order partial differential equations, i.e., the types which will occur later in the book. Yet even for this restricted class of partial differential equations, the classification can no longer proceed via a simple eigenvalue computation of an associated

matrix $\mathcal{A}(x)$. Instead, one has to make use of the notion of *symbol*, which is an associated polynomial function.

In order to introduce this concept, recall that any linear equation on a domain $\Omega \subset \mathbb{R}^d$ can be written in the form

$$Lu = f, \quad \text{where} \quad Lu = L(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u, \quad (1.39)$$

for coefficients $a_\alpha : \Omega \rightarrow \mathbb{R}$ and a right-hand side $f : \Omega \rightarrow \mathbb{R}$, where the argument vector is given by $x = (x_1, \dots, x_d) \in \Omega$. It will turn out to be extremely useful to associate the following polynomial function to the differential operator $L = L(x, D)$, which is called the symbol of L .

Definition 1.66 (Symbol of a differential operator). Consider a differential operator $L(x, D)$ on a domain $\Omega \subset \mathbb{R}^d$ as defined in (1.39), and assume that $a_\alpha \neq 0$ for some multi-index α with $|\alpha| = m$. Then the symbol associated with $L(x, D)$ is defined as the function

$$L(x, \xi) = \sum_{|\alpha| \leq m} a_\alpha(x) \xi^\alpha, \quad \text{for } x \in \Omega \quad \text{and} \quad \xi \in \mathbb{R}^d,$$

where we use the notation $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ for $\xi \in \mathbb{R}^d$ and any multi-index $\alpha \in \mathbb{N}_0^d$. In addition, we define the principal part of the symbol by

$$L^p(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x) \xi^\alpha, \quad \text{for } x \in \Omega \quad \text{and} \quad \xi \in \mathbb{R}^d.$$

In other words, the principal part of the symbol is only based on the highest-order derivatives of the differential operator.

One can easily see that if we define the principal part $L^p(x, D)$ of the differential operator $L(x, D)$ by

$$L^p(x, D)u = \sum_{|\alpha|=m} a_\alpha(x) D^\alpha u, \quad (1.40)$$

then the principal part of the symbol associated with $L(x, D)$ is exactly the symbol associated with the differential operator $L^p(x, D)$.

It is clear from the above definition that the symbol of a differential operator is a polynomial function of degree m in the variables ξ_1, \dots, ξ_d . At first glance, the introduction of the symbol of a differential operator must seem extremely arbitrary. However, from the point of view of Fourier transforms it is very natural. While we do not address Fourier transforms further in this book, we would like to mention that for a large class of functions $u : \mathbb{R}^d \rightarrow \mathbb{R}$ one can define their *Fourier transform* \hat{u} via

$$\hat{u}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} u(x) e^{-i\xi \cdot x} dx,$$

where $\xi \cdot x$ denotes the standard inner product of vectors in \mathbb{R}^d and i is the imaginary unit. It turns out that the Fourier transform uniquely characterizes the function u , i.e., there is a one-to-one correspondence between functions u and their Fourier transforms \hat{u} . The connection with the symbol of a differential operator can most easily be seen if we now assume that the coefficients a_α of the differential operator $L(x, D)$ defined in (1.39) are all

constant functions, i.e., they do not depend explicitly on x and we write $L(D) = L(x, D)$. In this case one can show that

$$L(i\xi)\hat{u}(\xi) \quad \text{is the Fourier transform of} \quad L(D)u(x).$$

In other words, the Fourier transform converts the application of a differential operator into a multiplication of complex-valued functions, and the respective factor is exactly the symbol of the differential operator evaluated at $i\xi$.

In order to motivate our later classification of higher-order equations, let us briefly return to the second-order case discussed in the previous section — but now using the notion of symbol. As before, we consider a base domain $\Omega \subset \mathbb{R}^d$, and on this domain a second-order linear differential operator of the form

$$Lu = L(x, D)u = \sum_{1 \leq i \leq j \leq d} A_{i,j}(x) u_{x_i x_j} + \sum_{1 \leq i \leq d} B_i(x) u_{x_i} + C(x) u, \quad (1.41)$$

see also (1.33). The principal part of this differential operator as defined in (1.40) is given by

$$L^p u = L^p(x, D)u = \sum_{1 \leq i \leq j \leq d} A_{i,j}(x) u_{x_i x_j}. \quad (1.42)$$

As we saw previously, the principal part of a linear second-order differential operator is the only part used in the definition of the matrix $\mathcal{A}(x)$. The next lemma illustrates the fundamental connection between this matrix and the principal part of the symbol associated with L .

Lemma 1.67 (Principal part of the symbol as quadratic form). *Consider a second-order linear partial differential operator $L(x, D)$ as in (1.41), and let $L^p(x, D)$ denote its principal part as in (1.42). Furthermore, define the symmetric matrix $\mathcal{A}(x)$ as in (1.34). Then the principal part of the symbol of $L(x, D)$ is given by the symmetric quadratic form*

$$L^p(x, \xi) = \xi^t \mathcal{A}(x) \xi,$$

where $x \in \Omega$ and $\xi \in \mathbb{R}^d$.

Proof. If we denote the entry of the matrix $\mathcal{A}(x)$ in the i -th row and j -th column by $\mathcal{A}_{i,j}(x)$, then we have

$$\mathcal{A}_{i,j}(x) = \begin{cases} A_{i,j}(x)/2 & \text{for } i < j, \\ A_{i,i}(x) & \text{for } i = j, \\ A_{j,i}(x)/2 & \text{for } i > j, \end{cases}$$

and this implies

$$\begin{aligned} \xi^t \mathcal{A}(x) \xi &= \sum_{i,j=1}^d \mathcal{A}_{i,j}(x) \xi_i \xi_j \\ &= \sum_{1 \leq i \leq d} A_{i,i}(x) \xi_i^2 + \sum_{1 \leq i < j \leq d} \frac{A_{i,j}(x)}{2} \xi_i \xi_j + \sum_{1 \leq j < i \leq d} \frac{A_{j,i}(x)}{2} \xi_i \xi_j \\ &= \sum_{1 \leq i \leq d} A_{i,i}(x) \xi_i^2 + \sum_{1 \leq i < j \leq d} A_{i,j}(x) \xi_i \xi_j = L^p(x, \xi), \end{aligned}$$

which furnishes the result. \square

As the reader may remember from linear algebra, symmetric quadratic forms such as the one above are most easily described using the d real eigenvalues of the defining symmetric matrix, in our case, the eigenvalues of the matrix $\mathcal{A}(x)$. If all of the eigenvalues of this matrix are strictly positive, then the quadratic form is called positive definite and we have $\xi^t \mathcal{A}(x) \xi > 0$ for all $\xi \neq 0$. Similarly, if all of the eigenvalues are strictly negative, then the quadratic form is called negative definite, and $\xi^t \mathcal{A}(x) \xi < 0$ for all $\xi \neq 0$. If the matrix has a zero eigenvalue, then the form is called degenerate, and if all eigenvalues are nonzero but have opposite signs, then the form is called indefinite. In the last case, the number $\xi^t \mathcal{A}(x) \xi$ can take any real value for $\xi \neq 0$.

In combination with Lemma 1.67, this discussion shows that a second-order linear partial differential equation associated with the operator $L(x, D)$ defined in (1.41) is

- elliptic, if $L^p(x, \xi)$ is positive or negative definite,
- parabolic, if $L^p(x, \xi)$ is degenerate, and
- hyperbolic, if $L^p(x, \xi)$ is indefinite.

This connection finally allows us to generalize some of our classifications to higher-order partial differential equations. As mentioned before, we only consider a few special situations, the first one being the subject of the following definition.

Definition 1.68 (Higher-order elliptic linear equations). Let $\Omega \subset \mathbb{R}^d$ and consider the m -th order linear partial differential operator $L(x, D)$ over Ω given by

$$L(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u, \quad \text{for } x \in \Omega.$$

Furthermore, let

$$L(x, D)u = f \quad \text{for some } f : \Omega \rightarrow \mathbb{R} \tag{1.43}$$

denote the linear partial differential equation associated with $L(x, D)$. This equation is called elliptic at $x \in \Omega$, if for all vectors $\xi \in \mathbb{R}^d$ we have

$$L^p(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x) \xi^\alpha \neq 0, \quad \text{as long as } \xi \neq 0.$$

The equation (1.43) is called elliptic, if it is elliptic at all points $x \in \Omega$. In the above two cases we sometimes just say briefly that the differential operator $L(x, D)$ is elliptic at x , or that the operator $L(x, D)$ is elliptic.

One can show that elliptic operators can only exist if the order m is an even number, see Review Question 1.5.18. In terms of examples, we have already seen in the previous section that the Laplace operator Δ in arbitrary dimensions is elliptic. The following example shows that this is also true for powers of the Laplacian.

Example 1.69 (Powers of the Laplacian). Let $\Omega \subset \mathbb{R}^d$ be an arbitrary domain, and consider the Laplacian Δ for functions $u : \Omega \rightarrow \mathbb{R}$ defined as usual by

$$\Delta u = \sum_{k=1}^d u_{x_k x_k}.$$

We claim first that the square Δ^2 of the Laplacian gives rise to elliptic fourth-order linear partial differential equations $\Delta^2 u = f$. For this note that

$$\Delta^2 u = \sum_{\ell=1}^d \frac{\partial^2}{\partial x_\ell^2} \left(\sum_{k=1}^d u_{x_k x_k} \right) = \sum_{\ell=1}^d \sum_{k=1}^d u_{x_k x_k x_\ell x_\ell},$$

and therefore the symbol of Δ^2 is given by

$$\sum_{\ell=1}^d \sum_{k=1}^d \xi_k^2 \xi_\ell^2 = \sum_{k=1}^d \xi_k^2 \cdot \sum_{\ell=1}^d \xi_\ell^2 = \left(\sum_{k=1}^d \xi_k^2 \right)^2 > 0 \quad \text{if } \xi \in \mathbb{R}^d \setminus \{0\}.$$

In other words, the operator Δ^2 is elliptic.

It is not difficult to generalize this example. In fact, a similar argument can be used to show that for any positive integer $n \in \mathbb{N}$ the n -th power Δ^n of the Laplacian has the symbol

$$\sum_{k_n=1}^d \cdots \sum_{k_1=1}^d \xi_{k_1}^2 \cdots \xi_{k_n}^2 = \left(\sum_{k=1}^d \xi_k^2 \right)^n > 0 \quad \text{if } \xi \in \mathbb{R}^d \setminus \{0\},$$

which again establishes ellipticity of this differential operator of order $2n$. ■

Definition 1.68 gives a complete characterization of elliptic linear problems in the higher-order case. Unfortunately, however, classifying parabolic and hyperbolic equations poses some serious challenges. In order to demonstrate this, we first consider two examples.

Example 1.70. On the one-dimensional domain $\Omega = (0, 1)$, consider the fourth-order partial differential equation

$$u_t = -u_{xxxx} \quad \text{for } t \geq 0 \quad \text{and } x \in \Omega = (0, 1),$$

and subject to homogeneous Dirichlet boundary conditions. We will see later that all solutions of this partial differential equation can be written in the form

$$u(t, x) = \sum_{k=1}^{\infty} c_k e^{-k^4 \pi^4 t} \sin(k \pi x),$$

with suitable coefficients c_k , for $k \in \mathbb{N}$. Functions of this form behave similarly to the solutions of the heat equation. Even if the function $u(0, \cdot)$ is not smooth, for any $t > 0$ the rapidly decaying factors $e^{-k^4 \pi^4 t}$ guarantee the smoothness of $u(t, \cdot)$ for any $t > 0$. Moreover, solutions converge to 0 as $t \rightarrow \infty$. In other words, the above partial differential equation exhibits two properties which are reminiscent of a *parabolic equation*, and in fact, it is therefore considered to be a linear parabolic equation. ■

Example 1.71. Again on the one-dimensional domain $\Omega = (0, 1)$, we now consider the fourth-order partial differential equation

$$u_{tt} = -u_{xxxx} \quad \text{for } t \geq 0 \quad \text{and } x \in \Omega = (0, 1),$$

and subject to homogeneous Dirichlet boundary conditions. In this case, all solutions of the partial differential equation can be written in the form

$$u(t, x) = \sum_{k=0}^{\infty} (c_k \cos(k^2 \pi^2 t) + d_k \sin(k^2 \pi^2 t)) \sin(k \pi x),$$

with suitable coefficients c_k and d_k , for $k \in \mathbb{N}_0$. In contrast to the previous example, functions of this form behave similarly to the solutions of the wave equation. They are oscillatory in nature, and any initial irregularity in the function $u(0, \cdot)$ will not disappear over time. In other words, the above partial differential equation exhibits properties which are reminiscent of a *hyperbolic equation*, and in fact, it is considered to be a linear hyperbolic equation. ■

The above two examples clearly illustrate the problems one is facing when trying to classify higher-order linear partial differential equations. While in the first- and second-order cases it suffices to only consider the principal part of the associated linear operator, this method cannot work in the general higher-order case. The differential equations in Examples 1.70 and 1.71 both have the same principal part, yet their characteristic solution behavior renders one equation parabolic and the other one hyperbolic. We would like to point out that this phenomenon is not only restricted to fourth-order equations. In fact, if one replaces the right-hand side $-u_{xxxx}$ in both examples by $(-1)^{m+1} \partial^{2m} u / \partial x^{2m}$, then the characteristic solution behavior of both partial differential equations remains unchanged.

Based on the above discussion, we will not attempt to provide a comprehensive classification of higher-order linear partial differential equations — which would be utterly outside the scope of this text. Rather, we concentrate on specific classes of higher-order parabolic and hyperbolic equations, which will suffice for our purposes. In fact, due to our later interest in nonlinear equations, the following definition distinguishes three classes of semilinear partial differential equations as elliptic, parabolic, and hyperbolic.

Definition 1.72 (Partial classification of semilinear equations). *Let the domain $\Omega \subset \mathbb{R}^d$ be given, let $m \in \mathbb{N}$ be an even integer, and consider an m -th order elliptic linear partial differential operator $L(x, D)$ over Ω given by*

$$L(x, D)u = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha u, \quad \text{for } x \in \Omega,$$

whose principal part of the symbol satisfies for all vectors $\xi \in \mathbb{R}^d$ and all $x \in \Omega$ the strict inequality

$$L^p(x, \xi) = \sum_{|\alpha|=m} a_\alpha(x) \xi^\alpha > 0, \quad \text{as long as } \xi \neq 0.$$

Note that in contrast to Definition 1.68, we assume that the principal part of the symbol associated with L is positive definite. Then we define the following.

- *The semilinear partial differential equation*

$$L(x, D)u + f(x, u, D^1 u, \dots, D^{m-1} u) = 0$$

is called elliptic, for any function f . In this equation, we use the abbreviation $D^k u$ for the collection of all k -th order derivatives which was introduced in Definition 1.64. This differential equation describes an unknown function $u = u(x)$, i.e., solutions are functions $u : \Omega \rightarrow \mathbb{R}$.

- *The semilinear partial differential equation*

$$u_t = (-1)^{1+m/2} L(x, D)u + f(t, x, u, D^1 u, \dots, D^{m-1} u)$$

is called parabolic, for any function f . This differential equation describes an unknown function $u = u(t, x)$, i.e., solutions are functions $u : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}$. Notice that both the elliptic operator $L(x, D)$ and the derivatives $D^k u$ only act on the x -variable, and not on the t -variable.

- The semilinear partial differential equation

$$u_{tt} + \alpha u_t = (-1)^{1+m/2} L(x, D)u + f(t, x, u, D^1 u, \dots, D^{m-1} u)$$

is called hyperbolic, for any function f . This differential equation describes an unknown function $u = u(t, x)$, i.e., solutions are functions $u : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}$. Notice that also here both the elliptic operator $L(x, D)$ and the derivatives $D^k u$ only act on the x -variable, and not on the t -variable.

If in either of the above three cases the function f does not explicitly depend on u and any of its derivatives, then the resulting is a linear equation, and we call the resulting equation elliptic, parabolic, or hyperbolic, respectively.

We would like to comment briefly on the fact that in the above definition, we do not allow general elliptic operators as in Definition 1.68, but only elliptic operators with positive definite principal symbol. In many cases, this restriction is purely notational, since if the elliptic operator L has a negative definite principal symbol, then $-L$ has a positive definite principal symbol. In other words, one would expect that one can simply replace L by $-L$. While this does make sense for elliptic semilinear and hyperbolic semilinear equations for $\alpha = 0$, in the parabolic semilinear case or the hyperbolic case with $\alpha \neq 0$ the situation is more delicate, and it does require the strange factor $(-1)^{1+m/2}$ together with a positive definite principal symbol for L . To see this, compare the heat equation

$$u_t = u_{xx},$$

which has a positive definite principal symbol, with the equation

$$u_t = -u_{xx},$$

in which the principal symbol is negative definite. This latter equation is called the *backward heat equation*. As we have mentioned earlier, the usual heat equation exhibits strong smoothing, i.e., if we start with a discontinuous initial condition $u(0, \cdot)$, the equation will produce smooth functions $u(t, \cdot)$ for $t > 0$. This implies that in principle, one cannot solve the heat equation for negative time. Now consider the backward heat equation. One can easily see that it arises from the standard heat equation if we replace t by $-t$. Thus, solving the backward heat equation in forward time is equivalent to solving the usual heat equation in backward time — and we just explained why this is not possible in general. Thus, in the parabolic case it matters greatly whether the elliptic operator L has a positive or negative definite principal symbol, and the choice made in Definition 1.72 turns out to be the correct one.

Example 1.73 (Cahn-Hilliard equation). One of the examples which will be treated later in this book is the Cahn-Hilliard equation. This is a fourth-order nonlinear partial differential equation of the form

$$u_t = -\Delta(\varepsilon^2 \Delta u + f(u)),$$

where $\varepsilon > 0$ is a small parameter, $f(u) = u - u^3$ is a smooth nonlinearity, and the function $u(t, x)$ is defined on $\mathbb{R}_0^+ \times \Omega$ for some domain $\Omega \subset \mathbb{R}^d$. This equation can be rewritten as

$$u_t = \underbrace{-\varepsilon^2 \Delta^2 u}_{=L(x,D)u} - \Delta f(u).$$

Note that according to Example 1.69, the linear part $L(x, D)u = \varepsilon^2 \Delta^2 u$ is elliptic with positive definite principal symbol. In addition, the remaining term on the right-hand side can be written as

$$-\Delta f(u) = -\sum_{k=1}^d \left(f''(u)u_{x_k}^2 + f'(u)u_{x_k x_k} \right) = -f''(u)(\nabla u \cdot \nabla u) - f'(u)\Delta u,$$

i.e., it only contains derivatives up to order two. This shows that the Cahn-Hilliard equation is a parabolic fourth-order nonlinear partial differential equation. ■

The above definition will suffice for most examples presented in this book. Possible exceptions will be treated in an ad hoc manner. The large variety seen in partial differential equations at first appears as a tangled forest of confusion. We believe that the classifications provided in this section will serve as a general guide, allowing the reader to identify and understand many of the trees in this forest.

1.4 • A Deeper Look: Stochastic Differential Equations

In this final section we give a brief introduction to ordinary differential equations which are subject to noise, i.e., to stochastic differential equations. Our presentation will by no means be a complete and fully rigorous treatment, but rather concentrate on the basic ideas, definitions, and results — and thus provide the foundation for the numerical study of simple stochastic partial differential equations later in the book. In addition, we will see that there are interesting connections between stochastic ordinary differential equations and certain elliptic and parabolic partial differential equations.

1.4.1 • Random Effects and Brownian Motion

After our introductory discussion of simple partial differential equations and their classification in the last two sections, we consider once again the initial value problem for ordinary differential equations in the form

$$x' = f(t, x) \quad \text{with} \quad x(t_0) = x_0.$$

We have already seen that under fairly weak assumptions on the function f , every such initial value problem has a unique solution $x : I \rightarrow \mathbb{R}^d$ which is defined on some interval I which contains t_0 in its interior. From a mathematical point of view, this fact was referred to as the *existence and uniqueness result for solutions of ordinary differential equations*. From a more philosophical point of view, this fact can be thought of as a *principle of determinism* — once the initial time t_0 and the initial state x_0 of a system satisfying the differential equation $x' = f(t, x)$ are known, the further temporal evolution of the system is completely determined by the differential equation. This principle can be simplified even more if the underlying differential equation does not explicitly depend on time, i.e., if we consider an initial value problem of the form

$$x' = g(x) \quad \text{with} \quad x(t_0) = x_0.$$

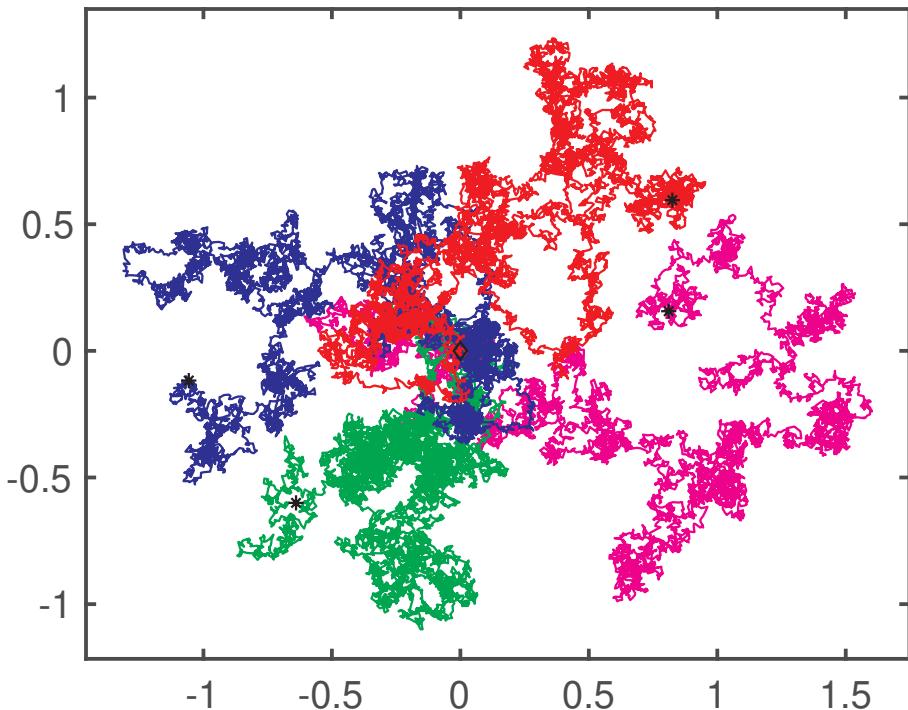


Figure 1.18. Four sample paths of Brownian motion in two space dimensions. All paths start at the origin which is marked by a diamond, their respective endpoints after one unit of time are marked by stars. While the paths share similar qualitative features, their specific shape is completely different in each case.

Since the time does not enter the formulation, one would hope that the specific choice of initial time does not influence the solution behavior. Assume therefore that $x(t)$ solves the *autonomous* initial value problem

$$x' = g(x) \quad \text{with} \quad x(0) = x_0,$$

where we choose the initial time at 0. Then one can easily verify that the shifted solution $y(t) = x(t - t_0)$ solves both the differential equation $y' = g(y)$ and the new initial condition $y(t_0) = x_0$. In other words, if an ordinary differential equation is autonomous, i.e., its right-hand side does not depend explicitly on time t , then the solution is uniquely determined by its initial state x_0 alone.

Irregular motion of pollen grains

In the above formulation, the principle of determinism is the basis for many differential equation models arising in the applied sciences. Yet, there are important phenomena which cannot easily be described on its basis. One example is the following observation by the botanist Robert Brown in 1827. Consider a container of water which is completely at rest and which has one grain of pollen floating on its surface. At first glance, the grain seems to be stationary in the absence of any fluid flow. However, under the microscope Brown noticed that the pollen grain was in fact moving, performing an extremely irregular zig-zag motion. Moreover, repetitions of the experiment resulted in pollen paths which

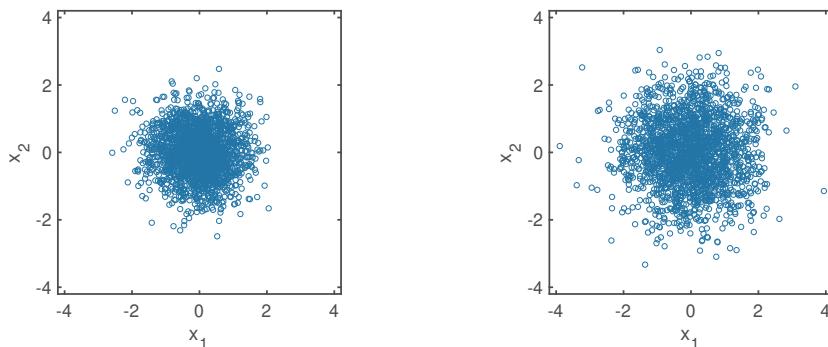


Figure 1.19. Distribution of the location of 2000 randomly generated Brownian paths in two dimensions at time $t = 1/2$ (left image) and $t = 1$ (right image). All Brownian paths originate at the origin.

were qualitatively similar, yet the detailed motion was different in every experiment. In order to give some idea of the qualitative form of this particle motion, which is nowadays called *Brownian motion*, Figure 1.18 contains four numerically simulated pollen grain paths, all originating at the same initial point which is marked by a diamond. One can readily see that the four resulting paths are significantly different. In other words, for the motion of a grain of pollen on water the principle of determinism does not seem to be valid. To provide some further intuition into the typical behavior of the pollen grains, Figure 1.19 shows the positions of 2000 numerically simulated grains of pollen after $1/2$ and 1 unit of time in the left and right image, respectively. In all of these simulations, the grain started out at the origin. Notice that the observed positions are distributed mostly within a circle of certain radius centered at the origin, that the radius of the circle seems to be larger for larger time, and that the positions towards the boundary of the circle occur less frequently.

At the time of his original experimental observation, Brown was unable to come up with a satisfactory description of this irregular pollen grain motion. Towards the end of the 19th century, however, Maxwell and Boltzmann, among others, proposed a kinetic theory of heat which is based on the idea that even if the water in a container is completely at rest from a macroscopic point of view, the water molecules themselves are constantly in motion due to thermodynamic effects. This constant motion of the water molecules leads to an enormous number of collisions with the pollen grain, thereby ultimately leading to the irregular Brownian motion of the particle. Even though the validity of the kinetic theory of heat was not universally accepted in the second half of the 19th century, it does in fact reopen the door to our above principle of determinism. Since the motion of the grain is due to the motion of the water molecules, an ordinary differential equations model for the complete system would have to include equations of motions for all of the involved particles — and the initial state would have to be described not only by the initial position of the pollen grain, but also by the initial locations and velocities of all involved water molecules. In other words, the different realizations of the motion shown in Figure 1.18 are due to different initial states of the water molecules.

Despite the explanation of Brownian motion provided by the kinetic theory of heat, we are still facing a major challenge in producing a useful mathematical model for the phenomenon. If one is interested in a model which only describes the motion of the grain, then one has to abandon the principle of determinism. In addition, it is not immediately

clear how such a model would incorporate the “typical” behavior of pollen paths. Yet, a model which is aimed at only describing the pollen motion would have the advantage of being low-dimensional, since it would only have to describe the evolution of the two position coordinates. On the other hand, if one is interested in a “deterministic” model in the sense of an ordinary differential equation, the model would necessarily have to include the initial positions and velocities of all involved water molecules — which clearly would lead to such a high-dimensional ordinary differential equation that it would be completely useless in practice.

A brief review of concepts from probability theory

It will be shown in the current section that the above problems can be resolved satisfactorily by the concept of a *stochastic differential equation*. The basic philosophy behind this concept is the attempt to study all possible Brownian paths at once by using the notion of probability space, which will now be introduced briefly. A *probability space* is given by a triple $(F, \mathcal{F}, \mathbb{P})$ consisting of the following:

- The set F is called the *sample space*, and it serves basically as an index set to access all possible outcomes of an experiment.⁴
- The set \mathcal{F} is a collection of subsets of F , which are called the *observable sets* or the *events* of the experiment. This collection carries the structure of a σ -algebra, i.e., we have $F \in \mathcal{F}$, and for any set $A \in \mathcal{F}$, its complement $F \setminus A$ is also contained in \mathcal{F} . Finally, for any countably infinite sequence of sets $(A_k)_{k \in \mathbb{N}} \subset \mathcal{F}$, their union $\bigcup_{k \in \mathbb{N}} A_k$ is in \mathcal{F} as well.
- Last but not least, the function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ is called the *probability measure*. It assigns a probability $\mathbb{P}(A) \in [0, 1]$ to every observable set $A \in \mathcal{F}$, which is interpreted as the likelihood of its occurrence in the experiment. The probability measure satisfies both $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(F) = 1$, and for every sequence of pairwise disjoint sets $(A_k)_{k \in \mathbb{N}} \subset \mathcal{F}$ one has to have

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Probability spaces are used to model experiments that involve an element of chance, and the prime example is that of a coin toss.

Example 1.74. Suppose we are performing n coin tosses, and record each time whether the coin shows heads (H) or tails (T). In this case, the sample space could consist of all possible outcomes of the experiment, which is given by the set of all n -tuples with entries equal to H or T, i.e., we have

$$F_1 = \{(f_1, \dots, f_n) : f_i \in \{H, T\}\} = \{H, T\}^n.$$

In this case, the collection \mathcal{F}_1 of observable sets can be defined as the set of all subsets of F_1 . Furthermore, since each of the 2^n possible outcomes is equally likely, we have

$$\mathbb{P}_1(\{(f_1, \dots, f_n)\}) = \frac{1}{2^n}, \quad \text{for every } (f_1, \dots, f_n) \in F_1,$$

⁴We would like to mention that in probability theory, the sample space is usually denoted by Ω . However, since this book is mainly about partial differential equations, and in the latter context Ω usually denotes the physical domain, we use the letter F to avoid confusion.

as well as

$$\mathbb{P}_1(A) = \frac{\#A}{2^n},$$

if $\#A$ denotes the number of elements of A . \blacksquare

Example 1.75. In the previous example, the probability space model for our coin toss was based on the detailed set of all possible outcomes of the experiment. If on the other hand we are only interested in the number of coins showing heads, then we could have instead chosen the sample space

$$F_2 = \{0, 1, \dots, n\}.$$

Choosing again the collection \mathcal{F}_2 of observable sets as the set of all subsets of F_2 , the probability measure associated with this experiment satisfies

$$\mathbb{P}_2(\{k\}) = \binom{n}{k} \frac{1}{2^n},$$

as well as

$$\mathbb{P}_2(A) = \frac{1}{2^n} \sum_{k \in A} \binom{n}{k}.$$

Note that due to the binomial formula we have $\mathbb{P}_2(F_2) = 1$. \blacksquare

As the above two examples show, in general there are many different ways of associating a probability space with a given experiment. Once a particular choice has been made, the tools of probability theory then allow one to determine the likelihood of certain events by computing the probability $\mathbb{P}(A)$ of an event $A \in \mathcal{F}$. Frequently, however, the probability space model for an experiment is only a first step towards answering a more complicated question. For example, one might be interested in computing certain numeric values from the experiments, such as the number of heads in a coin toss, or the number of tails, or the number of heads that were obtained during the last five tosses, or the length of the longest consecutive run of heads, etc.

In order to address this issue, probability theory uses the concept of a *random variable*, which is nothing but a function $X : F \rightarrow \mathbb{R}^d$ defined on the sample space F and taking values in some Euclidean space. In the general case, this function has to satisfy a fairly mild condition called *measurability*.⁵ If we then consider any “reasonable” subset $I \subset \mathbb{R}^d$, we can determine the probability that the random variable X takes values in I , i.e., we can compute

$$\mathbb{P}\{X \in I\} = \mathbb{P}(\{\omega \in F : X(\omega) \in I\}),$$

where we usually use the short-hand notation on the left-hand side rather than the (more precise but more cumbersome) expression on the right-hand side. The collection of sets I for which this expression makes sense is closely related to the notion of measurability of X . It turns out that I can be any open or closed subset of \mathbb{R}^d , as well as any set that can be represented as a countable union or intersection of such sets. More precisely, the set I has to be contained in the *Borel σ -algebra* \mathcal{B}^d on \mathbb{R}^d , which is the smallest σ -algebra on \mathbb{R}^d .

⁵In some sense, the concept of measurability makes sure that the random variable X interacts well with the underlying σ -algebra \mathcal{F} , and one might think of it as the minimal regularity which is required for working with random variables in a meaningful way. Yet, for our numerically oriented applications which are focused on a very specific situation we will be able to avoid detailed discussions of measurability questions.

that contains all open and closed sets. As long as we have $I \in \mathcal{B}^d$, the measurability of X will guarantee that also the associated set $\{X \in I\}$ is in fact contained in \mathcal{F} and therefore has an associated probability.

We illustrate the concept of a random variable briefly for the question “*How many coin tosses resulted in heads?*” for the experiment of n independent coin tosses. This will be done for each of the two probability spaces $(F_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(F_2, \mathcal{F}_2, \mathbb{P}_2)$ introduced above.

Example 1.76. If we use the probability space $(F_1, \mathcal{F}_1, \mathbb{P}_1)$ introduced in Example 1.74, then one can define a random variable X_1 as

$$X_1(\omega) = \#\{i \in \{1, \dots, n\} : f_i = H\} \quad \text{for } \omega = (f_1, \dots, f_n) \in F_1 = \{H, K\}^n.$$

In this case, the probability of observing exactly k heads is given by

$$\mathbb{P}\{X = k\} = \mathbb{P}\{X \in \{k\}\},$$

which can be evaluated to $\binom{n}{k}/2^n$. ■

Example 1.77. If on the other hand one employs the probability space $(F_2, \mathcal{F}_2, \mathbb{P}_2)$ from Example 1.75, then X_2 can be simply chosen as

$$X_2(\omega) = \omega \quad \text{for } \omega \in F_2 = \{0, 1, \dots, n\}.$$

Notice, however, that this probability space would make it impossible to consider the question “*How many of the first five coin tosses showed heads?*”, while this could easily be done in the setting of Examples 1.74 and 1.76. ■

One of the important notions associated with a random variable X is its *expected value* $\mathbb{E}(X)$. Intuitively speaking, the expected value of a random variable X is the average value of X that is observed if the experiment modeled by the probability space $(F, \mathcal{F}, \mathbb{P})$ is repeated many times. To be more specific, consider first the special case of a random variable which is an *indicator function*. For this, consider an event $A \in \mathcal{F}$ and define the function $\chi_A : F \rightarrow \mathbb{R}$ via

$$\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

We can interpret this random variable through the following experiment. Based on the probabilities underlying the probability space $(F, \mathcal{F}, \mathbb{P})$, randomly select a sample $\omega \in F$. If this sample lies in the given set A , we receive a payout of $1 = \chi_A(\omega)$, but if $\omega \notin A$, then the payout is $0 = \chi_A(\omega)$. If we perform this experiment many times, what is our expected payout? In this simple situation, it can easily be determined. The probability of selecting a sample in A is given by $\mathbb{P}(A)$, and this immediately implies that our expected average payout is

$$\mathbb{E}(\chi_A) = \mathbb{P}(A).$$

This example can readily be generalized. Suppose we are given events $A_1, \dots, A_\ell \subset \mathcal{F}$ and that our payout for selecting a sample in A_i is given by $a_i \in \mathbb{R}$. Then for a sample $\omega \in F$ the actual payout is represented by the random variable

$$X(\omega) = \sum_{i=1}^{\ell} a_i \chi_{A_i}(\omega),$$

and we leave it as an easy exercise to the reader to show that the expected, i.e., average payout per drawing is given by

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{\ell} a_i \chi_{A_i}\right) = \sum_{i=1}^{\ell} a_i \mathbb{P}(A_i) = \sum_{i=1}^{\ell} a_i \mathbb{E}(\chi_{A_i}).$$

The above formula defines the expected value for random variables which are arbitrary linear combinations of indicator functions. Notice that such linear combinations are piecewise constant functions, or step functions. If we are now given an arbitrary random variable $X : F \rightarrow \mathbb{R}$, then under suitable conditions one can approximate X by such linear combinations and define the expect value $\mathbb{E}(X)$ as the limit of the expected values of the approximations.

The acute reader might have realized that the above-described process sounds awfully similar to the construction of integrals in calculus. In fact, properly defining the expected value $\mathbb{E}(X)$ amounts to defining an integral of the random variable X over the probability space $(F, \mathcal{F}, \mathbb{P})$, and one usually writes

$$\mathbb{E}(X) = \int_F X d\mathbb{P}.$$

The details of this construction can be found in any textbook on measure theory, see for example [6, 10].

Beyond the above intuitive description, the following practical computation of the expected value will be important for us later on. Suppose that we have an experiment described by the random variable X over the probability space $(F, \mathcal{F}, \mathbb{P})$. Suppose further that we have performed the experiment a number of times and recorded the observed values of X every time. For example, say that n repetitions of the experiment have led to the outcomes $X(\omega_1), \dots, X(\omega_n)$. Then one can show that as n increases we have

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X(\omega_i).$$

This fact is the celebrated *strong law of large numbers*, which can be found in any standard textbook on probability theory. See for example [5, 10].

Before returning to our discussion of Brownian motion, we need to introduce one final probabilistic concept, namely that of *independence*. Consider a probability space $(F, \mathcal{F}, \mathbb{P})$, as well as two events $A, B \in \mathcal{F}$, and suppose we are interested in the likelihood of the occurrence of event B . Assume that after performing the underlying experiment, someone tells us that event A has occurred. Based on this additional piece of information, what is the new probability for the occurrence of the event B ? It is not hard to convince oneself that the new probability, which is called the *conditional probability of B given A* , and which is denoted by $\mathbb{P}(B|A)$, is given by

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

Essentially, this formula rescales the likelihood in such a way that the inclusion $A \subset B$ implies the identity $\mathbb{P}(B|A) = 1$ — since the occurrence of A in this case implies the occurrence of B . The notion of conditional probability naturally leads to the concept of independence. We call two events $A, B \in \mathcal{F}$ *independent*, if the occurrence of one event does

not affect the likelihood of the other, i.e., we have both $\mathbb{P}(B|A) = \mathbb{P}(B)$ and $\mathbb{P}(A|B) = \mathbb{P}(A)$. One can easily see that this is equivalent to requiring

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

The concept of independence can naturally be lifted to the case of random variables. Recall that we can think of a random variable as answering a certain question about the experiment, or, in more applied settings, as making a measurement. If one then considers two random variables $X, Y : F \rightarrow \mathbb{R}^d$, then these random variables are called *independent* as long as

$$\mathbb{P}\{X \in A \text{ and } Y \in B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}, \quad \text{for all } A, B \in \mathcal{B}^d,$$

where \mathcal{B}^d denotes the Borel σ -algebra on \mathbb{R}^d mentioned earlier. This definition formalizes the concept that answering the question encoded by the random variable X does in no way affect the outcome of the question encoded by Y . Similarly, a family of m random variables X_1, \dots, X_m is called *independent*, as long as

$$\mathbb{P}\{X_1 \in A_1 \text{ and } X_2 \in A_2 \text{ and } \dots \text{ and } X_m \in A_m\} = \prod_{k=1}^m \mathbb{P}\{X_k \in A_k\},$$

for all $A_k \in \mathcal{B}^d$, where $k = 1, \dots, m$. In this formula, the expression $\prod_{k=1}^m a_k$ denotes the product of the real numbers a_1, \dots, a_m .

Mathematical model for Brownian motion

After this lengthy detour into the realm of probability theory, we return to the story of pollen grains and Brownian motion. As we mentioned before, the desire to obtain a low-dimensional model for Brownian motion necessarily has to account for the multitude of possible paths that the pollen grain could take. Let us assume for the moment that we can construct a probability space $(F, \mathcal{F}, \mathbb{P})$ for the experiment performed by Robert Brown in 1827. Based on our discussion of random variables, the question “*What is the position of the pollen of grain after t units of time?*” can then be encoded by a random variable which is defined on F and takes values in \mathbb{R}^2 . Thus, it seems reasonable to try to construct a model for Brownian motion as a function

$$B : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^2,$$

where for each t the function $B(t, \cdot) : F \rightarrow \mathbb{R}^2$ denotes a random variable. In other words, we seek to model Brownian motion as a family of random variables which is indexed by the time t — and such an object is usually referred to as a *stochastic process* in probability theory. Using this interpretation, we can view the paths shown in Figure 1.18 as four specific paths of the form $\{B(t, \omega) : 0 \leq t \leq 1\} \subset \mathbb{R}^2$, for four specific but random choices of the argument $\omega \in F$. Similarly, the points shown in the left and right images of Figure 1.19 correspond to $B(1/2, \omega)$ and $B(1, \omega)$, respectively, for 2000 randomly chosen realizations $\omega \in F$.

But which conditions distinguish a Brownian motion from other stochastic processes? To answer this question we first seek to reduce the dimension of our problem even further. Since $B(t, \omega) \in \mathbb{R}^2$, we can write this stochastic process via its two coordinates as

$$B(t, \omega) = (B_1(t, \omega), B_2(t, \omega)) \in \mathbb{R}^2.$$

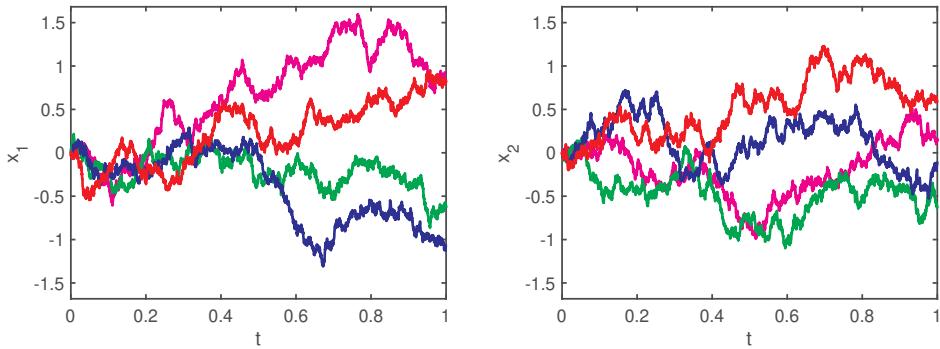


Figure 1.20. One-dimensional projections of the Brownian paths shown in Figure 1.18. The left image shows the x_1 -coordinate of each path as a function of time t , i.e., it shows the graph of $B_1(\cdot, \omega)$. The right image shows the corresponding x_2 -coordinates.

Then both B_1 and B_2 are real-valued stochastic processes. For the four Brownian paths of Figure 1.18 we show the first coordinate B_1 as a function of time t in the left image of Figure 1.20, the second components B_2 are shown in the right image. The graphs indicate that all of these one-dimensional projections share similar qualitative features. In fact, this is to be expected from the explanation of Brownian motion provided by the kinetic theory of heat. Since the underlying water molecules do not see distinguished coordinate axes, both the x_1 - and the x_2 -coordinate of Brownian motion should act completely independently, while still following the same underlying physical principles. In other words, the stochastic processes B_k for $k = 1, 2$ are independent realizations of *one-dimensional Brownian motion*, in which a particle is moving on the real line due to the effects of thermal fluctuations.

The first quantitative discussion of one-dimensional Brownian motion was presented by Albert Einstein in 1905. For this, we assume from now on that we have a one-dimensional Brownian motion $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ over a probability space $(F, \mathcal{F}, \mathbb{P})$. By applying the kinetic theory of heat that was mentioned earlier in this section, Einstein was able to describe the probabilities of events of the form $\mathbb{P}\{W(t, \cdot) \in I\}$ for arbitrary Borel sets $I \subset \mathbb{R}$. In fact, Einstein realized that these probabilities can be determined from a *density function* $p : \mathbb{R}_0^+ \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ via the integral formula

$$\mathbb{P}\{W(t, \cdot) \in I\} = \int_I p(t, x) dx \quad \text{for Borel sets} \quad I \subset \mathbb{R}.$$

One of the major contributions of Einstein was to show that this density function p has to satisfy the heat equation in the form

$$p_t = D p_{xx}, \quad \text{with} \quad D = \frac{2RT}{Nf}, \quad (1.44)$$

where R denotes the universal gas constant, T is the absolute temperature, f the friction coefficient, and N the Avogadro number.⁶ Since the density p has to generate a probabil-

⁶In fact, this explicit formula for the diffusion coefficient D made it possible to determine the Avogadro number from Brownian motion experiments, and led to one of the first experimental justifications of the kinetic theory!

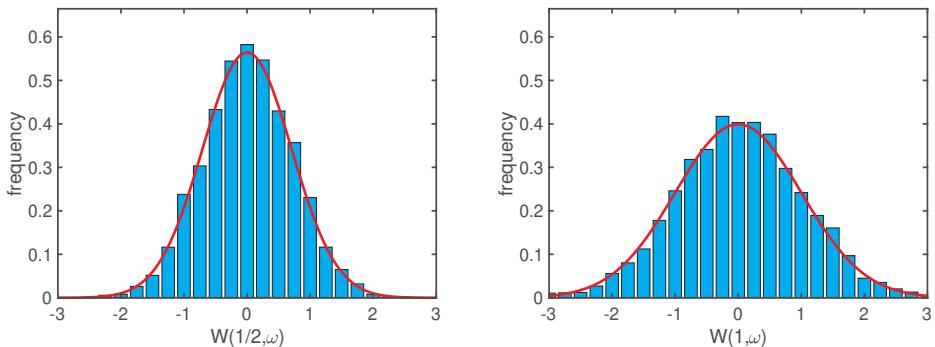


Figure 1.21. Distribution of the location of randomly generated Brownian paths in one space dimension at time $t = 1/2$ (left image) and $t = 1$ (right image). All Brownian paths originate at the origin. Each image shows a histogram based on 5000 sample paths. The actual underlying probability densities are shown in red.

ity measure, one also has to require that for all $t \in \mathbb{R}_0^+$ and $x \in \mathbb{R}$ one has

$$p(t, x) \geq 0, \quad \text{as well as} \quad \int_{-\infty}^{\infty} p(t, x) dx = 1.$$

Finally, in order to ensure that the particle originates at the origin, Einstein further deduced that

$$\lim_{t \rightarrow 0} p(t, x) = 0 \quad \text{for all} \quad x \neq 0.$$

It turns out that the above set of conditions leads to a unique solution for the density which is given by

$$p(t, x) = \frac{1}{\sqrt{4\pi D t}} e^{-x^2/(4Dt)}.$$

Thus, by relying solely on the kinetic theory of heat Einstein was able to show that if W denotes a one-dimensional Brownian motion, then one has to have

$$\mathbb{P}\{W(t, \cdot) \in [a, b]\} = \frac{1}{\sqrt{4\pi D t}} \int_a^b e^{-x^2/(4Dt)} dx, \quad (1.45)$$

for real numbers $a \leq b$. In other words, the likelihood for finding the particle between a and b at time t equals the area under the graph of $p(t, \cdot)$ between $x = a$ and $x = b$. This fact is illustrated numerically in Figure 1.21. The red curves in the left and right image of this figure show the density functions $p(1/2, \cdot)$ and $p(1, \cdot)$, respectively, for the diffusion coefficient $D = 1/2$. Superimposed on these curves we have shown histograms for the position of the Brownian path at times $t = 1/2$ and $t = 1$ which were obtained from numerically simulating 5000 one-dimensional Brownian motions. The histograms are scaled in such a way that the frequency of observing the particle between the left and right endpoint of a histogram bar equals the area of the bar. These images impressively demonstrate the validity of Einstein's derivation.

From a mathematical perspective, the identity (1.45) alone does not distinguish a stochastic process as a one-dimensional Brownian motion — and we will present the precise definition in the following section. However, random variables with density functions of the form given in (1.45) are central to probability theory, and we therefore close this section with the following definition.

Definition 1.78 (Normally distributed random variable). Let $(F, \mathcal{F}, \mathbb{P})$ denote a probability space, and let $X : F \rightarrow \mathbb{R}$ be a real-valued random variable. Then we say that X is normally distributed with mean μ and variance σ^2 , if for all real numbers $a \leq b$ we have

$$\mathbb{P}\{X \in [a, b]\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b e^{-(x-\mu)^2/(2\sigma^2)} dx,$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. Equivalently, we say that X is a Gaussian random variable with mean μ and variance σ^2 . In the special case $\mu = 0$ and $\sigma = 1$, we call X a standard normally distributed random variable.

In other words, the identity (1.45) states that for a one-dimensional Brownian motion W , the random variable $W(t, \cdot)$ is normally distributed with mean 0 and variance $2Dt$.

For later reference we note that on most current programming environments one can find random number generators which simulate normally distributed random variables. For example, in Matlab the command `randn` can be used to generate numbers which are normally distributed with mean zero and variance one, i.e., a standard normally distributed random variable. Furthermore, it is not hard to show that if X is a standard normally distributed random variable, then σX is normally distributed with mean 0 and variance σ^2 .

1.4.2 ■ Basic Properties of the Wiener Process

While the last section concentrated mainly on motivating and justifying the use of probability theory and stochastic processes for the study of certain natural phenomena such as Brownian motion, this section will mark the beginning of a more detailed mathematical study which will ultimately culminate in the definition of a stochastic differential equation. For now, however, we return to our discussion of Brownian motion. From a physics perspective, Albert Einstein identified one of the central ingredients of a Brownian motion — the fact that the position of the particle is normally distributed and that the variance of this distribution grows linearly with time. As we already mentioned in the last section, this property alone does not distinguish a stochastic process as a Brownian motion. From a mathematical point of view, Norbert Wiener was instrumental in formulating the following fundamental definition.

Definition 1.79 (Brownian motion, Wiener process). Let $(F, \mathcal{F}, \mathbb{P})$ denote a probability space, and consider a stochastic process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$. Then W is called Brownian motion or Wiener process, if the following conditions hold:

- (a) The process W starts at the origin, i.e., for all $\omega \in F$ we have $W(0, \omega) = 0$.
- (b) For every $\omega \in F$ the Brownian path $W(\cdot, \omega) : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ is a continuous function.
- (c) For every $0 \leq s < t$ the random variable $W(t, \cdot) - W(s, \cdot)$ is normally distributed with mean 0 and variance $t - s > 0$.
- (d) The process W has independent increments in the following sense. For every $m \in \mathbb{N}$ and arbitrary times $0 \leq t_0 < t_1 < \dots < t_m$ the random variables $W(t_k, \cdot) - W(t_{k-1}, \cdot)$ for $k = 1, \dots, m$ are independent.

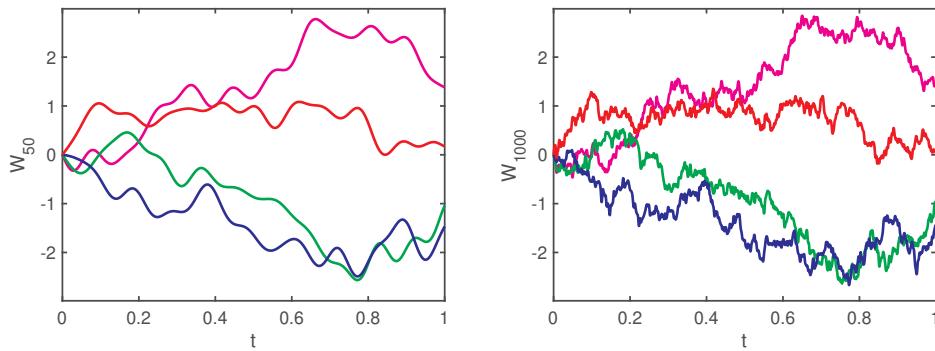


Figure 1.22. Approximations to four paths of the one-dimensional Wiener process using the series representation (1.46). The left image shows the approximations using the first 50 terms of the sum, the right image uses the first 1000 terms.

In addition, a vector-valued stochastic process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ over a probability space $(F, \mathcal{F}, \mathbb{P})$ is called Brownian motion or Wiener process, if each of the d component functions are one-dimensional Wiener processes as above which are independent.

Existence of Wiener processes

The above definition lists the conditions which make a stochastic process a Brownian motion, yet it does not make any claims of existence. In fact, showing that there actually is a probability space $(F, \mathcal{F}, \mathbb{P})$ and a stochastic process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ which satisfies all four conditions is highly nontrivial. Nevertheless, it can and has been done, but it is based on deep theorems of probability theory which are beyond the scope of our brief introduction to the subject.

A more direct approach was pursued by Norbert Wiener in his original existence proof. He showed that if X_k , for $k \in \mathbb{N}$, denotes an independent family of standard normally distributed random variables over a probability space $(F, \mathcal{F}, \mathbb{P})$, then the series

$$W(t, \omega) = \frac{\pi}{2\sqrt{2}} \sum_{k=1}^{\infty} \frac{X_k(\omega)}{k} \sin \frac{\pi k t}{2} \quad \text{for } 0 \leq t \leq 1, \quad \omega \in F, \quad (1.46)$$

converges, and in fact furnishes a Brownian motion for $0 \leq t \leq 1$. This stochastic process can then be extended to a Brownian motion on \mathbb{R}_0^+ by stringing together independent copies. At first glance, the above formula seems to be an excellent tool for approximating a Brownian motion, but the series converges extremely slowly — and we will present better numerical methods later in this book. To illustrate this fact, Figure 1.22 contains approximations to four paths of the one-dimensional Wiener process using the series representation (1.46). The left image shows the approximations using the first 50 terms of the sum, which clearly do not resemble the erratic zig-zag motions that were observed by Brown. One has to actually include on the order of 1000 terms, as shown in the right image, to obtain paths which are qualitatively acceptable.

Fundamental path properties

For the remainder of this section, we present a number of results providing insight into fundamental properties of Brownian motion, as defined in Definition 1.79. None of these

results will be proved in detail; rather, we concentrate on illustrating their effects. In this way, we hope that the reader can develop a sense for what Brownian motion really is. Proofs for all of these results can be found in standard texts on probability theory and stochastic differential equations, such as [1, 5, 10, 43].

We begin our discussion of the properties of the Wiener process with the following result, which identifies certain invariance and transformational properties.

Proposition 1.80 (Invariance properties). *Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ denote a possibly vector-valued Wiener process over the probability space $(F, \mathcal{F}, \mathbb{P})$. Then the following holds.*

- (a) *If $Q \in \mathbb{R}^{d \times d}$ denotes an arbitrary orthogonal matrix, then the stochastic process V defined by $V(t, \omega) = QW(t, \omega)$ is also a vector-valued Wiener process.*
- (b) *Each of the following formulas defines a vector-valued Wiener process:*
 - $V(t, \omega) = c W(t/c^2, \omega)$, for some real constant $c \neq 0$,
 - $V(t, \omega) = W(t+s, \omega) - W(s, \omega)$, for some real constant $s \geq 0$, and
 - $V(t, \omega) = t W(1/t, \omega)$.

For part (a), recall that a matrix $Q \in \mathbb{R}^{d \times d}$ is called orthogonal, if we have $QQ^t = Q^tQ = I$, i.e., if the inverse matrix Q^{-1} of Q exists and coincides with the transpose matrix Q^t .

This first result shows that Brownian motion exhibits a number of inherent symmetries and scaling properties. Part (a) of the lemma shows for example that Brownian motion in \mathbb{R}^d does not see any coordinate system in the sense that orthogonal transformations (i.e., rotations and reflections) of the process lead to another perfectly legitimate Brownian motion. Similarly, the properties in part (b) demonstrate certain scaling invariances of Brownian motion.

The second result addresses the sizes of excursions from the origin that are typically observed after t units of time.

Proposition 1.81 (Law of the iterated logarithm). *Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ denote a possibly vector-valued Wiener process over the probability space $(F, \mathcal{F}, \mathbb{P})$. Then there exists a set $\tilde{F} \in \mathcal{F}$ with $\mathbb{P}(\tilde{F}) = 1$ such that the following hold.*

- (a) *If the Wiener process is real-valued, i.e., we have $d = 1$, then both*

$$\limsup_{t \rightarrow \infty} \frac{W(t, \omega)}{\sqrt{2t \ln |\ln t|}} = 1 \quad \text{and} \quad \liminf_{t \rightarrow \infty} \frac{W(t, \omega)}{\sqrt{2t \ln |\ln t|}} = -1$$

hold for all $\omega \in \tilde{F}$.

- (b) *If the Wiener process is real-valued, i.e., we have $d = 1$, then both*

$$\limsup_{t \rightarrow 0^+} \frac{W(t, \omega)}{\sqrt{2t \ln |\ln t|}} = 1 \quad \text{and} \quad \liminf_{t \rightarrow 0^+} \frac{W(t, \omega)}{\sqrt{2t \ln |\ln t|}} = -1$$

hold for all $\omega \in \tilde{F}$.

- (c) *If the Wiener process is vector-valued, i.e., we have $d > 1$, then*

$$\limsup_{t \rightarrow \infty} \frac{|W(t, \omega)|}{\sqrt{2t \ln |\ln t|}} = 1 \quad \text{for all } \omega \in \tilde{F}$$

is satisfied.

In the following, we say that a property holds for almost all $\omega \in F$, if there exists a set $\tilde{F} \in \mathcal{F}$ with $\mathbb{P}(\tilde{F}) = 1$ such that the property holds for all $\omega \in \tilde{F}$.

Even though we present the above result only in the context of Brownian motion, the law of the iterated logarithm is one of the fundamental and far-reaching results of probability theory, which is valid under fairly weak hypotheses. In our situation, however, it provides a clear picture of how far Brownian motion can travel in a given amount of time.

To illustrate this, we first consider the behavior for large times t . By the definition of the limit superior and the limit inferior, one can see that the above proposition implies for every $\omega \in \tilde{F}$ and every $\varepsilon > 0$ the existence of a time $t_0 > 0$ such that

$$-(1 + \varepsilon) \sqrt{2t \ln |\ln t|} \leq W(t, \omega) \leq +(1 + \varepsilon) \sqrt{2t \ln |\ln t|} \quad \text{for all } t \geq t_0.$$

Furthermore, there exists an increasing sequence of times t_k with $t_k \rightarrow \infty$ as $k \rightarrow \infty$ such that

$$(-1)^{k+1} W(t_k, \omega) \geq (1 - \varepsilon) \sqrt{2t_k \ln |\ln t_k|},$$

which means that we have

$$W(t_k, \omega) \leq -(1 - \varepsilon) \sqrt{2t_k \ln |\ln t_k|} \quad \text{for even } k \in \mathbb{N},$$

as well as

$$W(t_k, \omega) \geq +(1 - \varepsilon) \sqrt{2t_k \ln |\ln t_k|} \quad \text{for odd } k \in \mathbb{N}.$$

What do these rather intimidating formulas actually mean? Since $\varepsilon > 0$ can be chosen arbitrarily small, they show that a typical path of Brownian motion will oscillate between the curves

$$-\sqrt{2t \ln |\ln t|} \quad \text{and} \quad \sqrt{2t \ln |\ln t|} \quad \text{as } t \rightarrow \infty,$$

and these curves basically grow like $\pm\sqrt{2t}$. On the one hand, while this implies a fairly slow growth of a typical Brownian path, it also shows that typical paths become unbounded. Moreover, typical Brownian paths do not just converge to $+\infty$ or $-\infty$. Rather, they oscillate infinitely many times between the above two curves. This behavior is indicated in Figure 1.23. Furthermore, since Brownian paths are continuous, the above oscillatory behavior forces a typical Brownian path to have an infinite number of large zeros at arbitrarily large times.

We now turn our attention to the behavior for small times. As one can easily see, Proposition 1.81 implies for every $\omega \in \tilde{F}$ and every $\varepsilon > 0$ the existence of a time $t_0 > 0$ such that

$$-(1 + \varepsilon) \sqrt{2t \ln |\ln t|} \leq W(t, \omega) \leq +(1 + \varepsilon) \sqrt{2t \ln |\ln t|} \quad \text{for all } 0 < t \leq t_0, \quad (1.47)$$

and there exists a decreasing sequence of times t_k such that

$$(-1)^{k+1} W(t_k, \omega) \geq (1 - \varepsilon) \sqrt{2t_k \ln |\ln t_k|} \quad \text{with} \quad \lim_{k \rightarrow \infty} t_k = 0^+. \quad (1.48)$$

As before, these inequalities imply that the Brownian path in fact has to oscillate between the curves

$$-\sqrt{2t \ln |\ln t|} \quad \text{and} \quad \sqrt{2t \ln |\ln t|} \quad \text{as } t \rightarrow 0^+,$$

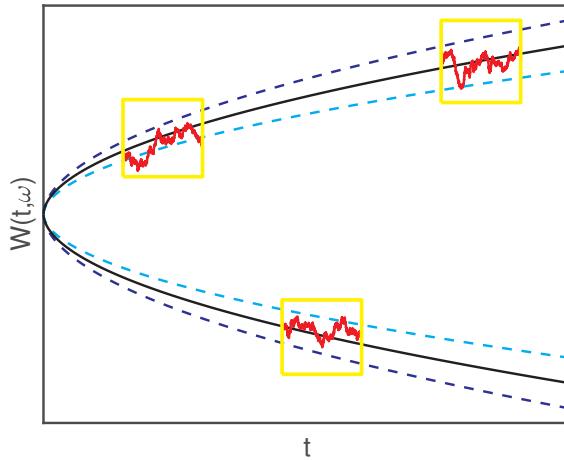


Figure 1.23. Illustration of the law of the iterated logarithm. The black curves are the graphs of the functions $\pm\sqrt{2t \ln |\ln t|}$, the dashed blue curves are for $\pm(1 + \varepsilon)\sqrt{2t \ln |\ln t|}$, and the dashed cyan curves for $\pm(1 - \varepsilon)\sqrt{2t \ln |\ln t|}$. As $t \rightarrow \infty$, a typical Brownian path $W(t, \omega)$, shown in red, has to oscillate between the upper and lower regions which are bounded by the cyan and blue dashed curves, i.e., it has to enter a sequence of “windows” which are indicated in yellow.

which also in this limit behave basically like $\pm\sqrt{2t}$. Together with the intermediate value theorem, this last observation guarantees the existence of a sequence of zeros of the path $W(\cdot, \omega)$ which converges to the zero at $t = 0$. It turns out that we can say even more about the set of zeros of a typical Brownian path.

Proposition 1.82 (Properties of the zero set). Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ denote a real-valued Wiener process over the probability space $(F, \mathcal{F}, \mathbb{P})$, and for every $\omega \in F$, let

$$\mathcal{Z}(\omega) = \{t \in \mathbb{R}_0^+ : W(t, \omega) = 0\}$$

denote the set of all zeros of the path $W(\cdot, \omega)$. Then there exists an observable set $\tilde{F} \in \mathcal{F}$ which has probability $\mathbb{P}(\tilde{F}) = 1$ such that for all $\omega \in \tilde{F}$ the following holds.

- (a) The set $\mathcal{Z}(\omega) \subset \mathbb{R}_0^+$ is closed and unbounded.
- (b) If $t_0 \in \mathcal{Z}(\omega)$ is a zero of $W(\cdot, \omega)$, then in every interval of the form $[t_0, t_0 + \varepsilon]$ there are infinitely many zeros of $W(\cdot, \omega)$, for any $\varepsilon > 0$. Similarly, if $t_0 > 0$ is a zero of $W(\cdot, \omega)$, then in every interval of the form $(t_0 - \varepsilon, t_0]$ there are infinitely many zeros of $W(\cdot, \omega)$, for any $\varepsilon > 0$. In other words, zeros of the path $W(\cdot, \omega)$ are never isolated.
- (c) The set $\mathcal{Z}(\omega) \subset \mathbb{R}_0^+$ has Lebesgue measure zero, i.e., for every $\varepsilon > 0$ there exists a sequence of bounded closed intervals $[a_k, b_k] \subset \mathbb{R}_0^+$ such that both

$$\mathcal{Z}(\omega) \subset \bigcup_{k=1}^{\infty} [a_k, b_k] \quad \text{and} \quad \sum_{k=1}^{\infty} (b_k - a_k) < \varepsilon.$$

In other words, the set $\mathcal{Z}(\omega)$ can be covered by intervals whose total length can be made arbitrarily small.

The above proposition is the first indicator of how strange typical Brownian paths really are. While there are many functions that are introduced in undergraduate classes which have infinitely many zeros, Brownian paths take this to the extreme. Whenever a Brownian path becomes zero, there are infinitely many zeros close by. Nevertheless, the set of zeros is basically negligible, since its “volume” (in the form of Lebesgue measure) is zero.

The accumulation properties of zeros are only the first implication of our formulas (1.47) and (1.48). If we combine these formulas with Proposition 1.80(b), then they can be used to describe the behavior of the difference quotient

$$\frac{W(t+h, \omega) - W(t, \omega)}{h}.$$

In fact, formula (1.47) implies for all $t > 0$ and all sufficiently small $h > 0$ the inequalities

$$-(1+\varepsilon) \sqrt{\frac{2 \ln |\ln h|}{h}} \leq \frac{W(t+h, \omega) - W(t, \omega)}{h} \leq +(1+\varepsilon) \sqrt{\frac{2 \ln |\ln h|}{h}},$$

and (1.48) yields the existence of a decreasing sequence of increments $h_k > 0$ such that

$$(-1)^{k+1} \frac{W(t+h_k, \omega) - W(t, \omega)}{h_k} \geq (1-\varepsilon) \sqrt{\frac{2 \ln |\ln h_k|}{h_k}} \quad \text{with} \quad \lim_{k \rightarrow \infty} h_k = 0^+.$$

As before, these inequalities show that for fixed $t > 0$, the difference quotient effectively oscillates between the two functions

$$-\sqrt{\frac{2 \ln |\ln h|}{h}} \quad \text{and} \quad +\sqrt{\frac{2 \ln |\ln h|}{h}} \quad \text{as} \quad h \rightarrow 0^+.$$

Yet, one can easily see that

$$\sqrt{\frac{2 \ln |\ln h|}{h}} \rightarrow +\infty \quad \text{as} \quad h \rightarrow 0^+,$$

which implies that the difference quotient $(W(t+h, \omega) - W(t, \omega))/h$ diverges in the limit $h \rightarrow 0^+$. In other words, this indicates that a typical Brownian path is nowhere differentiable. This is in fact true, as the following result shows.

Proposition 1.83 (Path regularity). *Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ denote a real-valued Wiener process over the probability space $(F, \mathcal{F}, \mathbb{P})$. Then there is a set $\tilde{F} \in \mathcal{F}$ with $\mathbb{P}(\tilde{F}) = 1$ such that for all $\omega \in \tilde{F}$ the following holds.*

- (a) *The path $W(\cdot, \omega)$ is nowhere differentiable.*
- (b) *For every $0 \leq \gamma < 1/2$ the path $W(\cdot, \omega)$ is Hölder-continuous with exponent γ at every $t_0 \geq 0$. In other words, for every $t_0 \geq 0$ and $0 \leq \gamma < 1/2$ there exists a constant $C > 0$ such that*

$$|W(t, \omega) - W(t_0, \omega)| \leq C |t - t_0|^\gamma$$

for all $t \in \mathbb{R}_0^+$ which are sufficiently close to t_0 .

- (c) For every $\gamma > 1/2$ the path $W(\cdot, \omega)$ is not Hölder-continuous with exponent γ at every point $t_0 \geq 0$.

If the result on the zero set of a typical Brownian path seemed strange, the above proposition takes this even further. Brownian paths are the complete opposite of the functions that are encountered in the standard undergraduate curriculum. While the reader will certainly have seen functions that are not differentiable everywhere (just think of piecewise defined functions such as the absolute value), Brownian paths are nowhere differentiable. In other words, we cannot associate a notion of velocity to a particle undergoing Brownian motion. In fact, due to the constant collisions which the particle experiences from the surrounding molecules, its “velocity” is always infinite! Despite this fact, Brownian paths are more regular than simple continuity. The stated Hölder continuity of order $\gamma < 1/2$ turns out to be extremely useful in the study of stochastic differential equations.

As our final result on the properties of a Brownian path we turn our attention to its monotonicity behavior, and to the related question of local maxima and minima. As a continuous function, every path $W(\cdot, \omega)$ has both a maximum and a minimum on every compact interval $[a, b] \subset \mathbb{R}_0^+$. As the reader might recall from calculus, however, the existence of these extreme values does not necessarily imply the existence of local maxima and minima, since the function might be monotone on $[a, b]$. For Brownian paths, this turns out to be impossible.

Proposition 1.84 (Local maxima and minima). *Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ denote a real-valued Wiener process over the probability space $(F, \mathcal{F}, \mathbb{P})$. Then there exists a set $\tilde{F} \in \mathcal{F}$ with probability $\mathbb{P}(\tilde{F}) = 1$ such that for all $\omega \in \tilde{F}$ the following holds.*

- (a) *For any interval $(a, b) \subset \mathbb{R}_0^+$ the path $W(\cdot, \omega)$ is neither monotone increasing nor monotone decreasing in (a, b) .*
- (b) *The set of points $t \in \mathbb{R}_0^+$ at which $W(\cdot, \omega)$ has a local maximum is countable and dense in \mathbb{R}_0^+ . All of these maxima are strict local maxima. In other words, in every interval $(a, b) \subset \mathbb{R}_0^+$ there are infinitely many strict local maxima of the path $W(\cdot, \omega)$.*
- (c) *The set of points $t \in \mathbb{R}_0^+$ at which $W(\cdot, \omega)$ has a local minimum is countable and dense in \mathbb{R}_0^+ . All of these minima are strict local minima.*

While part (a) of this proposition might seem surprising at first glance, it can easily be motivated using the kinetic theory of heat. Since a particle which undergoes Brownian motion experiences constant collisions from the surrounding molecules which approach from random directions, one has to expect that its direction is constantly changing.

1.4.3 • Differential Equations Perturbed by Additive Noise

After this long excursion into probability theory and the stochastic process known as Brownian motion or Wiener process, the reader might wonder what all of this has to do with differential equations. In the current section, we provide the missing link. For this, let us briefly return to the explanation of Brownian motion given by Albert Einstein. As most models are, his model is an idealization of a real-world phenomenon. While it does predict the correct overall behavior, it is only aimed at understanding the *position* of the particle, but it completely ignores its *velocity*. In fact, as we have seen in the last section,

the rigorous mathematical formulation of Brownian motion due to Norbert Wiener implies that a Brownian particle in Einstein's theory has infinite velocity at every point in time. In reality, this is impossible. Even though the particle will experience sudden velocity changes, its inertia will probably preclude these changes from being discontinuous. In other words, maybe a more refined model of Brownian motion should be formulated for the temporal evolution of its velocity, rather than the position itself.

The Langevin equation

A model of this type was proposed by the French physicist Paul Langevin in 1908. He studied the motion of a suspended particle in a fluid which is bombarded by the molecules of the surrounding solvent. His idea was to derive a differential equation for the velocity of the particle, and as Einstein before, he considered a one-dimensional equation for one component of the particle velocity vector. If $v(t)$ denotes this velocity at time t , then Langevin proposed the linear scalar differential equation

$$v' = -\alpha v + \sigma \xi(t), \quad \text{with} \quad v(0) = v_0, \quad (1.49)$$

which is nowadays called the *Langevin equation*. In this equation, the positive parameter $\alpha > 0$ describes the intensity of the frictional force $-\alpha v$ affecting the particle while it moves through the fluid, which is in fact the main force acting on the particle. The additive term $\sigma \xi(t)$, on the other hand, describes a residual highly irregular force experienced by the particle through the random bombardment of the solvent molecules. Based on our previous discussions it should not come as a surprise that Langevin thought of this residual force in a probabilistic way, even though the theory of probability was still in its infancy. The constant σ is used as a measure for the intensity of the residual force, and Langevin thought of ξ as *white noise*, i.e., a “stationary Gaussian stochastic process with mean zero and a constant spectral density on the entire real axis.” While we will refrain from giving precise definitions for the involved concepts, it suffices to say that white noise is a common model in the engineering literature for background noise which contains all frequencies with the same intensity.

Consequently, if we assume that the noise process $\xi = \xi(t, \omega)$ is actually a stochastic process $\xi : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ over a given probability space $(F, \mathcal{F}, \mathbb{P})$, then we can derive the explicit solution of the initial value problem (1.49) using the variation of parameters formula as

$$v(t, \omega) = e^{-\alpha t} \left(v_0 + \sigma \int_0^t e^{\alpha s} \xi(s, \omega) ds \right), \quad (1.50)$$

which is obtained separately for every $\omega \in F$. Unfortunately, however, the answer is not quite as simple as that...

White noise and the Wiener process

The problem with the above reasoning is that so far we have not discussed the white noise process ξ in more detail. In order for the integral in the velocity representation (1.50) to make sense for any fixed $\omega \in F$, we need to know that the paths $\xi(\cdot, \omega)$ are at least integrable. As it turns out, our problems already start considerably earlier than that — since from a mathematical point of view a white noise process with the above-cited properties cannot exist in the sense of a classical stochastic process. The fact that we will still be able to give the differential equation (1.49) a proper and rigorous interpretation, will be our first encounter with a stochastic differential equation.

Even though the white noise process ξ with the above properties cannot exist, and if one is not bothered by mundane questions such as existence, one can try to describe what probabilistic properties an “integrated version” of the white noise process should exhibit. Thus, if we formally define a new process via

$$W(t, \omega) = \int_0^t \xi(s, \omega) ds, \quad (1.51)$$

and if we formally lift the characteristics of ξ via standard properties of integration, then it can be shown that the process W should satisfy all the properties listed in Definition 1.79. In other words, the (nonexistent) stochastic process ξ which is referred to as white noise in the engineering literature, should be the (nonexistent) pathwise derivative of the Wiener process W .

At first glance, we have hit a dead end. However, just because the white noise process does not exist in the classical category of stochastic processes does not mean that it cannot be put on firm mathematical ground. In fact, it is possible to extend the notion of stochastic processes in a *distributional sense*, which leads to the notion of *generalized stochastic processes*. While a detailed description of this theory is beyond the scope of the current discussion, it suffices to say that it is possible to make sense of (1.51), albeit not in the standard sense of an integral. Once this has been accomplished, we are in a position to give (1.49) together with (1.51) a clear meaning. If we integrate both sides of (1.49) with respect to time over an interval $[0, t]$ we obtain

$$v(t) - v_0 = \int_0^t v'(s) ds = -\alpha \int_0^t v(s) ds + \sigma \underbrace{\int_0^t \xi(s, \omega) ds}_{=W(t, \omega)},$$

i.e., a solution of (1.49) is a stochastic process $v(t, \omega)$ which satisfies the integral equation

$$v(t, \omega) = v_0 - \alpha \int_0^t v(s, \omega) ds + \sigma W(t, \omega), \quad \text{for all } t \geq 0, \quad \omega \in F. \quad (1.52)$$

For this integral formula to make sense, we do not need that the paths of $v(t, \omega)$ are differentiable with respect to t . As long as the paths $v(\cdot, \omega)$ are continuous, one can evaluate the integral, and since the Wiener process has continuous paths, the right-hand side of (1.52) is indeed continuous with respect to t . In other words, even though the differential equation (1.49) cannot be interpreted within the classical theory, the integrated version (1.52) can. This is an instance of a *stochastic differential equation*. More generally, we have the following definition.

Definition 1.85 (Stochastic differential equation with additive noise). *For the following, consider an m -dimensional Wiener process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^m$ in the sense of Definition 1.79 over a probability space $(F, \mathcal{F}, \mathbb{P})$, and let ξ denote the associated formal white noise process given by*

$$\xi(t, \omega) = W'(t, \omega), \quad \text{in the generalized sense.} \quad (1.53)$$

Let $f : \mathbb{R}_0^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a continuously differentiable function, let $S \in \mathbb{R}^{d \times m}$ denote an arbitrary matrix, and consider an initial condition $x_0 \in \mathbb{R}^d$. Then we call the equation

$$x' = f(t, x) + S\xi(t, \omega) \quad \text{with} \quad x(0) = x_0 \quad (1.54)$$

a stochastic differential equation with additive noise.⁷

A stochastic process $x : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ is called a solution to the stochastic differential equation (1.54), if all paths $x(\cdot, \omega)$ are continuous with respect to time and if we have

$$x(t, \omega) = x_0 + \int_0^t f(s, x(s, \omega)) ds + SW(t, \omega) \quad (1.55)$$

for all $\omega \in F$.

If we denote the columns of the matrix S by $s_1, \dots, s_m \in \mathbb{R}^d$, and the components of ξ and W by $\xi_k \in \mathbb{R}$ and $W_k \in \mathbb{R}$, respectively, then one can rewrite the additive noise term in (1.54) as

$$S\xi(t, \omega) = \xi_1(t, \omega)s_1 + \dots + \xi_m(t, \omega)s_m,$$

and the last term in (1.55) reduces to

$$SW(t, \omega) = W_1(t, \omega)s_1 + \dots + W_m(t, \omega)s_m.$$

In other words, the additive noise term in (1.54) consists of a linear combination of m independent white noise processes.

The above definition does by no means cover all stochastic differential equations. Nevertheless, equations with additive noise are extremely important, and they will suffice for many of our applications later on. More specifically, we will generalize the concept of additive noise to the case of parabolic partial differential equations, and thereby address some interesting applied problems involving thermal noise.

Stochastic integrals with smooth integrands

As in the case of nonlinear ordinary differential equations, it is generally impossible to find explicit formulas for the solutions of a stochastic differential equation with additive noise. For this reason, the emphasis of this book is more on the development of numerical methods to approximate solution paths. Having said that, we do need a supply of test problems for which exact solutions are known, in order to test our numerical schemes later on. One of these is the original Langevin equation given by (1.49), and in the remainder of this section we will briefly sketch how its solution $v(t, \omega)$ can be computed. The resulting stochastic process is nowadays called *Ornstein-Uhlenbeck (velocity) process* after Leonard Ornstein and George Eugene Uhlenbeck, who in 1930 gave the first rigorous treatment of the Langevin equation. In fact, our discussion will show how solutions can be found for general linear stochastic differential equations with additive noise.

To begin our discussion, we return to the formal solution formula presented in (1.50). It was mentioned before that this formula cannot be interpreted in the sense of a classical integral for fixed $\omega \in F$, since the white noise process $\xi = W'$ does not exist as a classical stochastic process. However, it is possible to give this formula a precise mathematical meaning. For this we need the following definition.

⁷We would like to point out that our notation (1.54) for a stochastic differential equation with additive noise differs from the one generally used in stochastic analysis. In that context, the equation is usually written in the differentials notation given by

$$dx = f(t, x)dt + SdW \quad \text{with} \quad x(0) = x_0.$$

However, since stochastic differential equations will only play a minor part in this book, we adopt the “classical” ordinary differential equations notation.

Definition 1.86 (Stochastic integral with smooth integrand). Let $(F, \mathcal{F}, \mathbb{P})$ denote a probability space together with a scalar Wiener process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ in the sense of Definition 1.79. Furthermore, let $g : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ denote a stochastic process with continuously differentiable paths $g(\cdot, \omega)$ for all $\omega \in F$, and let $0 \leq a < b$ be arbitrary. Then the stochastic integral of g with respect to Brownian motion over the interval $[a, b]$ is the random variable which is defined by

$$\int_a^b g(s, \omega) dW(s, \omega) := g(b, \omega)W(b, \omega) - g(a, \omega)W(a, \omega) - \int_a^b g'(s, \omega)W(s, \omega) ds \quad (1.56)$$

for all $\omega \in F$, where $g'(\cdot, \omega)$ denotes the derivative of the path $g(\cdot, \omega)$. Notice that the integral on the right-hand side of (1.56) is defined in the classical sense for all arguments $\omega \in F$, since the integrand $g'(s, \omega)W(s, \omega)$ is continuous with respect to s . The expression on the left-hand side is our notation for the stochastic integral.

The specific form of the right-hand side of (1.56) can be motivated as follows. Consider the integral in (1.50), and then use $\xi = W'$ to formally apply the integration by parts formula, which furnishes

$$\int_a^b g(s, \omega)W'(s, \omega) ds = g(b, \omega)W(b, \omega) - g(a, \omega)W(a, \omega) - \int_a^b g'(s, \omega)W(s, \omega) ds .$$

However, justifying the validity of this formal procedure is far from trivial, and heavily relies on the fact that the integrand g has continuously differentiable paths. We will address the question of integrands with non-differentiable paths, such as for example the function $g(s, \omega) = W(s, \omega)$, briefly at the end of this section.

For later reference, we present without proof the following lemma from stochastic analysis, which shows that if the integrand in Definition 1.86 does not depend on ω , i.e., if g is a deterministic function, then the stochastic integral is in fact a normally distributed random variable.

Lemma 1.87 (Stochastic integral as Gaussian random variable). Consider a probability space $(F, \mathcal{F}, \mathbb{P})$ together with a scalar Wiener process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}$ in the sense of Definition 1.79. Furthermore, let $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ denote a continuously differentiable real-valued function, and let $0 \leq a < b$ be arbitrary. Then the stochastic integral of g with respect to Brownian motion over the interval $[a, b]$ as defined in Definition 1.86 is a Gaussian random variable with mean $\mu = 0$ and variance σ^2 given by

$$\sigma^2 = \int_a^b g(s)^2 ds .$$

See also Definition 1.78.

Solution formula for linear stochastic equations with additive noise

We are now finally in a position to present an explicit formula for the solutions of linear stochastic differential equations with additive noise. In order to formulate this result, we need the following notion from elementary differential equations. Consider a linear system of differential equations of the form

$$x' = Ax , \quad \text{for some } A \in \mathbb{R}^{d \times d} .$$

For each initial condition of the form $x(0) = x_0$ there exists a unique solution $x(t)$ defined for all $t \in \mathbb{R}$. Now define a matrix $\Phi(t) \in \mathbb{R}^{d \times d}$ in such a way that the k -th column of $\Phi(t)$ is the solution of the initial value problem

$$x' = Ax \quad \text{with} \quad x(0) = e_k \in \mathbb{R}^d,$$

where e_k denotes the k -th canonical basis vector in \mathbb{R}^d . Then we call $\Phi(t)$ the *fundamental matrix* associated with the linear equation $x' = Ax$. One can easily check that for arbitrary initial conditions $x_0 \in \mathbb{R}^d$ the function $x(t) = \Phi(t)x_0$ solves the initial value problem $x' = Ax$ with $x(0) = x_0$.

Example 1.88. In preparation for our discussion of the Ornstein-Uhlenbeck process, consider the two-dimensional linear differential equation

$$\frac{d}{dt} \begin{pmatrix} r \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\alpha \end{pmatrix} \begin{pmatrix} r \\ v \end{pmatrix}.$$

Then one can easily verify that the associated fundamental matrix is given by

$$\Phi(t) = \begin{pmatrix} 1 & (1-e^{-\alpha t})/\alpha \\ 0 & e^{-\alpha t} \end{pmatrix},$$

as long as we have $\alpha \neq 0$. ■

After these preparations, the following result presents an explicit solution formula for linear stochastic differential equations with additive noise.

Lemma 1.89 (Linear stochastic equation with additive noise). *For two given matrices $A \in \mathbb{R}^{d \times d}$ and $S \in \mathbb{R}^{d \times m}$, and an initial condition $x_0 \in \mathbb{R}^d$ consider the linear stochastic differential equation with additive noise given by*

$$x' = Ax + S\xi \quad \text{with} \quad x(0) = x_0,$$

where $\xi = W'$ for an m -dimensional Wiener process over a given probability space $(F, \mathcal{F}, \mathbb{P})$. Then the unique solution of this initial value problem is given by the stochastic process

$$x(t, \omega) = \Phi(t)x_0 + \sum_{k=1}^m \Phi(t) \int_0^t \Phi(\tau)^{-1} s_k dW_k(\tau, \omega),$$

where $s_1, \dots, s_m \in \mathbb{R}^d$ denote the columns of S , and $W_1, \dots, W_m \in \mathbb{R}$ the components of W .

The above lemma will allow us to determine a number of test problems for our numerical studies later on in this book. For now we use it to finish our story of Brownian motion with the study of the Langevin equation.

Example 1.90. As we mentioned at the beginning of this section, the Langevin equation (1.49) was proposed as a model for the temporal evolution of the velocity v of a Brownian particle. If we denote its position by $r(t)$, and assume that at time $t = 0$ the particle is at position r_0 and has velocity v_0 , then the pair (r, v) satisfies the linear stochastic differential equation

$$\frac{d}{dt} \begin{pmatrix} r \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\alpha \end{pmatrix} \begin{pmatrix} r \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} \xi, \quad \text{with} \quad \begin{pmatrix} r(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} r_0 \\ v_0 \end{pmatrix},$$

where ξ denotes the generalized derivative of a standard real-valued Wiener process. Then Lemma 1.89 together with Example 1.88 yields

$$\begin{aligned} r(t, \omega) &= r_0 + \frac{v_0}{\alpha} (1 - e^{-\alpha t}) + \frac{\sigma}{\alpha} W(t, \omega) - \frac{\sigma}{\alpha} e^{-\alpha t} \int_0^t e^{\alpha s} dW(s, \omega), \\ v(t, \omega) &= e^{-\alpha t} v_0 + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW(s, \omega). \end{aligned}$$

For any $t \geq 0$, Lemma 1.87 implies that the random variables $r(t, \cdot)$ and $v(t, \cdot)$ are Gaussian random variables. In fact, one can show that $r(t, \cdot)$ has mean $\mu_r(t)$ and variance $\sigma_r^2(t)$ given by

$$\mu_r(t) = r_0 + \frac{v_0}{\alpha} (1 - e^{-\alpha t}) \quad \text{and} \quad \sigma_r^2(t) = \frac{\sigma^2}{\alpha^2} t + \frac{\sigma^2}{2\alpha^3} (4e^{-\alpha t} - 3 - e^{-2\alpha t}),$$

and $v(t, \cdot)$ has mean $\mu_v(t)$ and variance $\sigma_v^2(t)$ given by

$$\mu_v(t) = e^{-\alpha t} v_0 \quad \text{and} \quad \sigma_v^2(t) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}).$$

The above formulas are of particular interest for the special case $r_0 = v_0 = 0$. Using the parameters α and σ , Langevin was able to express the constant D used in Einstein's formula (1.44) by the ratio

$$D = \frac{\sigma^2}{2\alpha^2},$$

which implies that the random variable $r(t, \cdot)$ is normally distributed with mean zero and variance

$$\sigma_r^2(t) = 2D t \left(1 + \frac{4e^{-\alpha t} - 3 - e^{-2\alpha t}}{2\alpha t} \right).$$

One can easily verify that the function in parentheses is increasing with respect to t . Moreover, it converges to 0 for $t \rightarrow 0$ and it converges to 1 as $t \rightarrow \infty$. Thus, for large enough times t , the *Ornstein-Uhlenbeck position process* $r(t, \cdot)$ has the same distribution as the Brownian motion process introduced by Einstein. In contrast to the latter, however, the formulation due to Langevin leads to a particle with a finite velocity. ■

1.4.4 • A First Glance at Ito's Stochastic Calculus

So far, our description of stochastic integration seems to have followed a strange pattern. Whenever we encounter an expression that cannot be explained in the classical sense, we just integrate, or use formulas from standard integration theory, until we have obtained a formally equivalent representation which only requires the continuity properties of the Wiener process and which does make sense in the classical theory. In other words, it seems that stochastic integration essentially is just a reformulation of classical integration theory.

It would be utterly dishonest to leave the reader with this impression, since it could not be further from the truth. Stochastic integration is completely different from the classical theory, and in fact, in most cases the integration rules that you have learned to appreciate from calculus are no longer valid in the stochastic context. We will spend the remainder of this section to give a brief discussion of what problems are encountered,

and how these can be resolved using stochastic calculus. Most of the presented material, however, will not be necessary for the purposes of this book — it is solely meant as a brief introduction into the central questions of stochastic integration. The only exceptions are the definition of a general stochastic differential equation and the examples at the very end of the section, since these will be used later on to discuss convergence rates for numerical approximation methods for stochastic differential equations, and for describing some surprising connections between stochastic ordinary differential equations and elliptic or parabolic partial differential equations.

The simplest stochastic integral

Our discussion will mostly revolve around the arguably simplest integral that is not covered by our Definition 1.86. Since in this definition we assumed that the paths of the integrand are differentiable with probability one, consider the integral

$$\int_a^b W(t, \omega) dW(t, \omega), \quad (1.57)$$

where W denotes a standard Wiener process. If one follows our discussion above, it seems tempting to replace the term $dW(t, \omega)$ by $W'(t, \omega)dt$, and then formally use integration by parts to obtain

$$\int_a^b W(t, \omega) W'(t, \omega) dt = W(b, \omega)^2 - W(a, \omega)^2 - \int_a^b W'(t, \omega) W(t, \omega) dt,$$

which then seems to imply

$$\int_a^b W(t, \omega) dW(t, \omega) \stackrel{?}{=} \frac{W(b, \omega)^2 - W(a, \omega)^2}{2}.$$

We will see in a moment that while it is possible to define the stochastic integral on the left-hand side, the result turns out not to be given by the formula on the right-hand side.

In order to explain why the above formula does not hold, we have to take a step back and ask ourselves which properties of an integral and of the associated integration theory are absolutely essential and have to be preserved at any cost. Obviously, we would like an integral of the form

$$\int_a^b X(t, \omega) dW(t, \omega)$$

to still be linear with respect to the integrand $X(t, \omega)$. One can readily verify that this linearity implies that the integral of the zero function $X(t, \omega) \equiv 0$ has to vanish, as well as the following partitioning property of the integral: Let $a = t_0 < t_1 < \dots < t_m = b$ denote a partition of $[a, b]$. Then the assumed linearity furnishes

$$\int_a^b X(t, \omega) dW(t, \omega) = \sum_{k=1}^m \int_{t_{k-1}}^{t_k} X(t, \omega) dW(t, \omega).$$

In addition to linearity, the second crucial property of an integral is its continuity in the following sense. If the integrand $X(t, \omega)$ is only slightly perturbed, the integral should also only change slightly.

By combining the linearity and continuity properties of the integral, one can obtain approximations for the value of the integral as follows. Assume for now that $X(t, \omega)$ is an arbitrary stochastic process with continuous paths. Furthermore, suppose that we are given a partition $a = t_0 < t_1 < \dots < t_m = b$ of the interval $[a, b]$ such that on each of the subintervals $[t_{k-1}, t_k]$ the function $X(\cdot, \omega)$ is essentially constant, and close to the value $X(t_k^*, \omega)$, for some $t_k^* \in [t_{k-1}, t_k]$. Then the linearity and continuity of the integral yield

$$\begin{aligned} \int_a^b X(t, \omega) dW(t, \omega) &= \sum_{k=1}^m \int_{t_{k-1}}^{t_k} X(t, \omega) dW(t, \omega) \\ &\approx \sum_{k=1}^m \int_{t_{k-1}}^{t_k} X(t_k^*, \omega) dW(t, \omega) \\ &= \sum_{k=1}^m X(t_k^*, \omega) \int_{t_{k-1}}^{t_k} 1 dW(t, \omega) \\ &= \sum_{k=1}^m X(t_k^*, \omega) (W(t_k, \omega) - W(t_{k-1}, \omega)), \end{aligned}$$

where the last equality is a consequence of Definition 1.86. In other words, the basic linearity and continuity assumptions imply that an integral can be approximated by Riemann-type sums.

Let us now return to the special case (1.57). Based on the above discussion it is tempting to define this integral as the limit

$$\int_a^b W(t, \omega) dW(t, \omega) = \lim_{\delta(P) \rightarrow 0} \sum_{k=1}^m W(t_k^*, \omega) (W(t_k, \omega) - W(t_{k-1}, \omega)), \quad (1.58)$$

where $a = t_0 < t_1 < \dots < t_m = b$ is an arbitrary partition P of the interval $[a, b]$, we choose $t_k^* \in [t_{k-1}, t_k]$, and $\delta(P) = \max_{k=1, \dots, m} |t_k - t_{k-1}|$ for the given partition P .

But does this limit exist? The answer to this question will finally open the door to stochastic analysis. It turns out that the existence of the limit in (1.58) is an extremely delicate matter:

- The limit on the right-hand side of (1.58) cannot be interpreted in an ω -wise fashion. More precisely, if we consider any $\omega \in F$, then one can show that the limit only exists if the path $W(\cdot, \omega)$ is a function with *bounded variation*. While the precise definition of this concept will not be of interest to us, one can show that any function of bounded variation has to be differentiable at almost all $t \geq 0$ — and we have already seen that the paths of the Wiener process are nowhere differentiable.
- In order to still give the limit in (1.58) a meaning, one has to consider the right-hand side as a sequence of random variables over the probability space $(F, \mathcal{F}, \mathbb{P})$ and use a convergence concept on the level of random variables. Without going into details, probability theory provides the concept of convergence in $L^2(\mathbb{P})$, and one can show that using this type of convergence the right-hand side indeed has a limit.
- Unfortunately, however, the location of the sampling points t_k^* does have a profound impact. The reader might recall from calculus that for the usual Riemann integral the choice of the $t_k^* \in [t_{k-1}, t_k]$ does not affect the resulting limit. This is far

from true in the stochastic setting. For example, if we choose a number $\alpha \in [0, 1]$, consider the interval $[a, b] = [0, T]$, and in (1.58) consider

$$t_k^* = t_{k-1} + \alpha(t_k - t_{k-1}),$$

then the $L^2(\mathbb{P})$ -limit in the right-hand side of (1.58) exists and is given by

$$\lim_{\delta(P) \rightarrow 0} \sum_{k=1}^m W(t_k^*, \omega)(W(t_k, \omega) - W(t_{k-1}, \omega)) = \frac{W(t, \omega)^2 - t}{2} + \alpha t.$$

Clearly this limit depends on the location of the points t_k^* as described by the parameter α !

In other words, when attempting to define the stochastic integral of Brownian motion with respect to itself, one cannot simply lift the classical Riemann sum approach to the space of square integrable random variables.

Stochastic Ito integration

It was one of the great contributions of the Japanese mathematician Kiyoshi Ito to realize that despite the above subtleties it is possible to define a stochastic integral in a meaningful way. He introduced the following definition.

Definition 1.91 (Stochastic Ito integral). Let $W(t, \omega)$ denote the Wiener process introduced in Definition 1.79 over some probability space $(F, \mathcal{F}, \mathbb{P})$, and let $\xi(t, \omega)$ denote the associated white noise process, i.e., the generalized derivative of the Wiener process. Furthermore, consider a suitable⁸ stochastic process $X : [a, b] \times F \rightarrow \mathbb{R}$ which has continuous paths with probability one, with underlying interval $[a, b] \subset \mathbb{R}_0^+$. Then the stochastic Ito integral of X with respect to W is defined as

$$\int_a^b X(\tau, \omega) dW(\tau, \omega) = \lim_{\delta(P) \rightarrow 0} \sum_{k=1}^m X(\tau_{k-1}, \omega)(W(\tau_k, \omega) - W(\tau_{k-1}, \omega)),$$

where the limit is taken in the normed linear space $L^2(\mathbb{P})$ of square-integrable random variables over $(F, \mathcal{F}, \mathbb{P})$. Moreover, in the above definition, the letter P denotes arbitrary partitions of the form $a = \tau_0 < \tau_1 < \dots < \tau_m = b$, and we define the maximal gap size of a partition as $\delta(P) = \max\{\tau_1 - \tau_0, \dots, \tau_m - \tau_{m-1}\}$.

If we replace the upper limit of integration by the independent variable t , then we interchangeably use the notations

$$Y(t, \omega) = Y(a, \omega) + \int_a^t X(\tau, \omega) dW(\tau, \omega) \quad \text{and} \quad Y' = X \xi(t, \omega)$$

to describe the fact that the stochastic process Y is obtained via stochastic integration of the stochastic process X with respect to W over the interval $[a, t]$.

Notice that in the Riemann sums which are used in the definition of the Ito integral, the integrand $X(t, \omega)$ is always evaluated at the *left endpoints* of the partition intervals $[t_{k-1}, t_k]$, which corresponds to the choice $\alpha = 0$ in the previous discussion of the integral of the Wiener process with respect to itself.⁹

⁸We refrain from presenting the precise assumptions on X which are necessary for the definition of the stochastic integral, and instead refer the interested reader to [1].

⁹The choice $\alpha = 1/2$ basically leads (we are imprecise here) to a different notion of stochastic integral which is called *Stratonovich integral*. Both the Stratonovich and the Ito theory have their merits, and there are ways to relate one to the other.

The stochastic Ito integral introduced in Definition 1.91 satisfies the linearity and continuity assumptions that we postulated earlier on, as well as the identity which we used in Definition 1.86. Unfortunately, however, many properties of the standard Riemann integral are no longer valid. Since a comprehensive discussion of the stochastic calculus associated with Ito's integral is beyond the scope of this book, we restrict ourselves to presenting the stochastic version of the classical chain rule, which is more commonly known as *Ito's formula*.

For this, we briefly recall the chain rule, yet in a version which is more suitable for our intended stochastic extension. Assume that the function $X : I \rightarrow \mathbb{R}$, defined on some interval $I \subset \mathbb{R}_0^+$, is continuously differentiable with derivative Q , i.e., assume that

$$X'(t) = Q(t) \quad \text{for all } t \in I.$$

In addition, let $S : I \times \mathbb{R} \rightarrow \mathbb{R}$ denote a continuously differentiable function of two variables t and x , and define a new function Y as

$$Y(t) = S(t, X(t)) \quad \text{for all } t \in I.$$

Then the chain rule states that the derivative of Y is given by

$$Y'(t) = \frac{\partial S}{\partial t}(t, X(t)) + \underbrace{\frac{\partial S}{\partial x}(t, X(t)) Q(t)}_{=X'(t)}.$$

In other words, we can abbreviate the classical chain rule by the implication

$$X' = Q \quad \text{and} \quad Y = S(t, X) \quad \text{implies} \quad Y' = \frac{\partial S}{\partial t}(t, X) + \frac{\partial S}{\partial x}(t, X) Q, \quad (1.59)$$

by omitting the argument t .

Turning our attention to the stochastic situation again, we now assume that we are given a stochastic process $X : I \times F \rightarrow \mathbb{R}$ in the form

$$X(t, \omega) = X(a, \omega) + \int_a^t Q(\tau, \omega) d\tau + \int_a^t R(\tau, \omega) dW(\tau, \omega)$$

for suitable stochastic processes Q and R which have continuous paths with probability one. Note that in this definition the first integral is a regular Riemann integral with respect to τ , while the second one is the stochastic integral defined in Definition 1.91, and we abbreviate this definition of X formally by writing

$$X' = Q + R \xi(t, \omega),$$

where $\xi(t, \omega)$ denotes the white noise process, i.e., the generalized derivative of the Wiener process $W(t, \omega)$. Now consider a function $S : I \times \mathbb{R} \rightarrow \mathbb{R}$, and define a new stochastic process Y as

$$Y(t, \omega) = S(t, X(t, \omega)) \quad \text{for all } t \in I \quad \text{and} \quad \omega \in F.$$

Is it possible to write this new stochastic process in the form

$$Y(t, \omega) = Y(a, \omega) + \int_a^t \tilde{Q}(\tau, \omega) d\tau + \int_a^t \tilde{R}(\tau, \omega) dW(\tau, \omega),$$

for suitable processes \tilde{Q} and \tilde{R} ? In other words, can we determine processes \tilde{Q} and \tilde{R} such that

$$Y' = \tilde{Q} + \tilde{R} \xi(t, \omega),$$

if we formally differentiate the integral equation as above? At first glance, it is tempting to assume that these processes should be given by

$$\tilde{Q}(t, \omega) = \frac{\partial S}{\partial t}(t, X(t)) + \frac{\partial S}{\partial x}(t, X(t)) Q(t, \omega),$$

$$\tilde{R}(t, \omega) = \frac{\partial S}{\partial x}(t, X(t)) R(t, \omega),$$

in accordance with the classical chain rule — but this turns out not to be quite true. Ito realized that while the above formula for \tilde{R} is in fact correct, the formula for \tilde{Q} is missing a term which depends on the *second* derivative of S with respect to X . More precisely, Ito proved the following result, which in view of later applications is formulated for the vector-valued case.

Theorem 1.92 (Ito's formula). *Let $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^m$ denote an m -dimensional Wiener process in the sense of Definition 1.79 over a probability space $(F, \mathcal{F}, \mathbb{P})$, and let $\xi(t, \omega)$ denote the associated white noise process, i.e., the componentwise generalized derivative of the Wiener process. Furthermore, for an interval $I \subset \mathbb{R}_0^+$ consider a stochastic process $X : I \times F \rightarrow \mathbb{R}^d$ of the form*

$$X(t, \omega) = X(a, \omega) + \int_a^t Q(\tau, \omega) d\tau + \int_a^t R(\tau, \omega) dW(\tau, \omega),$$

where $Q : I \times F \rightarrow \mathbb{R}^d$ and $R : I \times F \rightarrow \mathbb{R}^{d \times m}$ denote suitable stochastic processes which have continuous paths with probability one, i.e., assume that

$$X' = Q + R \xi(t, \omega).$$

Finally, assume that $S : I \times \mathbb{R}^d \rightarrow \mathbb{R}^k$ is a twice continuously differentiable function, define a new stochastic process Y as

$$Y(t, \omega) = S(t, X(t, \omega)) \quad \text{for all } t \in I \quad \text{and } \omega \in F,$$

and introduce the stochastic processes

$$\begin{aligned} \tilde{Q}(t, \omega) &= \frac{\partial S}{\partial t}(t, X(t)) + \frac{\partial S}{\partial x}(t, X(t)) Q(t, \omega) \\ &\quad + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 S}{\partial x_i \partial x_j}(t, X(t)) (R(t, \omega) R(t, \omega)^t)_{i,j}, \\ \tilde{R}(t, \omega) &= \frac{\partial S}{\partial x}(t, X(t)) R(t, \omega), \end{aligned}$$

where $(\partial S / \partial x)(t, x)$ denotes a $k \times d$ -matrix, $(\partial S / \partial t)(t, x)$ and $(\partial^2 S / \partial x_i \partial x_j)(t, x)$ are k -dimensional column vectors, and $R(t, \omega)^t$ denotes the transpose of the matrix $R(t, \omega)$. Then we have

$$Y(t, \omega) = Y(a, \omega) + \int_a^t \tilde{Q}(\tau, \omega) d\tau + \int_a^t \tilde{R}(\tau, \omega) dW(\tau, \omega),$$

which can be written in more compact form as

$$\begin{aligned} Y' = & \left(\frac{\partial S}{\partial t}(t, X) + \frac{\partial S}{\partial x}(t, X) Q + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 S}{\partial x_i \partial x_j}(t, X) (RR^t)_{i,j} \right) \\ & + \frac{\partial S}{\partial x}(t, X) R \xi(t, \omega). \end{aligned} \quad (1.60)$$

Notice that for the special case $R \equiv 0$ we recover the classical chain rule as in (1.59).

Before turning our attention to more general stochastic differential equations, we present a brief example which illustrates Ito's formula for the case $k = m = d = 1$ and for integrands depending on Brownian motion. For this, let $h : \mathbb{R} \rightarrow \mathbb{R}$ denote a continuously differentiable function with antiderivative H , and our goal is to compute the stochastic integral

$$\int_a^b h(W(\tau, \omega)) dW(\tau, \omega).$$

This can be accomplished by observing that $X = W$ satisfies

$$X' = \underbrace{0}_{=Q} + \underbrace{1}_{=R} \xi(t, \omega),$$

and an application of Theorem 1.92 with $S = H$ then shows that $Y = S(X)$ satisfies

$$Y' = \left(\underbrace{\frac{\partial S}{\partial t}(t, X)}_{=0} + \underbrace{\frac{\partial S}{\partial x}(t, X)}_{=h(X)} \underbrace{Q}_{=0} + \frac{1}{2} \underbrace{\frac{\partial^2 S}{\partial x^2}(t, X)}_{=h'(X)} \underbrace{R^2}_{=1} \right) + \underbrace{\frac{\partial S}{\partial x}(t, X)}_{=h(X)} \underbrace{R}_{=1} \xi(t, \omega),$$

which is equivalent to

$$H(W(b, \omega)) - H(W(a, \omega)) = \frac{1}{2} \int_a^b h'(W(\tau, \omega)) d\tau + \int_a^b h(W(\tau, \omega)) dW(\tau, \omega).$$

This finally implies

$$\int_a^b h(W(\tau, \omega)) dW(\tau, \omega) = H(W(b, \omega)) - H(W(a, \omega)) - \frac{1}{2} \int_a^b h'(W(\tau, \omega)) d\tau,$$

where the last integral is a standard pathwise Riemann integral. Notice that if the classical chain rule would still apply in the stochastic case, then this last integral would not be present. For the specific case $h(x) = x$ the above formula furnishes

$$\int_a^b W(\tau, \omega) dW(\tau, \omega) = \frac{W(b, \omega)^2 - W(a, \omega)^2}{2} - \frac{b-a}{2},$$

in accordance with our earlier discussion.

General stochastic differential equations

We close this section by introducing the concept of a general stochastic differential equation. In addition, we will demonstrate through examples how in certain situations Ito's formula can be used to obtain explicit solution representations. These examples will be used later on to test numerical approximation methods for stochastic differential equations. But first, we need to state the following central definition.

Definition 1.93 (General stochastic differential equation). Consider an m -dimensional Wiener process $W : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^m$ in the sense of Definition 1.79 over a probability space $(F, \mathcal{F}, \mathbb{P})$, and let $\xi(t, \omega)$ denote the associated white noise process, i.e., the componentwise generalized derivative of the Wiener process. In addition, we consider two smooth functions $f : \mathbb{R}_0^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $g : \mathbb{R}_0^+ \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$, i.e., the first function is vector- and the second one matrix-valued. Then a general stochastic differential equation is an equation of the form

$$x' = f(t, x) + g(t, x)\xi(t, \omega), \quad \text{with } x(t_0) = x_0, \quad (1.61)$$

where $(t_0, x_0) \in \mathbb{R}_0^+ \times \mathbb{R}^d$ specifies an arbitrary initial condition.

A stochastic process $x : \mathbb{R}_0^+ \times F \rightarrow \mathbb{R}^d$ with continuous paths is called a solution of (1.61) if it satisfies the integral equation

$$x(t, \omega) = x_0 + \int_{t_0}^t f(\tau, x(\tau, \omega)) d\tau + \int_{t_0}^t g(\tau, x(\tau, \omega)) dW(\tau, \omega). \quad (1.62)$$

In this equation, the first integral is a classical vector-valued Riemann integral, while the second one is a vector-valued stochastic Ito integral. In both cases, the integral is defined componentwise.

Stochastic differential equations of the above form have been extensively studied over the last few decades, and by now there exists a comprehensive theory which parallels the one for ordinary differential equations. For more details, we refer the reader to [1, 29, 54]. For the purposes of this text, we satisfy ourselves with the following examples. In order to keep the presentation simple, these only consider the scalar case $d = m = 1$.

Example 1.94. As a first example, we consider the general scalar linear stochastic differential equation

$$x' = ax + bx\xi(t, \omega), \quad \text{with } x(0) = x_0, \quad (1.63)$$

where a and b are arbitrary real constants. Motivated by the solution formula for nonautonomous linear scalar ordinary differential equations, it seems reasonable to expect that the solution of (1.63) has to involve the exponential of a linear combination of t and of $W(t, \omega)$. Thus, consider the stochastic process

$$Y(t, \omega) = e^{\alpha t + \beta W(t, \omega)} \quad \text{for constants } \alpha, \beta \in \mathbb{R}.$$

Using Ito's formula from Theorem 1.92 with $X = W$ and $S(t, x) = e^{\alpha t + \beta x}$, together with the fact that $X' = 0 + 1 \xi(t, \omega)$, one can then readily see that $Y = S(t, X)$ satisfies

$$Y' = \left(\underbrace{\frac{\partial S}{\partial t}(t, X) + \frac{\partial S}{\partial x}(t, X)}_{= \alpha Y} \underbrace{Q}_{= 0} + \frac{1}{2} \underbrace{\frac{\partial^2 S}{\partial x^2}(t, X)}_{= \beta^2 Y} \underbrace{R^2}_{= 1} \right) + \underbrace{\frac{\partial S}{\partial x}(t, X)}_{= \beta Y} \underbrace{R}_{= 1} \xi(t, \omega),$$

which can be simplified to

$$Y' = \left(\alpha + \frac{\beta^2}{2} \right) Y + \beta Y \xi(t, \omega).$$

In other words, the process $Y(t, \omega)$ does in fact solve a linear scalar stochastic differential equation of the form (1.63), for coefficients a and b which satisfy $a = \alpha + \beta^2/2$ and $b = \beta$. Solving these equations for α and β implies that the stochastic process

$$x(t, \omega) = x_0 e^{\left(\alpha - \frac{\beta^2}{2} \right)t + b W(t, \omega)} \quad (1.64)$$

solves (1.63). ■

Example 1.95. For our second example, we consider the nonlinear stochastic differential equation

$$x' = -a^2 \sin x \cos^3 x + a \cos^2 x \xi(t, \omega), \quad \text{with } x(0) = x_0, \quad (1.65)$$

where $a \in \mathbb{R}$ is a real constant. Even though it is not obvious at this point why we do so, in this case, we consider the stochastic process

$$Y(t, \omega) = \arctan(a W(t, \omega) + \tan x_0).$$

Again we use Theorem 1.92 with $X = W$, but now for the transformation

$$S(t, x) = \arctan(ax + \tan x_0).$$

As in the previous example, this implies that $Y = S(t, X)$ satisfies

$$\begin{aligned} Y' &= \left(\underbrace{\frac{\partial S}{\partial t}(t, X) + \frac{\partial S}{\partial x}(t, X)}_{=0} \underbrace{Q}_{=0} + \frac{1}{2} \underbrace{\frac{\partial^2 S}{\partial x^2}(t, X)}_{=-2a^2 \sin Y \cos^3 Y} \underbrace{R^2}_{=1} \right. \\ &\quad \left. + \underbrace{\frac{\partial S}{\partial x}(t, X)}_{=a \cos^2 Y} \underbrace{R}_{=1} \xi(t, \omega), \right. \end{aligned}$$

and this shows that the stochastic process

$$x(t, \omega) = \arctan(a W(t, \omega) + \tan x_0) \quad (1.66)$$

solves (1.65). Notice that this solution formula is not defined for initial conditions of the form $x_0 = \pi/2 + k\pi$, where $k \in \mathbb{Z}$. However, one can easily check that for these initial conditions, the solution process turns out to be constant, i.e., we have $x(t, \omega) = \pi/2 + k\pi$ for all $t \geq 0$ and $\omega \in F$. ■

Example 1.96. For our third and last example, we consider the nonlinear stochastic differential equation

$$x' = a^2(x^3 - x) + a(1 - x^2) \xi(t, \omega), \quad \text{with } x(0) = x_0, \quad (1.67)$$

where $a \in \mathbb{R}$ is a real constant. In this case, we consider the stochastic process

$$Y(t, \omega) = 1 - \frac{2(1-x_0)}{(1+x_0)e^{2aW(t,\omega)} + (1-x_0)}.$$

Again we use Theorem 1.92 with $X = W$ applied to the transformation

$$S(t, x) = 1 - \frac{2(1-x_0)}{(1+x_0)e^{2ax} + (1-x_0)}.$$

As in the previous examples, this implies that $Y = S(t, X)$ satisfies

$$\begin{aligned} Y' &= \left(\underbrace{\frac{\partial S}{\partial t}(t, X) + \frac{\partial S}{\partial x}(t, X)}_{=0} \underbrace{Q}_{=0} + \frac{1}{2} \underbrace{\frac{\partial^2 S}{\partial x^2}(t, X)}_{=2a^2(Y^3-Y)} \underbrace{R^2}_{=1} \right) \\ &\quad + \underbrace{\frac{\partial S}{\partial x}(t, X)}_{=a(1-Y^2)} \underbrace{R}_{=1} \xi(t, \omega), \end{aligned}$$

and this shows that the stochastic process

$$x(t, \omega) = 1 - \frac{2(1-x_0)}{(1+x_0)e^{2aW(t,\omega)} + (1-x_0)} \tag{1.68}$$

solves (1.67). In fact, this solution exhibits blow-up behavior. If the initial condition x_0 lies outside the interval $[-1, 1]$, then with probability one solutions will blow up in finite time. In contrast, for initial conditions between -1 and $+1$ the solutions exist for all positive times. We will study the solutions to this example in more detail later, in the context of numerical methods for stochastic differential equations. ■

1.5 • Review Questions and Problems

1.5.1 • Review Questions

Review Question 1.5.1. For each of the following ordinary differential equations, determine the order, and determine whether or not they are linear and autonomous. If the equation is linear, determine whether it is homogeneous.

- (a) $x''' + 4tx' + 6x + \cos t = 0$.
- (b) $x''x' + x \sin t = 0$.
- (c) The coupled system $x'y + 2y = 0$ and $y''x + tx' = 0$.
- (d) $5x'' - \cos(x') = 4$.

Review Question 1.5.2. For each of the following equations, determine their order, and determine whether they are linear or nonlinear. If an equation is linear, determine

whether it is homogeneous or nonhomogeneous. Note that in each case, u is a function of t and x , and c and ε are constants (independent of t and x).

$$\begin{array}{ll} \text{(a)} & u_{tt} = c^2 t^3 u_{xx} \\ \text{(b)} & u_{tt} = x^2 u_{xxt} + 1 \end{array} \quad \begin{array}{ll} \text{(c)} & u_t = -(\varepsilon^2 u_{xx} + u - u^3)_{xx} \\ \text{(d)} & u_{tt} - u_{xx}^2 = 3u + xt \end{array}$$

Review Question 1.5.3. For the function $F(x, y) = x^2y$ at the specific point $(1, 2)$, show that its gradient is given by $\nabla F(1, 2) = (4, 1)$. Then consider the planar domain $\Omega = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < y < 3\}$, and note that $(1, 2) \in \partial\Omega$. Find the unit outward normal vector to Ω at the point $(1, 2)$, and then determine the directional derivative $\partial F / \partial n(1, 2)$.

Review Question 1.5.4. In two and three dimensions, verify the formula $\operatorname{div}(\nabla G) = \Delta G$ (given in Theorem 1.37).

Review Question 1.5.5. Let $u(t, x)$ and $v(t, x)$ be two given solutions of the heat equation $w_t = kw_{xx}$ on the domain $\Omega \subset \mathbb{R}$, subject to homogeneous Dirichlet boundary conditions. Show that for any $a, b \in \mathbb{R}$, $w(t, x) = au(t, x) + bv(t, x)$ is also a solution to the heat equation with homogeneous Dirichlet boundary conditions.

Review Question 1.5.6. Assume that $y(t, x)$ and $z(t, x)$ both solve the initial-boundary value problem

$$\begin{aligned} u_t &= ku_{xx}, & u(t, 0) &= f_1(t), & u(t, L) &= f_2(t), & \text{for } t > 0, \\ u(0, x) &= g(x), & & & & & \text{for } 0 < x < L. \end{aligned}$$

Show that the difference $w(t, x) = y(t, x) - z(t, x)$ solves the heat equation subject to homogeneous Dirichlet boundary conditions and initial condition $w(0, x) = 0$ for all $x \in (0, L)$.

Review Question 1.5.7. Verify that the function $f(x, y) = \ln(x^2 + y^2)$ solves Laplace's equation for all $(x, y) \in \mathbb{R}^2$ except at $x = y = 0$.

Review Question 1.5.8. For the two harmonic functions $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$, verify the validity of the mean-value properties (1.15) and (1.16) directly, i.e., without using Theorem 1.42.

Review Question 1.5.9. Find a solution to the simple transport equation $u_t = 4u_x$ for $t > 0$ and $x \in \mathbb{R}$ subject to the initial condition $u(0, x) = xe^{-x}$.

Review Question 1.5.10. Let $u(t, x) = -(t + \frac{1}{x})^{-1}$. Show that u is a solution to the differential equation $u_t + x^2 u_x = 0$. What initial condition does u satisfy? Show that for any nonzero real number $\rho \in \mathbb{R}$, the function $u(t, x)$ is constant along the planar curve $\{(t, x) : x = 1/(\rho - t)\}$.

Review Question 1.5.11. Show that $(x - t)^2$ is a solution to the wave equation with $x \in \mathbb{R}$. What are the initial position and initial velocity functions?

Review Question 1.5.12. For the wave equation $u_{tt} = c^2 u_{xx}$ on the one-dimensional domain $\Omega = (0, L)$ verify that for every integer $n \in \mathbb{N}$ the function

$$u(t, x) = \cos \frac{n\pi x}{L} \cos \frac{cn\pi t}{L}$$

is a solution. Find some homogeneous boundary conditions which are satisfied by these solutions.

Review Question 1.5.13. Classify the following partial differential equations as parabolic, hyperbolic, or elliptic. If the differential equation depends on some of the independent variables, find the regions of parabolicity, hyperbolicity, and ellipticity. In each case, u is a function of the two variables explicitly mentioned in the equation.

- | | |
|--------------------------------------|----------------------------------------|
| (a) $u_{xx} + u_{xy} + u_{yy} = 0$ | (c) $u_{xx} - 6u_{xy} + 9u_{yy} = u_y$ |
| (b) $u_{xx} + 2u_{xy} + 3u_{yy} = u$ | (d) $y u_{xx} + x u_{yy} = 0$ |

Review Question 1.5.14. Assume that L is a linear differential operator, and assume that a given function $y(t, x)$ solves the equation $Lu = 0$. Assume further that $z(t, x)$ solves the equation

$$Lu = f(t, x). \quad (1.69)$$

Then show that for every real number $\alpha \in \mathbb{R}$ the function $w(t, x) = z(t, x) + \alpha y(t, x)$ solves (1.69).

Review Question 1.5.15. In general, a function F on a vector space is said to be *linear* if for all u, v in the domain of F and all $c \in \mathbb{R}$ we have both

$$F(u + v) = F(u) + F(v) \quad \text{and} \quad F(cu) = cF(u).$$

Show that the following differential operators are linear by this definition.

- (a) Let F have as its domain all continuously differentiable functions $u(t, x)$ and define $F(u) = u_t - au_x$, where $a \in \mathbb{R}$. Show that F is linear.
- (b) Let F have as its domain all twice continuously differentiable functions $u(t, x)$. Define $F(u) = u_{xx} + u_{yy}$, and then show that F is linear.
- (c) Give an example of a function F with domain being the continuously differentiable functions $u(t, x)$ such that F is not linear.

Review Question 1.5.16. For each of the following equations, determine whether the function is linear, semilinear, quasilinear, or fully nonlinear.

- | | |
|-----------------------------------------|------------------------------------------|
| (a) $u_t + uu_{xx} + x^2 t u_{xxt} = 0$ | (c) $u_{tt} = c^2 u_{xx} u_t + 5t^2 + 1$ |
| (b) $u_{tt} - u_{xx}^2 = 3u + xt$ | (d) $u_{ttt} + t^2 u = e^{x^2 t} u_{xx}$ |

Review Question 1.5.17. Classify the following semilinear partial differential equations as parabolic, hyperbolic, or elliptic. If the differential equation depends on some of the independent variables, find the regions of parabolicity, hyperbolicity, and ellipticity. In each case, u is a function of the two variables explicitly mentioned in the equation.

- | | |
|--------------------------------------------|------------------------------------------------|
| (a) $u_{tt} = u_{xx} + u(1-u)$ | (c) $u_t + xu_{xx} + u_t u_x = 3x + t^2$ |
| (b) $u_{xx} + (1+x^2)u_{yy} + u_x u_y = 0$ | (d) $u_{xx} + 3u_{xy} + u_{yy} + u(1-u_y) = 0$ |

Review Question 1.5.18. Show that if an m -th order linear partial differential operator is elliptic, then m has to be an even number. In other words, there are no elliptic differential operators of odd order.

1.5.2 • Problems

Problem 1.5.1. Use the gradient method to find an outward unit normal vector at a point on the boundary of the following regions. Hint: In all cases, the boundary of the given region can be expressed as a level set of the function given when defining the set.

- (a) The interior of the ellipse $\{(x, y) \in \mathbb{R}^2 : x^2/25 + y^2/36 < 1\}$.
- (b) The interior of the ellipsoid $\{(x, y, z) \in \mathbb{R}^3 : x^2/4 + y^2/9 + z^2/81 < 1\}$.
- (c) The interior of the cylinder $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 < 16\}$.
- (d) The region $\{(x, y) \in \mathbb{R}^2 : y - 3x^2 > 0\}$. Hint: Verify that the gradient of the function $G(x, y) = y - 3x^2$ is an inward normal vector. Thus the negative of the gradient is an outward normal vector.

Problem 1.5.2. For each function F , domain Ω , and point on $\partial\Omega$, find $\partial F/\partial n$, where $\partial F/\partial n$ is defined in Definition 1.33.

- (a) Function $F(x, y) = (x^2 + y^2)$, domain $\Omega = \{(x, y) : x^2/4 + y^2/9 < 1\}$, at the boundary point $(x_0, y_0) = (2, 0)$.
- (b) Function $F(x, y) = \ln x + y^2$, domain $\Omega = \{(x, y) : 3y - x^2 < 17\}$, at the boundary point $(x_0, y_0) = (2, 7)$.
- (c) Function $F(x, y, z) = 5x^2 + y^2 + xyz$, domain $\Omega = \{(x, y, z) : x^2 + y^2 < 16\}$, at the boundary point $(x_0, y_0, z_0) = (2, 2\sqrt{3}, 3)$.

Problem 1.5.3. Let u denote a solution to the heat equation $u_t = k u_{xx}$ subject to homogeneous Dirichlet boundary conditions on $\Omega = (0, L)$, and define

$$E(t) = \int_0^L u(t, x)^2 dx \quad \text{for } t \geq 0.$$

Show that the function $E(t)$ is decreasing as a function of time t by showing that the derivative of this function satisfies $dE(t)/dt \leq 0$ for all $t > 0$. The function $E(t)$ is called the energy for the heat equation. Note that unlike for the wave equation, the energy function is not conserved along solutions but instead only decreasing.

Problem 1.5.4. This problem generalizes Problem 1.5.3 to higher dimensions. Let $\Omega \subset \mathbb{R}^d$ and assume that u satisfies the heat equation $u_t = k \Delta u$ on Ω subject to homogeneous Dirichlet boundary conditions. Define $E(t) = \int_{\Omega} u(t, x)^2 dx$. Show that $E(t)$ is a decreasing function of time t by verifying $dE(t)/dt \leq 0$ for all $t \geq 0$.

Problem 1.5.5. Assume that u satisfies Laplace's equation on a domain $\Omega \subset \mathbb{R}^3$ and that v is a continuously differentiable function such that $v(x) = 0$ for all $x \in \partial\Omega$. Use the divergence theorem to show that

$$\int_{\Omega} \nabla u(x) \cdot \nabla v(x) dx = 0.$$

Problem 1.5.6. Verify directly, i.e., without using the connection with complex analytic functions, that the polynomials u and v defined in (1.14) are harmonic. This implies that for $n \in \mathbb{N}$ an even integer, the real and imaginary parts of the complex analytic function $f(z) = z^n$ solve Laplace's equation.

Problem 1.5.7. Let $n \in \mathbb{N}$ be an odd integer, and consider the complex analytic function $f(z) = z^n$. Find explicit formulas for the functions $u, v : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined via real and imaginary parts through the formula

$$f(x + iy) = (x + iy)^n = u(x, y) + iv(x, y) \quad \text{for all } (x, y) \in \mathbb{R}^2.$$

Then verify directly that the polynomials u and v are harmonic.

Problem 1.5.8. Let $u : B_R(x_0, y_0) \rightarrow \mathbb{R}$ denote a continuous function defined on the open ball $B_R(x_0, y_0) \subset \mathbb{R}^2$ with radius R and center $(x_0, y_0) \in \mathbb{R}^2$. Show that then

$$\lim_{r \rightarrow 0} \left(\frac{1}{2\pi r} \int_{\partial B_r(x_0, y_0)} u \, ds \right) = u(x_0, y_0),$$

without using the statement of Theorem 1.42.

Problem 1.5.9. For the two functions $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$, find all their local and global maxima and minima on each of the following sets:

- (a) $\overline{\Omega} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 2\}$,
- (b) $\overline{\Omega} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 2\}$,
- (c) $\overline{\Omega} = \{(x, y) \in \mathbb{R}^2 : x^2 - 2x + y^2 - 2y + 1 \leq 0\}$,
- (d) $\overline{\Omega} = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x + 2y \leq 2\}$.

In each case, visualize the set $\overline{\Omega}$ together with the location of all extreme values.

Problem 1.5.10. Use the method of characteristics to find a solution for the generalized transport equation for the unknown $u(t, x)$ given by

$$u_t + 2xu_x = 0$$

for $x \in \mathbb{R}$ and $t > 0$, with initial condition $u(0, x) = e^{-x^2}$. Describe the asymptotic behavior of the solution for $t \rightarrow \infty$.

Problem 1.5.11. Use the method of characteristics to find a solution for the generalized transport equation for $u(t, x)$ given by

$$u_t + x^2u_x = 0$$

for $x > 0$ and $t > 0$, with initial condition $u(0, x) = \cos x$ and subject to $u(t, 0) = 1$. Describe the asymptotic behavior of the solution for $t \rightarrow \infty$. Does the problem have a solution for all $x \in \mathbb{R}$?

Problem 1.5.12. The goal of this problem is to use the divergence theorem to derive a higher-dimensional version of the transport equation. Consider a fluid flowing through a domain $\Omega \subset \mathbb{R}^d$, where $d = 1, 2, 3$ with velocity given by a prescribed vector $v(t, x)$, and let $u(t, x)$ be the concentration of dye in the fluid.

- (a) Write down an equation for the mass of dye in a region $D \subset \Omega$, and then determine the derivative dM/dt .

- (b) Since dye cannot be created or destroyed, the rate of change of mass in D is equal to the negative of the outflow on the boundary of D . At each point along the boundary, this outflow is given by the component of $u(t,x)v(t,x)$ in the outward normal direction. Use this to write down a formula for the net outflow of dye on the boundary of D .
- (c) Use the divergence theorem to write this as an integral on the domain D .
- (d) Equate the negative outflow given in (c) with the expression in (a). Use this to write down a higher-dimensional transport equation.

Problem 1.5.13. Let u be a solution of the wave equation $u_{tt} = c^2 u_{xx}$ on the domain $\Omega = (0, 1)$, subject to homogeneous Neumann boundary conditions.

- (a) Define

$$E(t) = \int_0^1 (u_t(t, x)^2 + c^2 u_x(t, x)^2) dx .$$

Show that $E(t)$ is constant, i.e., it does not depend on the time t . That is, show that the derivative satisfies $dE(t)/dt = 0$. Hint: The derivation is similar to the derivation shown in the case of the wave equation with homogeneous Dirichlet boundary conditions.)

- (b) Use (a) to show that the only differentiable function satisfying the wave equation $u_{tt} = c^2 u_{xx}$ such that $u(0, x) = 0$ and $u_t(0, x) = 0$ for all $x \in (0, 1)$, as well as $u_x(t, 0) = u_x(t, 1) = 0$ for all $t > 0$ is the trivial function $u(t, x) \equiv 0$. Hint: Find $E(0)$ for a function with these initial conditions.

Problem 1.5.14. Let u and v be given solutions to the wave equation $u_{tt} = c^2 u_{xx}$ subject to homogeneous Neumann boundary conditions and with the two initial conditions $u(0, x) = v(0, x) = f(x)$ and $u_t(0, x) = v_t(0, x) = g(x)$.

- (a) Show that $w(t, x) = u(t, x) - v(t, x)$ is also a solution to the wave equation subject to homogeneous Neumann boundary conditions. What are the initial conditions for w ?
- (b) Use part (a) of this problem combined with Problem 1.5.13(b) to show that a differentiable solution to the wave equation with homogeneous Neumann boundary conditions and initial conditions $f(x)$ and $g(x)$ must be unique.

Problem 1.5.15. Give a geometric description of the regions of the plane where the equation $u_{xx} + 2y u_{xy} + x(2-x) u_{yy} = u$ is parabolic, elliptic, and hyperbolic. Hint: the curve of points where the equation is parabolic divides the regions where the equation is elliptic and hyperbolic.

Problem 1.5.16. Show that if $X : F \rightarrow \mathbb{R}$ is a standard normally distributed random variable over a probability space $(F, \mathcal{F}, \mathbb{P})$, then σX is normally distributed with mean 0 and variance σ^2 . Can you create a random variable which is normally distributed with mean μ and variance σ^2 ?

Problem 1.5.17. Verify the solution formulas for the Ornstein-Uhlenbeck position and velocity processes $r(t, \omega)$ and $v(t, \omega)$, respectively, which were given in Example 1.90.

Problem 1.5.18. Using Lemma 1.87, show that the Ornstein-Uhlenbeck velocity process $v(t, \omega)$ has the variance given in Example 1.90. Can you use this result also to find the variance of the position process r ?

Problem 1.5.19. Using the properties of the Wiener process, describe the asymptotic behavior of the solutions (1.64) of the general scalar linear stochastic differential equation (1.63). How does this behavior differ from the solution behavior of the underlying deterministic equation which corresponds to $b = 0$?

Problem 1.5.20. Using Ito's formula, verify in detail that the stochastic process given in (1.66) solves the stochastic differential equation (1.65). Verify also that for every $k \in \mathbb{Z}$ the constant process $x(t, \omega) = \pi/2 + k\pi$ is indeed a solution of (1.65).

Problem 1.5.21. Using Ito's formula, verify in detail that the stochastic process given in (1.68) solves the stochastic differential equation (1.67). Using the properties of the Wiener process, show that for every initial condition $x_0 \notin [-1, 1]$ solutions blow up in finite time with probability one. Can you describe the typical solution behavior for initial conditions $x_0 \in (-1, 1)$? What can you say for $x_0 = \pm 1$?

Chapter 2

Separation of Variables

In this chapter we turn our attention to analytical methods for finding explicit solutions for certain partial differential equations. Though these methods only apply to a restricted set of linear problems, the theory underlying these methods is the basis for any advanced study of nonlinear problems. In addition, the methods also lead to the development of a large class of numerical methods for solving partial differential equations, as we discuss in subsequent chapters.

The analytical methods presented in this chapter are centered around two main ideas. The first of these is called *separation of variables*, which is aimed at reducing the study of partial differential equations on higher-dimensional domains to the case of boundary value problems on one-dimensional domains. The second idea is concerned with *series representations of solutions* and will lead to a study of Fourier-type series and complete orthonormal sets in function spaces. Combined, both ideas will be used throughout the remainder of this book in a variety of settings.

Our discussion of these topics starts out by considering generalized Fourier series expansions. There is an important connection between Fourier series and partial differential equations. In fact, Jean Baptiste Joseph Fourier discovered the eponymous series in the course of writing an 1807 memoir on the theory of heat propagation, and generalizations of Fourier series lie at the heart of many analytical and numerical techniques presented in this book. We then address Sturm-Liouville boundary value problems in one dimension. We combine these concepts to describe the method of separation of variables using the classical partial differential equations introduced in the last chapter. Particular emphasis is placed on studying a variety of different types of boundary conditions, as well as on demonstrating the usefulness of suitable coordinate transformations.

2.1 • Generalized Fourier Series Expansions

As we mentioned above, one of the main analytical techniques of the present chapter will enable us in many situations to obtain explicit solution formulas for certain classes of partial differential equations on higher-dimensional domains. As will become clear very soon, many of these solution formulas are most easily presented as convergent series which are based on suitable “basis functions.” In this section, we describe the functional-analytic framework for these series expansions, which will lead to the concepts of generalized Fourier series and complete orthonormal sets.

2.1.1 • Orthogonal Functions

Our discussion of initial value problems for ordinary differential equations in the previous chapter mainly centered on the question of existence and uniqueness of solutions — and we have seen that under fairly mild assumptions these can always be achieved. The situation turns out to be completely different in the case of boundary value problems. Even under reasonable assumptions, boundary value problems can have infinitely many solutions. In such a case, it is of utmost importance to be able to represent solutions in a compact and nonredundant way. To make this point clearer, consider the case of a linear equation of the form

$$Ax = 0, \quad \text{for some matrix } A \in \mathbb{R}^{m \times d}.$$

The set of all vectors $x \in \mathbb{R}^d$ which satisfy this system forms a linear subspace of the vector space \mathbb{R}^d , and this subspace is called the nullspace of the matrix. If the nullspace contains more than just the trivial solution $x = 0$, then it automatically contains infinitely many solutions. Nevertheless, one can describe the nullspace in a compact way using linear algebra. Since the nullspace is a vector space, it contains a basis. In other words, one can determine a finite number of vectors $v_1, \dots, v_k \in \mathbb{R}^d$ with $k \leq d$ which are linearly independent and such that every solution of the linear system $Ax = 0$ can be written as a linear combination

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_k v_k$$

for real numbers $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$. Thus, even though the actual solution set is infinite, it can be conveniently represented by a finite set of vectors.

In the context of boundary value problems we will be interested in finding solutions defined on a compact interval I in \mathbb{R} . Does it make sense to talk about a basis representation of a solution set in this situation? To see that it does, we refer to the linear algebra concept of a vector space. One can easily see that the collection of all functions $f : I \rightarrow \mathbb{R}$ forms a vector space if we define the addition of two functions f and g pointwise via

$$(f + g)(x) = f(x) + g(x) \quad \text{for all } x \in I, \tag{2.1}$$

and the multiplication by a scalar $\alpha \in \mathbb{R}$ via

$$(\alpha f)(x) = \alpha f(x) \quad \text{for all } x \in I. \tag{2.2}$$

Thus, the natural vector space operations in \mathbb{R}^d have natural counterparts in the space of functions. Similarly, one can talk of linear combinations of functions, as well as linear independence, and these concepts naturally lead to the concept of a basis of a subspace.

Rather than presenting all of the above in detail, we will restrict our attention to a special kind of basis from linear algebra, namely to the special case of an orthogonal basis. For this, recall that the *norm* of a vector u is defined by

$$\|u\| = \sqrt{u \cdot u} = \sqrt{\sum_{k=1}^d u_k^2}. \tag{2.3}$$

The symbol “ \cdot ” denotes the dot product of the vector with itself. Two vectors u and v are called *orthogonal* or *perpendicular*, if their dot product vanishes, i.e., if we have $u \cdot v = 0$.

Any collection of nonzero pairwise orthogonal vectors is automatically linearly independent, and can therefore serve as a candidate for a basis. In fact, even more is true: Any subspace of the vector space \mathbb{R}^d has bases which consist of pairwise orthogonal vectors.

It turns out to be very useful to extend the notion of orthogonality to functions. This will allow us to construct basis representations which are based on orthogonality. We do so by extending the idea of the dot product on \mathbb{R}^d to the space of functions. The function analog of the dot product is called the *inner product*, defined as follows:

Definition 2.1 ($L^2(I)$, inner product, and orthogonal functions). Let $I \subset \mathbb{R}$ denote an arbitrary interval. Then we define the vector space $L^2(I)$ as

$$L^2(I) = \left\{ f : I \rightarrow \mathbb{R} : f \text{ is integrable and } \int_I |f(x)|^2 dx < \infty \right\}.$$

In this vector space, the sum of two functions and the scalar multiplication are defined pointwise as in (2.1) and (2.2), respectively. For two functions $f, g \in L^2(I)$, we define the inner product of f and g by

$$\langle f, g \rangle = \int_I f(x)g(x) dx.$$

For every function $f \in L^2(I)$, its L^2 -norm is defined as

$$\|f\|_{L^2(I)} = \left(\int_I |f(x)|^2 dx \right)^{1/2} = \sqrt{\langle f, f \rangle}.$$

Finally, two functions f and g in $L^2(I)$ are called orthogonal, if $\langle f, g \rangle = 0$.

The vector space $L^2(I)$ defined here is named after the mathematician Henri Lebesgue. Notice that in the above definition we do not make any assumption on the boundedness or closedness of the interval I . In particular, the choices $I = \mathbb{R}$ and $I = (0, 1)$ are certainly admissible. In such cases, the notion of integrability is to be understood in the sense of improper integrals.

The next two examples illustrate these concepts. They show that both the functions in the vector space $L^2(I)$ and the orthogonality of such functions is quite dependent on the interval I .

Example 2.2 (The vector space $L^2(I)$). Consider the two functions

$$f(x) = 1 \quad \text{and} \quad g(x) = x.$$

For the interval $I_1 = [-1, 1]$, both f and g are in the vector space $L^2(I_1)$, since

$$\|f\|_{L^2(I_1)}^2 = \int_{I_1} |f(x)|^2 dx = \int_{-1}^1 1^2 dx = 2 < \infty,$$

$$\|g\|_{L^2(I_1)}^2 = \int_{I_1} |g(x)|^2 dx = \int_{-1}^1 x^2 dx = 2/3 < \infty.$$

Furthermore, f and g are orthogonal in $L^2(I_1)$, since

$$\langle f, g \rangle_{L^2(I_1)} = \int_{-1}^1 x dx = 0.$$

Similarly, the same functions f and g are contained in the vector space $L^2(I_2)$ for the interval $I_2 = [0, 1]$, because $\|f\|_{L^2(I_2)}^2 = 1 < \infty$, and $\|g\|_{L^2(I_2)}^2 = 1/3 < \infty$. However, f and g are not orthogonal in $L^2(I_2)$, since $\langle f, g \rangle_{L^2(I_2)} = 1/2 \neq 0$.

As an even more extreme contrast, for the interval $I_3 = \mathbb{R}$ neither f nor g are contained in the vector space $L^2(I_3)$, since

$$\begin{aligned}\int_{I_3} |f(x)|^2 dx &= \int_{-\infty}^{\infty} 1^2 dx = \infty, \\ \int_{I_3} |g(x)|^2 dx &= \int_{-\infty}^{\infty} x^2 dx = \infty.\end{aligned}$$

In other words, even for the same functional formulas, membership in a space $L^2(I)$ depends crucially on the underlying domain I . ■

Example 2.3 (Functions in $L^2(I)$). If one chooses $I = (1, \infty)$ and considers the functions $f(x) = e^{-x^2}$ and $g(x) = 1/\sqrt{x}$, then we have $f \in L^2(1, \infty)$, but $g \notin L^2(1, \infty)$. Similarly, if one chooses $I = (0, 1)$ and considers the functions $f(x) = 1/\sqrt[4]{x}$ and $g(x) = 1/\sqrt{x}$, then we have $f \in L^2(0, 1)$, but $g \notin L^2(0, 1)$. ■

In Definition 2.1 it is stated that the collection of functions given by $L^2(I)$, together with pointwise addition and scalar multiplication is actually a vector space. This statement means that for arbitrary $f, g \in L^2(I)$ and $\alpha \in \mathbb{R}$ we have both

$$\alpha f \in L^2(I) \quad \text{and} \quad f + g \in L^2(I).$$

While the former inclusion follows easily from

$$\int_I |\alpha f(x)|^2 dx = \alpha^2 \int_I |f(x)|^2 dx < \infty,$$

the latter inclusion follows from the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ via

$$\int_I |f(x) + g(x)|^2 dx \leq 2 \int_I |f(x)|^2 dx + 2 \int_I |g(x)|^2 dx < \infty.$$

Finally, notice that for any two functions $f, g \in L^2(I)$ the inequality $2ab \leq a^2 + b^2$ implies that their product $f g$ is integrable, since

$$\int_I |f(x)g(x)| dx \leq \frac{1}{2} \int_I |f(x)|^2 dx + \frac{1}{2} \int_I |g(x)|^2 dx < \infty.$$

We now present several important examples of families of orthogonal functions. All of these will be considered extensively later on.

Example 2.4 (The sine family). For our first example, let $L > 0$ be an arbitrary length, and consider the interval $I = [0, L]$. On this interval, consider the sequence of functions $g_m : I \rightarrow \mathbb{R}$ defined by

$$g_m(x) = \sin \frac{m\pi x}{L} \quad \text{for } m \in \mathbb{N}.$$

This sequence has the property that for any pair of positive integers $j \neq k$ the functions g_j and g_k are orthogonal, i.e., we have

$$\langle g_j, g_k \rangle = 0 \quad \text{for all positive integers } j \neq k.$$

To verify this, recall the trigonometric identity

$$\sin a \sin b = \frac{\cos(a-b) - \cos(a+b)}{2},$$

which implies for $j \neq k$ the identity

$$\begin{aligned} \langle g_j, g_k \rangle &= \int_0^L \sin \frac{j\pi x}{L} \sin \frac{k\pi x}{L} dx \\ &= \frac{1}{2} \int_0^L \left(\cos \frac{(j-k)\pi x}{L} - \cos \frac{(j+k)\pi x}{L} \right) dx \\ &= \frac{1}{2} \left(\frac{L}{\pi(j-k)} \sin \frac{(j-k)\pi x}{L} - \frac{L}{\pi(j+k)} \sin \frac{(j+k)\pi x}{L} \right) \Big|_0^L = 0. \end{aligned}$$

In addition, we have

$$\|g_k\|_{L^2(I)} = \sqrt{\langle g_k, g_k \rangle} = \sqrt{\frac{L}{2}},$$

which follows from

$$\langle g_k, g_k \rangle = \int_0^L \left(\sin \frac{k\pi x}{L} \right)^2 dx = \frac{1}{2} \int_0^L \left(1 - \cos \frac{2k\pi x}{L} \right) dx = \frac{L}{2},$$

where we used the trigonometric identity $2\sin^2 a = 1 - \cos(2a)$. ■

Example 2.5 (The cosine family). In this example we consider the interval $I = [0, L]$ of length $L > 0$, and the set of functions

$$b_m(x) = \cos \frac{m\pi x}{L} \quad \text{for nonnegative integers } m \in \mathbb{N}_0.$$

Note that $b_0(x) = 1$. Using a similar proof as used above in Example 2.4 one can easily verify that for any nonnegative integers $j \neq k$ the functions b_j and b_k are orthogonal. Furthermore, we have the identities

$$\|b_k\|_{L^2(I)} = \sqrt{\frac{L}{2}} \quad \text{for arbitrary } k \in \mathbb{N}, \quad \text{as well as} \quad \|b_0\|_{L^2(I)} = \sqrt{L}.$$

Notice the difference in norms for $k = 0$ and $k \geq 1$. ■

Example 2.6 (The Fourier family). Let $I = [0, L]$ for $L > 0$, and consider a family consisting of both the functions

$$G_m(x) = \sin \frac{2\pi mx}{L} \quad \text{for } m \in \mathbb{N},$$

and the functions

$$H_m(x) = \cos \frac{2\pi mx}{L} \quad \text{for } m \in \mathbb{N}_0.$$

We call this set of functions the Fourier family, since it is the set of functions used for a classical Fourier series (as discussed in a subsequent section of this chapter). As in the

previous two examples one can show that all of these functions are pairwise orthogonal, i.e., for all $j, k \in \mathbb{N}$ and all $m, n \in \mathbb{N}_0$ we have

$$\langle G_j, G_k \rangle = \langle G_j, H_n \rangle = \langle H_m, H_n \rangle = 0 \quad \text{for } j \neq k, j \neq n, \text{ and } m \neq n.$$

Furthermore, the norms of these functions satisfy

$$\|G_k\|_{L^2(I)} = \|H_k\|_{L^2(I)} = \sqrt{\frac{L}{2}} \quad \text{for } k \in \mathbb{N}, \quad \text{and} \quad \|H_0\|_{L^2(I)} = \sqrt{L}.$$

At first glance, it seems that this example is just combining the functions from the previous two examples. However, this is not quite true. One can easily see that

$$G_k = g_{2k} \quad \text{and} \quad H_m = h_{2m} \quad \text{for all } k \in \mathbb{N} \quad \text{and} \quad m \in \mathbb{N}_0.$$

In other words, the families G_k and H_m only use “half” of the functions in each of the previous examples. It turns out that this is in fact necessary to achieve pairwise orthogonality, since for example one has $\langle g_1, h_0 \rangle = 2L/\pi \neq 0$. ■

It was already mentioned earlier that in the case of Euclidean space \mathbb{R}^d , the notion of orthogonality is frequently used to construct vector space bases with nice properties. One of the crucial properties in this context is the fact that sets of pairwise orthogonal vectors are automatically linearly independent. This remains true for the case of orthogonal functions, as the following lemma shows.

Lemma 2.7 (Orthogonality and linear independence). *Let $I \subset \mathbb{R}$ denote an arbitrary interval and consider the space $L^2(I)$ introduced in Definition 2.1. Furthermore, let $B \subset L^2(I)$ denote any set of functions with the following properties:*

- (a) *For any $f \in B$ we have $\|f\|_{L^2(I)} > 0$.*
- (b) *For arbitrary $f, g \in B$ with $f \neq g$ we have $\langle f, g \rangle = 0$.*

Such a set is called an orthogonal set. Then the set B is linearly independent. In other words, if $f_1, \dots, f_m \in B$ are arbitrary and if coefficients $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ are chosen in such a way that

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) + \dots + \alpha_m f_m(x) = 0 \quad \text{for all } x \in I,$$

then one has to have $\alpha_1 = \alpha_2 = \dots = \alpha_m = 0$.

Proof. Assume B is a set satisfying the two properties in the lemma. Suppose further that the constants $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ are chosen in such a way that

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) + \dots + \alpha_m f_m(x) = 0 \quad \text{for all } x \in I.$$

Now fix an integer $k \in \{1, \dots, m\}$ and take the inner product of both sides with the function f_k . This yields

$$0 = \langle 0, f_k \rangle = \left\langle \sum_{i=1}^m \alpha_i f_i, f_k \right\rangle = \sum_{i=1}^m \alpha_i \langle f_i, f_k \rangle = \alpha_k \|f_k\|_{L^2(I)}^2,$$

since $\langle f_i, f_k \rangle = 0$ whenever we have $i \neq k$. Due to $\|f_k\|_{L^2(I)} > 0$ the last identity immediately implies $\alpha_k = 0$, and since $k \in \{1, \dots, m\}$ was arbitrary the result follows. □

With the above result, we have identified the first essential ingredient for constructing bases in vector spaces — linear independence. The second ingredient will be the subject of the next section.

2.1.2 • Complete Orthonormal Sets

We now turn our attention to the concept of bases. For this, let us first recall some basic facts for Euclidean space \mathbb{R}^d . A set B is called a basis in the classical sense if B is linearly independent and every element of the vector space can be written as a finite linear combination of elements in B . A key result of linear algebra is that since \mathbb{R}^d is d -dimensional, any linearly independent set of d vectors is automatically a basis of \mathbb{R}^d . Since orthogonality implies linear independence, this immediately shows that any set of d pairwise orthogonal nonzero vectors is automatically a basis of \mathbb{R}^d . For example, in the case $d = 3$ the three vectors

$$v_1 = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad v_3 = \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix}$$

form a basis of \mathbb{R}^3 , since one can easily verify $v_1 \cdot v_2 = v_1 \cdot v_3 = v_2 \cdot v_3 = 0$.

In the case of the function space $L^2(I)$ things are a bit more complicated. First of all, as we have seen in the previous three examples, this vector space is no longer finite-dimensional, but rather *infinite-dimensional* — since we have found infinite sets which are orthogonal and therefore linearly independent. In addition, it turns out that the notion of basis that is used in linear algebra is not adequate for the study of infinite-dimensional spaces. To see this, note that none of the three families of functions introduced in the previous examples can be a basis of the space $L^2(I)$ with $I = [0, L]$ in the classical sense, since every finite linear combination of these functions is infinitely many times differentiable, whereas the space $L^2(I)$ contains nondifferentiable functions. (See Review Question 2.7.4.) This shows that any classical basis for $L^2(I)$ would necessarily have to include nondifferentiable functions!

In order to resolve the above issues, one has to abandon the idea of exact representation of an element of $L^2(I)$ by finite linear combinations of basis functions. This leads to the following definition generalizing the classical concept of a basis.

Definition 2.8 (Complete orthogonal sets). Let $I \subset \mathbb{R}$ denote an arbitrary interval, and consider the vector space $L^2(I)$ introduced in Definition 2.1. We call a subset B of $L^2(I)$ a complete orthogonal set, if it satisfies the following three properties:

- (a) For every $\varphi \in B$ we have $\|\varphi\|_{L^2(I)} > 0$.
- (b) For every pair $\varphi, \psi \in B$ with $\varphi \neq \psi$ we have $\langle \varphi, \psi \rangle = 0$.
- (c) The set of all finite linear combinations of functions in B are dense in $L^2(I)$ with respect to the $L^2(I)$ -norm. In other words, for any $f \in L^2(I)$ and every $\varepsilon > 0$, there exist functions $\varphi_1, \dots, \varphi_m \in B$ and real coefficients $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ such that

$$\left\| f - \sum_{i=1}^m \alpha_i \varphi_i \right\|_{L^2(I)} < \varepsilon.$$

If in addition we have $\|\varphi\|_{L^2(I)} = 1$ for all $\varphi \in B$, then B is called a complete orthonormal set.

If a function is continuous, then the first property in the definition above implies that it is not identically zero, see Problem 2.7.1. A set with the second property is an

orthogonal set. We now focus on the third part of the definition. A set with this third property is called a *complete set*. It states that if we measure the distance between two functions $f, g \in L^2(I)$ by the norm $\|f - g\|_{L^2(I)}$ of their difference, then any given function $f \in L^2(I)$ can be approximated arbitrarily well by linear combinations of *some* functions in B . In light of our discussion before the above definition, one would certainly expect that in order to achieve better and better approximations, one has to increase the number m of functions from B used in the linear combination.

It is clear that one can always turn a complete orthogonal set into a complete orthonormal set by just normalizing the functions φ_k . However, to show that a given set of functions does in fact form a complete orthogonal set is usually much harder. We now give a few examples of complete orthonormal sets.

Example 2.9 (The sine family). As we have seen in Example 2.4, if we consider the interval $I = [0, L]$ for some $L > 0$, then the set $\{g_m : m \in \mathbb{N}\}$ with $g_m(x) = \sin(m\pi x/L)$ consists of pairwise orthogonal functions. In addition, this set is a complete orthogonal set in $L^2(I)$. (See Example 2.39 for the proof of completeness.) As a consequence, the functions

$$\varphi_m(x) = \sqrt{\frac{2}{L}} \sin \frac{m\pi x}{L} \quad \text{for } m \in \mathbb{N}$$

form a complete orthonormal set in $L^2(I)$. ■

Example 2.10 (The cosine family). The functions in Example 2.5 constitute a complete orthogonal set in $L^2(I)$ for $I = [0, L]$. (See Example 2.40 for the proof of completeness.) Therefore the functions

$$\varphi_0(x) = \sqrt{\frac{1}{L}}, \quad \text{and} \quad \varphi_m(x) = \sqrt{\frac{2}{L}} \cos \frac{m\pi x}{L} \quad \text{for } m \in \mathbb{N}$$

form a complete orthonormal set in $L^2(I)$. ■

Example 2.11 (The Fourier family). The family in Example 2.6 is a complete orthogonal set, which gives rise to the following complete orthonormal set $\{\varphi_k : k \in \mathbb{N}_0\}$ in $L^2(I)$, for $I = [0, L]$ for $L > 0$, where $\varphi_0(x) = \sqrt{1/L}$,

$$\varphi_{2k-1}(x) = \sqrt{\frac{2}{L}} \sin \frac{2k\pi x}{L}, \quad \text{and} \quad \varphi_{2k}(x) = \sqrt{\frac{2}{L}} \cos \frac{2k\pi x}{L}$$

for all $k \in \mathbb{N}$. Looking ahead, this family has a special place in the study of generalized Fourier series introduced in Theorem 2.14. Namely, the generalized Fourier series with respect to this complete orthonormal set are known as *classical Fourier series*. ■

Example 2.12 (The family of monomials). So far, all complete orthonormal sets have been based on trigonometric functions, but there are other types. Consider for example the monomials given by $p_k(x) = x^k$, for $k \geq 0$. The Stone-Weierstraß theorem of mathematical analysis shows that the set $\{p_k : k \geq 0\}$ is a complete set in $L^2([-1, 1])$. However, this set is not orthogonal. (See Review Question 2.7.5.) Example 2.44 demonstrates that it is possible to use a method analogous to the Gram-Schmidt orthogonalization method of linear algebra to find a complete orthonormal set of polynomials called the Legendre polynomials. ■

Best approximations

The definition of completeness says it is possible to approximate an $L^2(I)$ -function f to arbitrary accuracy with a linear combination of functions φ_i from a complete orthogonal set. Let us now turn the completeness statement around and ask the following question: For fixed $f \in L^2(I)$ and fixed $\varphi_1, \dots, \varphi_m \in B$, what is the best-possible approximation of f by a linear combination of $\varphi_1, \dots, \varphi_m$? In other words, for which choice of the unknown coefficients $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ is the expression

$$\rho(\alpha_1, \dots, \alpha_m) = \left\| f - \sum_{i=1}^m \alpha_i \varphi_i \right\|_{L^2(I)}$$

minimal? In order to answer this question, note that by the definition of the inner product we obtain the validity of the following linearity properties

$$\langle \beta_1 g_1 + \beta_2 g_2, h \rangle = \langle h, \beta_1 g_1 + \beta_2 g_2 \rangle = \beta_1 \langle g_1, h \rangle + \beta_2 \langle g_2, h \rangle$$

for arbitrary $g_1, g_2, h \in L^2(I)$ and $\beta_1, \beta_2 \in \mathbb{R}$. With this identity one can then rewrite the square of ρ . This gives, in combination with the fact that $\langle \varphi_i, \varphi_j \rangle = 0$ for $i \neq j$, the identity

$$\begin{aligned} \rho(\alpha_1, \dots, \alpha_m)^2 &= \left\| f - \sum_{i=1}^m \alpha_i \varphi_i \right\|_{L^2(I)}^2 = \left\langle f - \sum_{i=1}^m \alpha_i \varphi_i, f - \sum_{j=1}^m \alpha_j \varphi_j \right\rangle \\ &= \langle f, f \rangle - \sum_{j=1}^m \alpha_j \langle f, \varphi_j \rangle - \sum_{i=1}^m \alpha_i \langle \varphi_i, f \rangle + \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j \langle \varphi_i, \varphi_j \rangle \\ &= \|f\|_{L^2(I)}^2 - 2 \sum_{i=1}^m \alpha_i \langle f, \varphi_i \rangle + \sum_{i=1}^m \alpha_i^2 \|\varphi_i\|_{L^2(I)}^2 \\ &= \|f\|_{L^2(I)}^2 - \sum_{i=1}^m \frac{\langle f, \varphi_i \rangle^2}{\|\varphi_i\|_{L^2(I)}^2} + \sum_{i=1}^m \left(\alpha_i \|\varphi_i\|_{L^2(I)} - \frac{\langle f, \varphi_i \rangle}{\|\varphi_i\|_{L^2(I)}} \right)^2, \end{aligned}$$

where the last equality is obtained by completing the square. If we now define

$$\gamma_i = \frac{\langle f, \varphi_i \rangle}{\|\varphi_i\|_{L^2(I)}^2} \quad \text{for all } i = 1, \dots, m,$$

then the above identity can be rewritten as

$$\rho(\alpha_1, \dots, \alpha_m)^2 = \|f\|_{L^2(I)}^2 - \sum_{i=1}^m \gamma_i^2 \|\varphi_i\|_{L^2(I)}^2 + \sum_{i=1}^m (\alpha_i - \gamma_i)^2 \|\varphi_i\|_{L^2(I)}^2.$$

From this we can now easily read off the minimal value for ρ . In particular, notice that the first two terms on the right-hand side do not depend on the coefficients α_i , and that the last term on the right-hand side is always nonnegative. Thus, in order to achieve the best-possible approximation, one has to choose the α_i in such a way that the third term vanishes, and this happens exactly for $\alpha_i = \gamma_i$. This implies the following result.

Lemma 2.13 (Best-possible approximation). *Let $I \subset \mathbb{R}$ denote an arbitrary interval, and consider the vector space $L^2(I)$ introduced in Definition 2.1. Furthermore, let $f \in L^2(I)$ be*

arbitrary, and consider functions $\varphi_1, \dots, \varphi_m \in L^2(I)$ with $\|\varphi_i\|_{L^2(I)} > 0$ for all $i = 1, \dots, m$, as well as

$$\langle \varphi_i, \varphi_j \rangle = 0 \quad \text{for all } i, j = 1, \dots, m \quad \text{with } i \neq j.$$

Then the expression

$$\rho(\alpha_1, \dots, \alpha_m) = \left\| f - \sum_{i=1}^m \alpha_i \varphi_i \right\|_{L^2(I)} \quad \text{where } \alpha_1, \dots, \alpha_m \in \mathbb{R}$$

achieves its minimal value if and only if we choose $\alpha_i = \gamma_i$, where

$$\gamma_i = \frac{\langle f, \varphi_i \rangle}{\|\varphi_i\|_{L^2(I)}^2} \quad \text{for all } i = 1, \dots, m. \quad (2.4)$$

The resulting minimal value is then given by

$$\min_{\alpha_1, \dots, \alpha_m \in \mathbb{R}} \rho(\alpha_1, \dots, \alpha_m) = \sqrt{\|f\|_{L^2(I)}^2 - \sum_{i=1}^m \gamma_i^2 \|\varphi_i\|_{L^2(I)}^2}. \quad (2.5)$$

Furthermore, the inequality

$$\sum_{i=1}^m \gamma_i^2 \|\varphi_i\|_{L^2(I)}^2 \leq \|f\|_{L^2(I)}^2 \quad (2.6)$$

holds, and it is called Bessel's inequality.

After these preparations, we can now return to our discussion of complete orthogonal sets. Assume that we are given a complete orthogonal set in $L^2(I)$ which is countable, i.e., assume that we have

$$B = \{\varphi_k \in L^2(I) : k \in \mathbb{N}\}.$$

Now let $f \in L^2(I)$ be arbitrary. Then for every $m \in \mathbb{N}$ we can define the approximation

$$f_m = \sum_{k=1}^m \gamma_k \varphi_k \quad \text{with } \gamma_k \text{ as in (2.4).}$$

Do these approximations improve if we increase m ? According to Lemma 2.13 we have

$$\|f - f_m\|_{L^2(I)} = \sqrt{\|f\|_{L^2(I)}^2 - \sum_{i=1}^m \gamma_i^2 \|\varphi_i\|_{L^2(I)}^2}$$

for all $m \in \mathbb{N}$, and the right-hand side of this identity clearly decreases with increasing m . In fact, our definition of completeness guarantees that the right-hand side converges to zero as $m \rightarrow \infty$. This fact, together with a few other important consequences, is collected in the following theorem which we state without proof.

Theorem 2.14 (Generalized Fourier series). *Let $I \subset \mathbb{R}$ denote an interval, and consider the vector space $L^2(I)$ introduced in Definition 2.1. Furthermore, let $B \subset L^2(I)$ be any complete orthogonal set in the sense of Definition 2.8. Then the following hold.*

(a) The set B is countably infinite, i.e., there exist functions φ_k for $k \in \mathbb{N}$ such that

$$B = \{\varphi_k \in L^2(I) : k \in \mathbb{N}\}.$$

(b) Let $f \in L^2(I)$ be arbitrary and define

$$\gamma_k = \frac{\langle f, \varphi_k \rangle}{\|\varphi_k\|_{L^2(I)}^2} \quad \text{for all } k \in \mathbb{N}. \quad (2.7)$$

Then we have

$$\lim_{m \rightarrow \infty} \left\| f - \sum_{k=1}^m \gamma_k \varphi_k \right\|_{L^2(I)} = 0. \quad (2.8)$$

(c) In the situation of (b), one also has the identity

$$\|f\|_{L^2(I)} = \sqrt{\sum_{k=1}^{\infty} \gamma_k^2 \|\varphi_k\|_{L^2(I)}^2}. \quad (2.9)$$

In the following, we will write

$$f \stackrel{L^2}{=} \sum_{k=1}^{\infty} \gamma_k \varphi_k$$

to indicate that the convergence is in the $L^2(I)$ -sense, and call the right-hand side the generalized Fourier series of f with respect to B .

The results of Theorem 2.14 take a particularly nice form if B is a complete orthonormal set, i.e., if the functions in B are all normalized to unit length. In this case, the generalized Fourier series of a function $f \in L^2(I)$ is given by

$$f \stackrel{L^2}{=} \sum_{k=1}^{\infty} \langle f, \varphi_k \rangle \varphi_k, \quad \text{where} \quad \langle f, \varphi_k \rangle = \int_I f(x) \varphi_k(x) dx,$$

and we have

$$\|f\|_{L^2(I)} = \sqrt{\sum_{k=1}^{\infty} \langle f, \varphi_k \rangle^2}.$$

We illustrate this approximation method with some examples. In each of our examples, we start with a complete orthogonal set $\{\varphi_k\}$ and a function $f(x)$, and we compute the best approximation coefficients γ_k defined in (2.7). We compare the function f to the finite sum

$$f_N(x) = \sum_{k=1}^N \gamma_k \varphi_k, \quad (2.10)$$

showing that in many cases this sum is an excellent approximation of f . Further details regarding this finite sum approximation are discussed on page 123.

Example 2.15 (Approximating functions with the sine family). For our first example consider the interval $I = [0, 1]$ and the complete orthonormal set

$$\varphi_k(x) = \sqrt{2} \sin(k\pi x) \quad \text{for } k \in \mathbb{N}$$

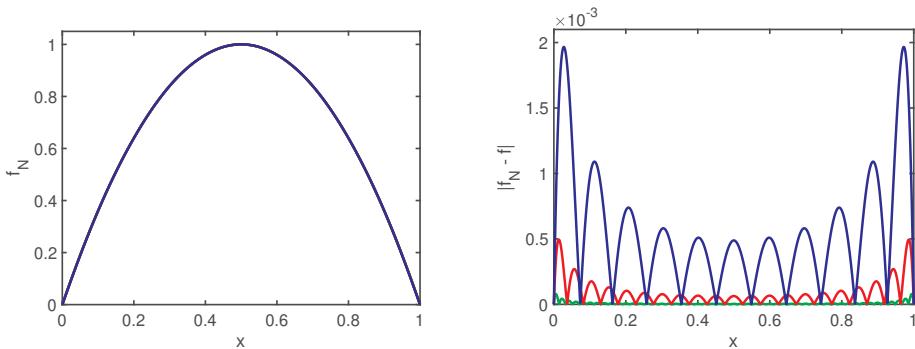


Figure 2.1. Generalized Fourier series approximations using the complete orthonormal set from Example 2.9 with $L = 1$ to approximate the function $f(x) = 4x(1-x)$ as shown in Example 2.15. The image in the left column contains the approximating sums $\sum_{k=1}^N \gamma_k \varphi_k$ for $N = 10, 20, 50$ in blue, red, and green, respectively. The figure in the right column shows the corresponding errors $|f_N(x) - f(x)|$. Notice the difference in the vertical scales of the right figures.

from Example 2.9. Furthermore, consider the function f

$$f(x) = 4x(1-x) \quad \text{for all } x \in [0, 1].$$

Then the coefficients γ_k of the generalized Fourier series of f with respect to $\{\varphi_k : k \in \mathbb{N}\}$ are found using (2.7), and they are given by

$$\begin{aligned} \gamma_k &= \frac{\langle 4x(1-x), \varphi_k(x) \rangle}{\|\varphi_k\|_{L^2(0,1)}^2} = \frac{\int_0^1 4x(1-x) \sqrt{2} \sin(k\pi x) dx}{1^2} \\ &= \frac{8\sqrt{2}(1-(-1)^k)}{k^3 \pi^3} = \begin{cases} 0 & , \quad \text{for even } k , \\ \frac{16\sqrt{2}}{k^3 \pi^3} & , \quad \text{for odd } k . \end{cases} \end{aligned}$$

The resulting approximations f_N , as defined in (2.10) on page 119, are shown in the left image of Figure 2.1 for the cases $N = 10, 20, 50$, and the right image depicts the corresponding graphs of $|f_N(x) - f(x)|$. The approximations of f through the partial Fourier sums is fantastic. Namely, we can see in the image that for every value of x in the interval $[0, 1]$ the inequalities

$$|f_{10}(x) - f(x)| \leq 2 \cdot 10^{-3}, \quad |f_{20}(x) - f(x)| \leq 0.5 \cdot 10^{-3}, \quad \text{and} \quad |f_{50}(x) - f(x)| \leq 0.1 \cdot 10^{-3}$$

hold. One can speculate from these observations that for each N , there is a corresponding maximum error E_N such that $|f_N(x) - f(x)| \leq E_N$ for all $x \in [0, 1]$, and that this maximal error satisfies $E_N \rightarrow 0$ as $N \rightarrow \infty$. That is, as N increases the functions f_N appear to converge uniformly to f on all of $[0, 1]$. We will see soon that this example does indeed have uniform convergence. However, Example 2.17 below will show that the convergence of solutions is not always so nice.

While we have computed these coefficients in closed form, doing so is not always easy or even possible. If one is only interested in the comparison between $f_N(x)$ and $f(x)$, it suffices to compute γ_k numerically. This is straightforward using the Matlab `integral` command. For example, the following line of code computes γ_3 :

```
f = @(x) sqrt(2).*sin(3.*pi.*x).*4.*x.* (1-x)
integral(f,0,1)
```

where $@(x)$ indicates the variable to be integrated, and the symbol $.*$ denotes multiplication. ■

Example 2.16 (Approximating a function with the cosine family). In this example, we consider again the interval $I = [0, 1]$, using the complete orthonormal set $\{\varphi_k\}_{k \in \mathbb{N}_0}$ from Example 2.10, such that

$$\varphi_k(x) = \begin{cases} 1 & \text{for } k=0, \\ \sqrt{2}\cos(k\pi x) & \text{for } k>0. \end{cases}$$

The function f is defined as

$$f(x) = x \quad \text{for all } x \in [0, 1].$$

The coefficients γ_k of the generalized Fourier series of f with respect to $\{\varphi_k : k \in \mathbb{N}_0\}$ are found again using (2.7). These coefficients are

$$\gamma_0 = \int_0^1 x \, dx = \frac{1}{2},$$

while for $k \in \mathbb{N}$ we have

$$\gamma_k = \int_0^1 x \sqrt{2}\cos(k\pi x) \, dx = -\frac{\sqrt{2}(1-(-1)^k)}{k^2\pi^2} = \begin{cases} 0 & , \text{ for even } k, \\ -\frac{2\sqrt{2}}{k^2\pi^2} & , \text{ for odd } k. \end{cases}$$

The resulting approximations f_N , as defined in (2.10), are shown in the left image of Figure 2.2 for the cases $N = 10, 20, 50$, and the right image depicts the errors $|f_N(x) - f(x)|$. As in the previous example, the approximating sums seem to converge uniformly to f on all of $[0, 1]$. ■

Example 2.17 (An approximation with slow convergence). We now consider the approximation using the sine family $\{\varphi_k : k \in \mathbb{N}\}$ of the function f given by

$$f(x) = 1 \quad \text{for all } x \in [0, 1].$$

One can easily verify that the coefficients γ_k of the generalized Fourier series of f with respect to $\{\varphi_k : k \in \mathbb{N}\}$ are given by

$$\gamma_k = \int_0^1 \sqrt{2}\sin(k\pi x) \, dx = \frac{\sqrt{2}(1-(-1)^k)}{k\pi} = \begin{cases} 0 & , \text{ for even } k, \\ \frac{2\sqrt{2}}{k\pi} & , \text{ for odd } k. \end{cases}$$

Some resulting approximations f_N , as defined in (2.10), are shown in the left image of Figure 2.3 for the cases $N = 10, 20, 50$. While these approximations seem to get closer to the function $f(x) = 1$ as N increases, one does have to measure this closeness in an averaged L^2 -sense. To illustrate this, the right image of Figure 2.3 shows the corresponding

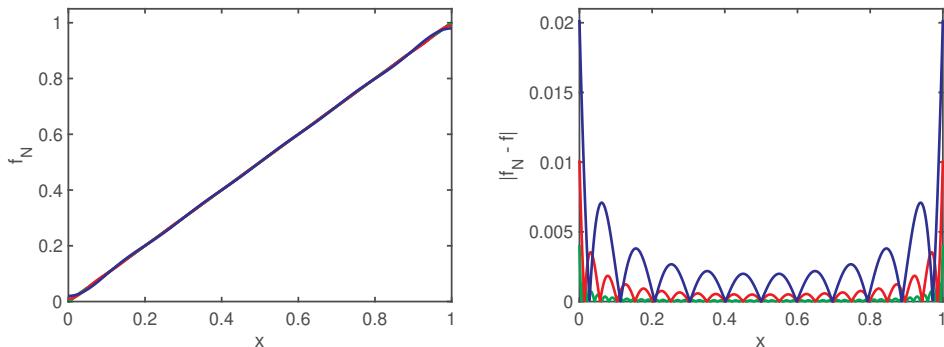


Figure 2.2. Generalized Fourier series approximations using the complete orthonormal set from Example 2.10 with $L = 1$ for the function $f(x) = x$, as described in Example 2.16. The image on the left contains the approximating sums $\sum_{k=1}^N \gamma_k \varphi_k$ for $N = 10, 20, 50$ in blue, red, and green, respectively. The figure on the right shows the corresponding errors $|f_N(x) - f(x)|$.

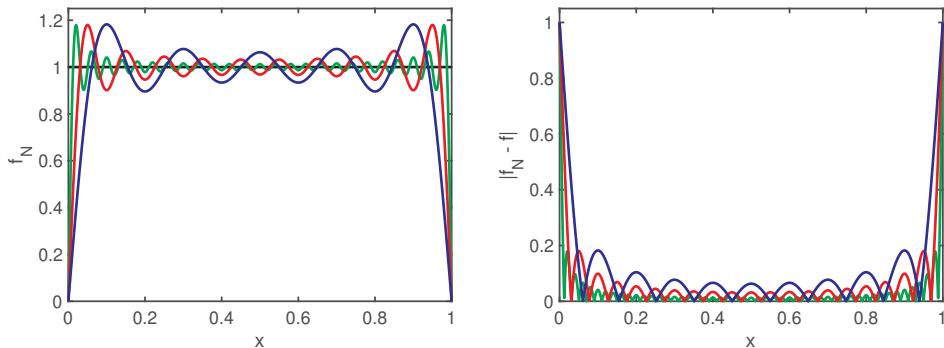


Figure 2.3. Generalized Fourier series approximations using the complete orthonormal set from Example 2.9 with $L = 1$ to approximate the function and $f(x) = 1$ as shown in Example 2.17. The image in the left column contains the approximating sums $\sum_{k=1}^N \gamma_k \varphi_k$ for $N = 10, 20, 50$ in blue, red, and green, respectively. The figure in the right column shows the corresponding errors $|f_N(x) - f(x)|$.

functions $|f_N(x) - f(x)|$. It is evident from this figure that the area under the squares of these graphs seems to decrease to zero. Nevertheless, the function values $f_N(x)$ do not converge to $f(x) = 1$ for all $x \in [0, 1]$, since we clearly have $f_N(0) = f_N(1) = 0$ for all values of $N \in \mathbb{N}$. One can show, however, that for all remaining x -values we do in fact obtain

$$\lim_{N \rightarrow \infty} f_N(x) = 1 \quad \text{for all } 0 < x < 1,$$

albeit with a slow rate of convergence. One might wonder what distinguishes this case from the previous cases, which are observed to have excellent convergence rates. The intuition for why the convergence is so slow comes from observing the mismatch of boundary conditions between the sine family and the function f : For every function $\varphi_k(x)$ in the sine family one has $\varphi_k(0) = \varphi_k(1) = 0$, yet $f(0) = f(1) = 1$. This type of oscillatory behavior near the boundary is characteristic in the case of mismatched boundary conditions, and is known as the Gibbs phenomenon. ■

Example 2.18 (Uniqueness of the generalized Fourier series). In this example, we illustrate the uniqueness of the generalized Fourier series. Consider the generalized Fourier series for the function $f(x) = 4\sqrt{2}\sin(2\pi x) + 3\sqrt{2}\sin(5\pi x)$ using the sine family. Note that due to $\varphi_k(x) = \sqrt{2}\sin(k\pi x)$ one has $f(x) = 4\varphi_2(x) + 3\varphi_5(x)$. Thus, one generalized Fourier series for f would be

$$f(x) = \sum_{k=1}^{\infty} c_k \varphi_k(x) \quad \text{where} \quad c_k = \begin{cases} 4 \text{ for } k = 2, \\ 3 \text{ for } k = 5, \\ 0 \text{ otherwise.} \end{cases}$$

Would we have obtained a different answer if we used the best-approximation formula? No! Theorem 2.13 guarantees that the best-possible approximation is unique. We illustrate this statement concretely using the linearity of the inner product, which implies the identity

$$\gamma_k = \langle f, \varphi_k \rangle = \langle 4\varphi_2 + 3\varphi_5, \varphi_k \rangle = 4\langle \varphi_2, \varphi_k \rangle + 3\langle \varphi_5, \varphi_k \rangle.$$

Due to the orthogonality of the functions φ_k , the first term of the right-hand side is nonzero only when $k = 2$, and the second term is nonzero only when $k = 5$. Thus, since $\|\varphi_k\|_{L^2(0,1)} = 1$, the two methods give the same answer, i.e., we have $\gamma_k = c_k$. ■

Convergence of the best approximation

According to Theorem 2.14, the approximation for f converges to f in the L^2 sense. In contrast, Example 2.17 shows that the convergence properties may not be particularly nice. So what does the L^2 -equality between the function f and its generalized Fourier series really mean? As was mentioned before, this equality has to be understood in the sense of convergence in the L^2 -norm, i.e., in an averaged sense. More precisely, assume again that the sequence $\{\varphi_k : k \in \mathbb{N}\}$ is a complete orthonormal set. If we define $f_N(x)$ as in (2.10) by the finite sum

$$f_N(x) = \sum_{k=1}^N \gamma_k \varphi_k(x), \quad \text{where} \quad \gamma_k = \langle f, \varphi_k \rangle = \int_I f(x) \varphi_k(x) dx,$$

then we have

$$\int_I (f(x) - f_N(x))^2 dx \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty.$$

We would like to point out, however, that in general there are values of x such that

$$\lim_{N \rightarrow \infty} f_N(x) \neq f(x) \quad \text{as} \quad N \rightarrow \infty.$$

In fact, there are even examples where at some x -values the sequence of partial sum function values $(f_N(x))_{N=1}^{\infty}$ does not converge at all, let alone to $f(x)$.

It is clear from these examples that the question of when a generalized Fourier series of a function f in fact satisfies the equality

$$f(x) = \sum_{k=1}^{\infty} \gamma_k \varphi_k(x) \quad \text{for all} \quad x \in I$$

is rather delicate. This is true even for the case of classical Fourier series, i.e. obtained by using the Fourier family introduced in Example 2.11.

Already at the time of their discovery, Fourier himself was interested in understanding when a Fourier series converges to its limit function pointwise. Since all functions φ_k in the complete orthonormal set from Example 2.11 satisfy $\varphi_k(0) = \varphi_k(L)$, any pointwise limit has to also satisfy this constraint. Let us therefore assume that

$$f \in L^2([0, L]) \quad \text{satisfies} \quad f(0) = f(L).$$

Is it always true that the Fourier series of f converges to f pointwise? It was shown early on in the 19th century that this is the case if f is continuously differentiable. And in fact, the general belief at the time was that it should also be true for all continuous functions f . However, in 1876 Paul David Gustav du Bois-Reymond gave an example of a continuous function f , whose Fourier series actually diverges at one point — thereby disproving the conjecture. Only fairly recently, in 1966, Lennart Carleson showed that for any function $f \in L^2([0, L])$ with $f(0) = f(L)$ the classical Fourier series converges to f “almost everywhere” in $[0, L]$.¹⁰

Despite these subtleties, generalized Fourier series will play a central role in many later parts of this book. Several of these applications will require us to be able to establish the uniform convergence of a given generalized Fourier series. In these situations, the following sufficient criterion is essential.

Theorem 2.19 (Uniform convergence of the generalized Fourier series). *Let $I \subset \mathbb{R}$ denote an arbitrary interval and consider the space $L^2(I)$ introduced in Definition 2.1. Furthermore, let $\{\varphi_k(x) : k \in \mathbb{N}\}$ denote a complete orthogonal set as in Definition 2.8, let $f \in L^2(I)$ be arbitrary, and denote the corresponding generalized Fourier series of f by*

$$f \stackrel{L^2}{=} \sum_{k=1}^{\infty} \gamma_k \varphi_k.$$

If there exists a sequence of nonnegative reals $M_k \geq 0$ with

$$\sum_{k=1}^{\infty} M_k < \infty,$$

and such that for some integer k_0 we have

$$|\gamma_k \varphi_k(x)| \leq M_k \quad \text{for all} \quad k \geq k_0 \quad \text{and} \quad x \in I,$$

then the generalized Fourier series of f converges uniformly on the interval I . If in addition all functions φ_k are continuous, then so is the limit function of the generalized Fourier series.

Proof. The proof of this theorem follows directly from the celebrated Weierstraß M -test, which for completeness we now state. \square

Theorem 2.20 (Weierstraß M -test). *Let $D \subset \mathbb{R}^n$ be arbitrary, and let $f_n : D \rightarrow \mathbb{R}$ denote an arbitrary function for every $n \in \mathbb{N}$. Furthermore, assume that there exists a sequence $(M_n)_{n \in \mathbb{N}}$ of nonnegative real numbers with*

$$\sum_{n=1}^{\infty} M_n < \infty,$$

¹⁰The precise definition of the notion of almost everywhere was given in Section 1.4. However, in case you have skipped our discussion of stochastic differential equations: It certainly suffices to have an intuitive understanding of this concept.

and such that

$$|f_n(x)| \leq M_n \quad \text{for all } x \in D \quad \text{and } n \in \mathbb{N}.$$

Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly on D to a function $f : D \rightarrow \mathbb{R}$. In particular, if all the functions f_n are continuous, then so is f .

Notice that in the theorem preceding the Weierstraß M -test we have refrained from stating the identity

$$f(x) = \sum_{k=1}^{\infty} \gamma_k \varphi_k(x) \quad \text{for all } x \in I. \quad (2.11)$$

This is due to the fact that in general, many functions lead to the same generalized Fourier series. To see this, take for example the function $f(x) = 0$ on I , with trivial generalized Fourier series. Since functions in $L^2(I)$ only have to be integrable, any function g which is equal to f for all but countably many x -values in I is an element of $L^2(I)$ with the same generalized Fourier series, but of course it does not equal the series everywhere. One can easily see that if all the functions φ_k are continuous, and if in addition f is continuous, then in the situation of the above theorem we do in fact have (2.11).

In addition, we would like to point out that if the functions φ_k in Theorem 2.19 are *uniformly bounded*, i.e., if there exists a constant $M > 0$ such that

$$|\varphi_k(x)| \leq M \quad \text{for all } x \in I \quad \text{and } k \in \mathbb{N},$$

then the theorem applies as long as the generalized Fourier coefficients satisfy

$$\sum_{k=1}^{\infty} |\gamma_k| < \infty. \quad (2.12)$$

This uniform boundedness is given for the complete orthonormal sets introduced in Examples 2.4, 2.5, and 2.6.

Armed with Theorem 2.19 and the simplified criterion (2.12), we can now revisit our earlier examples. Notice that the generalized Fourier coefficients of $f(x) = 4x(1-x)$ in Example 2.15 satisfy

$$|\gamma_k| \leq \frac{C}{k^3} \quad \text{for some constant } C > 0,$$

and the convergence of the series $\sum_{k=1}^{\infty} k^{-3}$ therefore implies the uniform convergence of the generalized Fourier series. Similarly, one can easily show that the series for $f(x) = x$ in Example 2.16 converges uniformly. Notice, however, that Theorem 2.19 does not apply to the series for $f(x) = 1$ in Example 2.17, since the corresponding generalized Fourier coefficients decay only like $1/k$, and the harmonic series $\sum_{k=1}^{\infty} k^{-1}$ is divergent.

The Weierstraß M -test presented in Theorem 2.20 gives conditions which ensure the continuity of a series of continuous functions. As our main interest in this book is the study of partial differential equations, we will also need to be able to determine when a series of differentiable functions is differentiable. For this, we make use of the following result from real analysis, which is stated without proof.

Theorem 2.21 (Differentiability of a series). *Let $D \subset \mathbb{R}$ be closed, and let $f_n : D \rightarrow \mathbb{R}$ denote an arbitrary continuously differentiable function for every $n \in \mathbb{N}$. Furthermore, assume that the series $\sum_{n=1}^{\infty} f_n$ converges at some point a in D , and $\sum_{n=1}^{\infty} f'_n$ converges uniformly on the domain D . Then $F = \sum_{n=1}^{\infty} f_n$ converges uniformly on D . Furthermore, F is differentiable on D , and $F' = \sum_{n=1}^{\infty} f'_n$.*

2.1.3 • Extensions and Generalizations

In the last few sections we have presented a self-contained introduction to generalized Fourier series, centered around the need for expressing arbitrary functions in the space $L^2(I)$ through a suitable basis of functions. This has led to the concept of complete orthonormal sets — which will accompany us throughout the remainder of the book. Yet, in the interest of clarity, we have restricted our discussion to the case of functions which are defined on a subset $I \subset \mathbb{R}$. As it turns out, this restriction can easily be relaxed. Also, our choice of inner product in the function space $L^2(I)$ is not always suitable for a given problem, and one would therefore like to be able to consider more general inner products, which still preserve the validity of the results from the previous sections. These two extensions will be discussed in the remainder of the present section.

Inner products with weight functions

We begin by generalizing the notion of inner product in the space $L^2(I)$. In the standard Euclidean space \mathbb{R}^d , the canonical scalar product is only one of many possibilities. In fact, for any symmetric and positive definite matrix $A \in \mathbb{R}^{d \times d}$ the definition

$$x \cdot y = x^t A y \quad \text{for all } x, y \in \mathbb{R}^d$$

provides another inner product¹¹. One can still define the notions of orthogonality and norm as before, and in fact the basic properties of the terms obtained in this way remain unchanged. In some sense, introducing a different scalar product as above can be seen as transforming the notion of length, or as weighing different coordinate directions in different ways.

The situation is analogous for the case of functions in $L^2(I)$. As we will see soon, the need for other inner products arises naturally in the context of Sturm-Liouville problems. For this, we need the following definition.

Definition 2.22 (Weighted inner products). *Let $I \subset \mathbb{R}$ denote an arbitrary interval, and let $w : I \rightarrow \mathbb{R}^+$ denote a strictly positive piecewise continuous function. Then we define the weighted L^2 -space $L_w^2(I)$ by*

$$L_w^2(I) = \left\{ f : I \rightarrow \mathbb{R} : f \text{ is integrable and } \int_I |f(x)|^2 w(x) dx < \infty \right\}.$$

In this vector space, the sum of two functions and the scalar multiplication are defined pointwise as in (2.1) and (2.2), respectively. For two functions $f, g \in L_w^2(I)$, we define the weighted inner product of f and g by

$$\langle f, g \rangle_w = \int_I f(x) g(x) w(x) dx.$$

For every function $f \in L_w^2(I)$, its L_w^2 -norm is defined as

$$\|f\|_{L_w^2(I)} = \left(\int_I |f(x)|^2 w(x) dx \right)^{1/2} = \sqrt{\langle f, f \rangle_w}.$$

¹¹As mentioned earlier, the notation t denotes the transpose of a matrix. Therefore, for a column vector x , the notation x^t denotes the corresponding row vector. See also Remark 1.20, where we outlined our convention regarding vectors.

Finally, two functions f and g in $L_w^2(I)$ are called L_w^2 -orthogonal, if $\langle f, g \rangle_w = 0$.

It is clear from the above definition that this new concept reduces to the previous one if we use the weight function $w(x) \equiv 1$. In general, different weight functions lead to different orthogonal sets, as the following simple example shows.

Example 2.23. Consider the interval $I = [0, 1]$ and the weight function

$$w(x) = e^{2x} \quad \text{for all } x \in I = [0, 1].$$

Then one can check that the infinite set of functions given by

$$\varphi_n(x) = e^{-x} \sin(n\pi x) \quad \text{for } n \in \mathbb{N}$$

is an orthogonal set, i.e., for all $i \neq j$ one has $\langle \varphi_i, \varphi_j \rangle_w = 0$. Yet these functions are not orthogonal with respect to the standard inner product on $L^2(I)$. ■

Example 2.24 (Hermite polynomials). Consider the interval $I = \mathbb{R}$ and the weight function

$$w(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Then the polynomial functions

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad \text{and} \quad H_3(x) = x^3 - 3x$$

are a set of orthogonal functions contained in $L_w^2(I)$. To show this statement, we combine the fact that

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

with integration by parts to show that each function H_k for $k = 0, 1, 2, 3$ is contained in $L_w^2(I)$. In fact, one can show that $\langle H_k, H_k \rangle_w = k!$ for all k . Moreover, if j is an odd integer, then $x^j e^{-x^2/2}$ is an odd function, and one obtains the identity $\int_{-\infty}^{\infty} x^j e^{-x^2/2} dx = 0$. This fact can then be used to show that each distinct pair of the functions H_k is orthogonal. The functions H_k are the first four *Hermite polynomials*, which form a complete orthogonal set of functions on \mathbb{R} with respect to the weight function w . ■

We close this section with the following reassuring remark.

Remark 2.25. All the results and properties concerning the standard L^2 -inner product remain valid for the weighted L_w^2 -inner product. This applies to all results of both Section 2.1.1 and Section 2.1.2. However, in order to keep our notation as simple as possible, we will not formulate these results again for the weighted inner product $\langle \cdot, \cdot \rangle_w$.

Higher-dimensional base domains

Our second extension is concerned with the base domain of the space $L^2(I)$. One can fairly easily see that all of the results described in Sections 2.1.1 and 2.1.2 rely heavily on the fact that the functions in this space are real-valued (although this could be generalized). However, for the base domain we need hardly any specific structure at all. It is therefore not surprising that the results can be extended to the case of functions of several variables

on higher-dimensional base domains. This leads to the following definition, in which we change notation slightly, including the change of base domain from an interval I to a higher-dimensional region Ω .

Definition 2.26 ($L^2(\Omega)$ and orthogonal functions). Let $\Omega \subset \mathbb{R}^d$ denote an arbitrary open or closed set.¹² Then we define the vector space $L^2(\Omega)$ as

$$L^2(\Omega) = \left\{ f : \Omega \rightarrow \mathbb{R} : f \text{ is integrable and } \int_{\Omega} |f(x)|^2 dx < \infty \right\}.$$

In this vector space, the sum of two functions and the scalar multiplication are defined pointwise as in the case of one-dimensional base domains $I \subset \mathbb{R}$. For two functions $f, g \in L^2(\Omega)$, we define the inner product of f and g by

$$\langle f, g \rangle = \int_{\Omega} f(x)g(x) dx.$$

For every function $f \in L^2(\Omega)$, its L^2 -norm is defined as

$$\|f\|_{L^2(\Omega)} = \left(\int_{\Omega} |f(x)|^2 dx \right)^{1/2} = \sqrt{\langle f, f \rangle}.$$

Finally, two functions f and g in $L^2(\Omega)$ are called orthogonal, if $\langle f, g \rangle = 0$.

In the above definition, other than the slight change of notation, nothing much changes, as the following remark asserts.

Remark 2.27. All results and properties concerning the standard $L^2(I)$ -inner product remain valid for the inner product in $L^2(\Omega)$, regardless of the dimension of the domain $\Omega \subset \mathbb{R}^d$. This applies to all results of both Section 2.1.1 and Section 2.1.2, as well as the concept of weighted inner products discussed above.

2.2 ■ Sturm-Liouville Boundary Value Problems

As we mentioned earlier, one of the main analytical techniques of the present chapter will enable us in many situations to reduce partial differential equations on higher-dimensional domains to the study of associated boundary value problems on one-dimensional domains — which are nothing but boundary value problems for ordinary differential equations. Thus, in this section we return to the case of ordinary differential equations. In particular, we will study a class of problems called Sturm-Liouville boundary value problems.

2.2.1 ■ Symmetry and Regular Sturm-Liouville Problems

As shown in the previous section, complete orthonormal sets are the natural generalization of the notion of basis to the infinite-dimensional setting. However, apart from a few specific examples, we have not presented any general idea of how such sets of functions

¹²We need some property on the base domain which ensures that the notion of integral makes sense. In particular, we want to make sure that all “reasonable functions” such as continuous functions, and in particular constant functions, are integrable. At the very least, for bounded Ω we have to be able to integrate the function which is constant 1, which then furnishes the notion of area, volume, etc. of the set Ω . For our purposes, the notions of openness and closedness suffice, even though this could be generalized.

can actually be found. In this and the following section we address how to find a complete orthonormal set. As we will see, complete orthonormal sets of functions arise naturally in the context of boundary value problems associated with second-order linear ordinary differential equations.

Second-order boundary value problems

We begin with a brief discussion of such boundary value problems. For this, consider a bounded interval $I = [a, b] \subset \mathbb{R}$ and a differential equation of the form

$$a_2(x)u'' + a_1(x)u' + a_0(x)u = 0 \quad (2.13)$$

for an unknown function $u : [a, b] \rightarrow \mathbb{R}$, where the coefficient functions a_0 , a_1 , and a_2 are continuous on the interval $I = [a, b]$. In addition, we assume that the coefficient function a_2 has no zero in I , i.e., it is either strictly positive or strictly negative on all of I . This differential equation has infinitely many solutions. In fact, one can show that if u_1 and u_2 are a pair of so-called *fundamental solutions*,¹³ then every solution of (2.13) is of the form $c_1u_1 + c_2u_2$, for suitable $c_1, c_2 \in \mathbb{R}$. This immediately implies that in order to single out a unique solution, one needs to specify two additional constraints. As discussed in Section 1.1.1 this is usually accomplished by an initial condition of the form $u(a) = u_0$ and $u'(a) = u_1$ (or alternatively $u(b) = u_0$ and $u'(b) = u_1$), for given constants $u_0, u_1 \in \mathbb{R}$, i.e., one imposes two conditions at one endpoint of the interval I .

It is natural to wonder whether it is also possible to impose conditions on the function values of u or u' at both endpoints of I , which leads to the concept of a *boundary value problem*. For the purposes of this section, we consider boundary conditions of the form

$$\alpha_1u(a) + \alpha_2u'(a) = 0 \quad \text{and} \quad \beta_1u(b) + \beta_2u'(b) = 0, \quad (2.14)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$. More general boundary conditions will be discussed later. In order to develop some basic intuition regarding the solvability of (2.13) subject to the boundary conditions (2.14), we start with two simple examples.

Example 2.28 (A boundary value problem with only the trivial solution). Consider the second-order linear differential equation

$$u'' - u = 0 \quad \text{on the interval} \quad I = [0, 1], \quad (2.15)$$

subject to the boundary conditions

$$u(0) = 0 \quad \text{and} \quad u'(1) = 0. \quad (2.16)$$

Clearly this boundary value problem falls into the category discussed above. In order to find all solutions, one first needs to find a set of fundamental solutions of the differential equation (2.15). One can easily see that the functions $u_1(x) = e^x$ and $u_2(x) = e^{-x}$ are linearly independent solutions, and therefore a fundamental set. This implies that every solution of (2.15) is of the form

$$u(x) = c_1e^x + c_2e^{-x} \quad \text{for some} \quad c_1, c_2 \in \mathbb{R}.$$

¹³Recall that two solutions u_1 and u_2 of (2.13) are called *fundamental solutions*, if they are linearly independent, i.e., if $c_1, c_2 \in \mathbb{R}$ are two constants such that $c_1u_1(x) + c_2u_2(x) = 0$ for all $x \in [a, b]$, then one necessarily has $c_1 = c_2 = 0$. This is equivalent to the linear independence of the vectors $(u_1(x), u'_1(x))$ and $(u_2(x), u'_2(x))$ in \mathbb{R}^2 for some $x \in [a, b]$. Finally, one can show that u_1 and u_2 are linearly independent if and only if $(u_1(x), u'_1(x))$ and $(u_2(x), u'_2(x))$ are linearly independent for all $x \in [a, b]$.

Plugging the above form of u into the boundary conditions (2.16) then furnishes the linear system

$$c_1 + c_2 = 0 \quad \text{and} \quad c_1 e - c_2 e^{-1} = 0$$

for the unknown coefficients c_1 and c_2 . Since this system only has the solution $c_1 = c_2 = 0$, the boundary value problem (2.15), (2.16) has the unique solution u with $u(x) = 0$ for all $x \in [0, 1]$. We refer to this solution as the *trivial solution*. ■

Example 2.29 (A boundary value problem with infinitely-many solutions).

Consider again the second-order linear differential equation

$$u'' - u = 0 \quad \text{on the interval} \quad I = [0, 1], \quad (2.17)$$

but this time subject to the boundary conditions

$$u'(0) = 0 \quad \text{and} \quad (1 - e^2)u(1) + (1 + e^2)u'(1) = 0. \quad (2.18)$$

As in the previous example, the functions $u_1(x) = e^x$ and $u_2(x) = e^{-x}$ form a fundamental set for (2.17), and thus every solution of this differential equation is of the form

$$u(x) = c_1 e^x + c_2 e^{-x} \quad \text{for some} \quad c_1, c_2 \in \mathbb{R}.$$

Plugging this form of u into the boundary conditions (2.18) now yields the linear system

$$c_1 - c_2 = 0 \quad \text{and} \quad (1 - e^2)(c_1 e + c_2 e^{-1}) + (1 + e^2)(c_1 e - c_2 e^{-1}) = 0,$$

which is equivalent to the system

$$c_1 = c_2 \quad \text{and} \quad c_1 \left(\underbrace{(1 - e^2)(e + e^{-1}) + (1 + e^2)(e - e^{-1})}_{=0} \right) = 0.$$

In other words, for every constant $c \in \mathbb{R}$ the function $u(x) = ce^x + ce^{-x}$ is a solution to the boundary value problem (2.17), (2.18). ■

In the following example, we consider a family of boundary value problems depending on a parameter λ . We show that each boundary value problem in the family has either infinitely many solutions, or only the trivial solution. As we develop the theory further, we will return to this example again as an illustration.

Example 2.30 (A family of boundary value problems). For fixed $\lambda \in \mathbb{R}$ and $L > 0$, consider the boundary value problem

$$u'' + \lambda u = 0, \quad \text{subject to} \quad u(0) = u(L) = 0.$$

In order to solve this boundary value problem, we have to distinguish three cases, depending on the sign of the parameter λ . This is due to the fact that the fundamental set of solutions for the above differential equation is different in each case.

- (1) If $\lambda < 0$, then the fundamental solutions to the differential equation are e^{px} and e^{-px} for $p = \sqrt{-\lambda}$, i.e., all solutions of the differential equation are of the form

$$u(x) = c_1 e^{px} + c_2 e^{-px}, \quad \text{where} \quad p = \sqrt{-\lambda}.$$

By our first boundary condition, we have

$$c_1 u_1(0) + c_2 u_2(0) = c_1 + c_2 = 0,$$

implying that $c_1 = -c_2$, and our solution is of the form $u(x) = c_1(e^{px} - e^{-px})$. By our second boundary condition, we have $u(L) = c_1(e^{pL} - e^{-pL}) = 2c_1 \sinh(pL)$. However, the function $\sinh(s)$ is only zero for $s = 0$, and we have assumed $p > 0$ and $L > 0$. Therefore $c_1 = 0$, and this implies that the only solution u is the trivial solution.

- (2) If $\lambda = 0$, then the solution to the differential equation is of the form $u(x) = c_1 + c_2 x$. By the first boundary condition, $u(0) = c_1 = 0$. By the second boundary condition, one obtains $u(L) = c_2 L = 0$. Therefore, the only solution is once again trivial in this case, since $L > 0$ implies $c_2 = 0$.
- (3) If $\lambda > 0$, then the fundamental solutions are $\cos(\sqrt{\lambda}x)$ and $\sin(\sqrt{\lambda}x)$, and therefore all solutions are of the form $u(x) = c_1 \cos(\sqrt{\lambda}x) + c_2 \sin(\sqrt{\lambda}x)$. The first boundary condition yields $c_1 = 0$, and this implies

$$u(x) = c_2 \sin(px) \quad \text{where} \quad p = \sqrt{\lambda}.$$

The second boundary condition requires $pL = \ell\pi$, i.e., in this case we need

$$p = \frac{\ell\pi}{L} \quad \text{for} \quad \ell \in \mathbb{N},$$

and we obtain the solutions $u(x) = c_2 \sin(\ell\pi x/L)$ for arbitrary $c_2 \in \mathbb{R}$. If on the other hand one has $pL \neq \ell\pi$, then there is only the trivial solution.

To summarize our findings, we have shown that there exist a countably infinite number of λ -values

$$\lambda_\ell = \frac{\ell^2\pi^2}{L^2} \quad \text{for} \quad \ell \in \mathbb{N},$$

such that for each such value λ_ℓ and for each $c_2 \in \mathbb{R}$,

$$\varphi_\ell(x) = c_2 \sin \frac{\ell\pi x}{L}$$

is a solution of the boundary value problem. That is, for each λ_ℓ there are infinitely-many solutions of the problem. Note that each such solution is a multiple of a member of the previously-discussed sine family. On the other hand, for every $\lambda \in \mathbb{R}$ such that $\lambda \neq \lambda_\ell$, the only possible solution is the trivial solution $u \equiv 0$. ■

These examples show that in contrast to the case of initial conditions, imposing two boundary conditions does not necessarily imply the existence of a unique solution to the boundary value problem (2.13), (2.14). Nevertheless, the examples do give us a general approach for determining the set of all solutions.

For this, we first consider the differential equation (2.13) by itself. It follows from the theory of linear ordinary differential equations that there are two linearly independent solutions u_1 and u_2 of (2.13) such that all solutions of this equation are of the form

$$u(x) = c_1 u_1(x) + c_2 u_2(x), \quad \text{where} \quad c_1, c_2 \in \mathbb{R}.$$

If we now try to determine which of these solutions satisfy the boundary conditions in (2.14), we have to solve the system

$$\begin{aligned} c_1 (\alpha_1 u_1(a) + \alpha_2 u'_1(a)) + c_2 (\alpha_1 u_2(a) + \alpha_2 u'_2(a)) &= 0, \\ c_1 (\beta_1 u_1(b) + \beta_2 u'_1(b)) + c_2 (\beta_1 u_2(b) + \beta_2 u'_2(b)) &= 0. \end{aligned}$$

In other words, a pair (c_1, c_2) gives a solution of the boundary value problem if and only if the vector (c_1, c_2) is contained in the nullspace of the matrix

$$\begin{pmatrix} \alpha_1 u_1(a) + \alpha_2 u'_1(a) & \alpha_1 u_2(a) + \alpha_2 u'_2(a) \\ \beta_1 u_1(b) + \beta_2 u'_1(b) & \beta_1 u_2(b) + \beta_2 u'_2(b) \end{pmatrix} \in \mathbb{R}^{2 \times 2}. \quad (2.19)$$

That is, there are nontrivial solutions of the boundary value problem if and only if the above matrix is singular. In Examples 2.28, 2.29, and 2.30(3) this matrix is given by

$$A_1 = \begin{pmatrix} 1 & 1 \\ e & -e^{-1} \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & -1 \\ 2e & -2e \end{pmatrix}, \quad \text{and} \quad A_3 = \begin{pmatrix} 1 & 0 \\ \cos(L\sqrt{\lambda}) & \sin(L\sqrt{\lambda}) \end{pmatrix},$$

respectively. Thus, the boundary value problem in Example 2.28 has the unique solution $(c_1, c_2) = 0$ since A_1 is invertible, and the boundary value problem in Example 2.29 has a one-dimensional solution space given by $(c_1, c_2) = c(1, 1)$ for $c \in \mathbb{R}$ since A_2 has a one-dimensional nullspace. The boundary problem in Example 2.30(3) has a nontrivial solution exactly when A_3 is singular, i.e., when $\sin(L\sqrt{\lambda}) = 0$. In fact, we can now describe the structure of the set of solutions of the boundary value problems (2.13), (2.14).

Proposition 2.31 (Number of solutions of a boundary value problem). Consider a bounded interval $I = [a, b] \subset \mathbb{R}$, and assume that the functions a_0 , a_1 , and a_2 are continuous on the interval $I = [a, b]$. In addition, we assume that the function a_2 has no zero in I , i.e., it is either strictly positive or strictly negative on all of I . Consider the boundary value problem given by the second-order linear differential equation

$$a_2(x)u'' + a_1(x)u' + a_0(x)u = 0 \quad (2.20)$$

for an unknown function $u : [a, b] \rightarrow \mathbb{R}$, subject to the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0, \quad (2.21)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$. Then exactly one of the following two statements holds.

- (a) The boundary value problem (2.20), (2.21) has only the trivial solution $u = 0$.
- (b) There exists a nontrivial function $\varphi : [a, b] \rightarrow \mathbb{R}$ such that all solutions of the boundary value problem (2.20), (2.21) are of the form $c\varphi$, where $c \in \mathbb{R}$ is arbitrary.

In other words, the solution set of (2.20), (2.21) consists either of the trivial solution, or of a one-dimensional solution space which can be represented as the multiples of a nontrivial solution φ .

Proof. Let u_1 and u_2 denote a fundamental set of solutions for (2.20), and let $A \in \mathbb{R}^{2 \times 2}$ denote the matrix in (2.19).

Due to our discussion preceding the proposition, a function $u(x) = c_1 u_1(x) + c_2 u_2(x)$ solves the boundary value problem (2.20), (2.21) if and only if the vector (c_1, c_2) is contained in the nullspace of A . Thus, if the matrix A is invertible, then the boundary value

problem has only the trivial solution, which corresponds to $(c_1, c_2) = (0, 0)$, and statement (a) holds. On the other hand, if A is singular but not the zero matrix, its nullspace is one-dimensional, and one can easily see that (b) is satisfied.

In order to complete the proof we only have to show that A can never be the zero matrix. For this, assume that for some $\alpha_1, \alpha_2, \beta_1$, and β_2 we have $A = 0 \in \mathbb{R}^{2 \times 2}$. Due to the definition of A , this is equivalent to the linear system

$$\underbrace{\begin{pmatrix} u_1(a) & u'_1(a) \\ u_2(a) & u'_2(a) \end{pmatrix}}_{=B} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Since u_1 and u_2 form a fundamental set, both the 2×2 -submatrix in the upper left corner and the 2×2 -submatrix in the lower right corner of B are invertible. This, shows that the matrix B is invertible as well, which implies that we have to have $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$. This contradicts our choice of these constants, and completes the proof. \square

Notice that the proof of the proposition also shows that both of the rows of the matrix in (2.19) have to be nontrivial, i.e., at least one entry in each row is nonzero.

Symmetric linear differential operators

Before returning to our discussion of complete orthonormal sets, we would like to take another look at boundary value problems from a different perspective. For this, consider again the situation of Proposition 2.31. We define the *general second-order linear differential operator* L by

$$(Lu)(x) = a_2(x)u''(x) + a_1(x)u'(x) + a_0(x)u(x) \quad \text{for } x \in I = [a, b] \subset \mathbb{R}, \quad (2.22)$$

and the *space \mathcal{B} of admissible functions* by

$$\mathcal{B} = \{u \in C^2[a, b] : u \text{ satisfies (2.21)}\}. \quad (2.23)$$

Then solving the boundary value problem (2.20), (2.21) can be restated as finding a function $u \in \mathcal{B}$ which satisfies $Lu = 0$.

On a purely formal level, the equation $Lu = 0$ resembles a linear system of equations of the form $Az = 0$, where A denotes an arbitrary square matrix and z a corresponding vector. If one is interested in finding nontrivial solutions $z \neq 0$, this equation reduces to finding a nontrivial nullspace of the matrix A . There is, however, another point of view which will prove to be useful. Any vector $z \neq 0$ which satisfies $Az = 0$ can also be thought of as an *eigenvector* of the matrix A which corresponds to the *eigenvalue* $\lambda = 0$, since it solves the equation $Az = \lambda z$. As the reader may recall from linear algebra, eigenvalues and eigenvectors lie at the heart of any deeper understanding of matrices. In certain situations, for example for symmetric matrices A , nontrivial eigenvectors can be used to construct orthogonal bases for the underlying vector space — and therefore it is natural to wonder whether nontrivial solutions $\varphi \in \mathcal{B}$ of boundary value problems of the form $L\varphi = \lambda\varphi$ can be used to construct complete orthonormal sets. In keeping with the linear algebra notation, if such a function φ exists for a number λ , we call λ an *eigenvalue* of the linear differential operator L with corresponding *eigenfunction* φ .

As we will see in the next section, eigenvalue problems associated with boundary value problems can indeed provide orthonormal sets of functions. Yet as in the case of matrices A , additional conditions on the underlying differential operator L are necessary in general, and deriving such a condition is the subject of the remainder of this section.

It was mentioned earlier that if $A \in \mathbb{R}^{d \times d}$ is a symmetric matrix, then there always exists a basis of \mathbb{R}^d which consists of eigenvectors of A . But how could we generalize the notion of symmetry to differential operators? In the matrix case, A being a symmetric matrix means that A has to equal its transpose matrix A^t , but it is not immediately clear how such a characterization could be generalized to differential operators. Notice, however, that if A is a symmetric matrix and $w, z \in \mathbb{R}^d$ are arbitrary, then we have

$$(Aw) \cdot z = (Aw)^t z = w^t A^t z = w \cdot (A^t z) = w \cdot (Az).$$

In this sequence of equations, $z_1 \cdot z_2 = z_1^t z_2$ denotes the dot product of two arbitrary column vectors $z_1, z_2 \in \mathbb{R}^d$. It is not hard to see that if, conversely, a matrix A satisfies the above equation, then it has to be symmetric — just pick w and z as two canonical basis vectors in \mathbb{R}^d . Together, we have shown that a matrix

$$A \in \mathbb{R}^{d \times d} \text{ is symmetric if and only if } (Aw) \cdot z = w \cdot (Az) \text{ for all } w, z \in \mathbb{R}^d. \quad (2.24)$$

Together with our previous discussion of the space of square integrable functions and the associated inner product of functions, the following definition should now not come as a surprise.

Definition 2.32 (Symmetric operator). *Let $I = [a, b] \subset \mathbb{R}$ denote a bounded interval, and consider the space $L^2(I)$ of square integrable functions and the inner product $\langle \cdot, \cdot \rangle$ as in Definition 2.1. Furthermore, let $\mathcal{B} \subset L^2(I)$ be some subspace of admissible functions, and consider a mapping $L : \mathcal{B} \rightarrow L^2(I)$ which assigns a function $Lu \in L^2(I)$ to every $u \in \mathcal{B}$. Then we say that the mapping L is a symmetric operator, or simply symmetric, if*

$$\langle Lu, v \rangle = \langle u, Lv \rangle \quad \text{for all } u, v \in \mathcal{B}. \quad (2.25)$$

In order to illustrate the above definition, consider the differential operator

$$Lu = u'' \quad \text{for } u \in C^2[a, b].$$

Then for arbitrary functions $u, v \in C^2[a, b]$ we have

$$\langle Lu, v \rangle = \int_a^b u''(x)v(x)dx \quad \text{and} \quad \langle u, Lv \rangle = \int_a^b u(x)v''(x)dx,$$

and integrating by parts twice furnishes

$$\begin{aligned} \langle Lu, v \rangle &= \int_a^b u''(x)v(x)dx = - \int_a^b u'(x)v'(x)dx + (u'(b)v(b) - u'(a)v(a)) \\ &= \int_a^b u(x)v''(x)dx + (u'(b)v(b) - u'(a)v(a)) - (u(b)v'(b) - u(a)v'(a)) \\ &= \langle u, Lv \rangle + (u'(b)v(b) - u'(a)v(a) - u(b)v'(b) + u(a)v'(a)). \end{aligned}$$

In other words, the operator L is symmetric if and only if the expression in parentheses vanishes. This can be accomplished by further restricting the set of admissible functions via boundary conditions. For example, it is easy to see that both

$$\begin{aligned} Lu = u'' &\quad \text{for } u \in \mathcal{B} = \{v \in C^2[a, b] : v(a) = v(b) = 0\} \quad \text{and} \\ Lu = u'' &\quad \text{for } u \in \mathcal{B} = \{v \in C^2[a, b] : v'(a) = v'(b) = 0\} \end{aligned}$$

lead to symmetric operators. However, not all differential operators lead to symmetric operators no matter which set of boundary conditions are used. To demonstrate this, consider the differential operator

$$Lu = e^x u'' \quad \text{for } u \in C^2[a, b].$$

Then for arbitrary functions $u, v \in C^2[a, b]$ one obtains after two-fold integration by parts the expression

$$\langle Lu, v \rangle = \langle u, L^* v \rangle + (B(b) - B(a)),$$

where

$$\begin{aligned} (L^* v)(x) &= e^x v''(x) + 2e^x v'(x) + e^x v(x) \quad \text{and} \\ B(x) &= e^x (u'(x)v(x) - u(x)v'(x) - u(x)v(x)). \end{aligned}$$

While the choice of appropriate boundary conditions can be used to enforce both $B(a) = 0$ and $B(b) = 0$, it is clear that in general we still have $L \neq L^*$!

A closer look at the previous two examples shows that the non-constant factor e^x in front of u'' is responsible for the lack of symmetry of the second operator, since integration by parts introduces additional expressions involving v or v' . This can be avoided if we consider differential operators of the following specific form.

Definition 2.33 (Differential operator in divergence form). Let $I = [a, b] \subset \mathbb{R}$ be a bounded interval, and assume that $q : I \rightarrow \mathbb{R}$ is continuous. In addition, let $p : I \rightarrow \mathbb{R}$ be a continuously differentiable function without zeros. Then a differential operator L of the form

$$(Lu)(x) = \frac{d}{dx} (p(x)u'(x)) + q(x)u(x) \tag{2.26}$$

is called a differential operator in divergence form. One can easily see that this is equivalent to the representation

$$Lu = p(x)u'' + p'(x)u' + q(x)u, \tag{2.27}$$

where $u \in C^2[a, b]$ is arbitrary.

The importance of differential operators in divergence form stems from the following result. It shows that such differential operators are automatically symmetric, provided we impose appropriate boundary conditions on the involved functions.

Theorem 2.34 (Divergence form implies symmetry). Let L be a differential operator in divergence form as introduced in Definition 2.33, and consider the set \mathcal{B} of admissible functions given by

$$\mathcal{B} = \{u \in C^2[a, b] : u \text{ satisfies (2.29) below}\}, \tag{2.28}$$

i.e., functions in \mathcal{B} satisfy the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0, \quad (2.29)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$. Then we have

$$\langle Lu, v \rangle = \langle u, Lv \rangle \quad \text{for all } u, v \in \mathcal{B}. \quad (2.30)$$

In other words, the operator L is symmetric on \mathcal{B} .

Proof. Let $u, v \in \mathcal{B}$ be arbitrary. Then integration by parts furnishes

$$\begin{aligned} \langle Lu, v \rangle &= \int_a^b \left(\frac{d}{dx} (p(x)u'(x)) v(x) + q(x)u(x)v(x) \right) dx \\ &= p(x)u'(x)v(x)|_a^b - \int_a^b p(x)u'(x)v'(x) dx + \int_a^b q(x)u(x)v(x) dx \\ &= (p(x)u'(x)v(x) - p(x)u(x)v'(x))|_a^b \\ &\quad + \int_a^b \left(\frac{d}{dx} (p(x)v'(x)) u(x) + q(x)u(x)v(x) \right) dx \\ &= p(x)(u'(x)v(x) - u(x)v'(x))|_a^b + \langle u, Lv \rangle. \end{aligned}$$

Thus, it suffices to prove that

$$u'(a)v(a) = u(a)v'(a) \quad \text{and} \quad u'(b)v(b) = u(b)v'(b).$$

We only establish the first identity, the second one follows similarly. Assume first that $\alpha_1 \neq 0$. Since u and v satisfy (2.29), we have both

$$u(a) = -\frac{\alpha_2}{\alpha_1} u'(a) \quad \text{and} \quad v(a) = -\frac{\alpha_2}{\alpha_1} v'(a),$$

which therefore implies

$$u'(a)v(a) = -\frac{\alpha_2}{\alpha_1} u'(a) v'(a) = u(a)v'(a).$$

Similarly, if $\alpha_2 \neq 0$, then

$$u'(a) = -\frac{\alpha_1}{\alpha_2} u(a) \quad \text{and} \quad v'(a) = -\frac{\alpha_1}{\alpha_2} v(a),$$

which now yields

$$u'(a)v(a) = -\frac{\alpha_1}{\alpha_2} u(a) v(a) = u(a)v'(a).$$

Due to $(\alpha_1, \alpha_2) \neq (0, 0)$ this completes the proof. \square

The importance of the above theorem is that it shows that for differential operators L in divergence form, and subject to the boundary conditions of the above form, one can hope to construct complete orthonormal sets by solving the eigenvalue problem

$$L\varphi = \lambda\varphi$$

with $\varphi \in \mathcal{B} \setminus \{0\}$. But what about second-order linear differential operators of the general form given in (2.13)? In this case, the differential operator L is defined by

$$Lu = a_2(x)u'' + a_1(x)u' + a_0(x)u,$$

and the coefficient function a_1 is usually not the derivative of a_2 — in fact, a_2 does not even have to be differentiable. As it turns out, one can always multiply such an operator by a suitable function so that the new differential operator is of divergence form. More precisely, we have the following result.

Lemma 2.35 (Conversion to divergence form). *Let $I = [a, b] \subset \mathbb{R}$ denote a bounded interval, and assume that the functions a_0 , a_1 , and a_2 are continuous on I . In addition, assume that a_2 has no zero in I , i.e., it is either strictly positive or strictly negative on all of I . Finally, consider the second-order linear differential operator L defined by*

$$Lu = a_2(x)u'' + a_1(x)u' + a_0(x)u.$$

Then the differential operator

$$L_{\text{div}} u = \frac{\exp\left(\int \frac{a_1(x)}{a_2(x)} dx\right)}{a_2(x)} Lu$$

is in divergence form. More precisely, we have

$$(L_{\text{div}} u)(x) = \frac{d}{dx} (p(x)u'(x)) + q(x)u(x),$$

where

$$p(x) = \exp\left(\int \frac{a_1(x)}{a_2(x)} dx\right) \quad \text{and} \quad q(x) = \frac{a_0(x)}{a_2(x)} \exp\left(\int \frac{a_1(x)}{a_2(x)} dx\right).$$

Notice that the function a_1/a_2 is defined and continuous on all of I , since a_2 has no zero in this interval.

Proof. According to the above definitions we have

$$L_{\text{div}} u = p(x)u'' + \frac{a_1(x)p(x)}{a_2(x)}u' + \frac{a_0(x)p(x)}{a_2(x)}u.$$

Due to $p'(x) = a_1(x)p(x)/a_2(x)$ this immediately furnishes the result. \square

While the lemma does show that every second-order linear differential operator can be transformed into an operator in divergence form, it is not immediately clear whether this procedure should make it possible to construct a complete orthonormal set of eigenfunctions — where at the moment we are still assuming that this can be done for symmetric operators. To see the potential complication, consider again the situation of Lemma 2.35. In order to find eigenfunctions φ of the differential operator L , one needs to solve equations of the form

$$L\varphi = \lambda\varphi \quad \text{with} \quad \varphi \in \mathcal{B},$$

where the boundary conditions are encoded in \mathcal{B} . For example, if $p(x) = 1$ and $q(x) = 0$, we get the symmetric operator $Lu = u''$. As we have shown in Example 2.30, for boundary values $u(0) = u(1) = 0$, the resulting eigenvalue problem $Lu = -\lambda u$ has nontrivial

solutions for $\lambda = \ell^2\pi^2/L^2$, and these nontrivial solutions are the elements of the sine family. However, it is not clear if this example is a special case or if it can be generalized. Since L is not symmetric in general, it seems imprudent to expect a complete orthonormal set of eigenfunctions for this equation. Yet, Lemma 2.35 shows that there exists a differential operator L_{div} which is in divergence form, and which satisfies $L_{\text{div}}u = c(x)Lu$ for some function c without zeros in I . Thus, the above eigenvalue problem is equivalent to

$$L_{\text{div}}\varphi = \lambda c(x)\varphi \quad \text{with} \quad \varphi \in \mathcal{B}.$$

Even though the operator on the left-hand side is now symmetric, this equation is no longer an eigenvalue problem in the classical sense — and this fact seems to indicate that for general second-order linear differential operators L one cannot expect a complete orthonormal set of eigenfunctions. The next section shows that this concern is not warranted, and we are able to state conditions for which there is a complete orthonormal set of eigenfunctions and give details on the properties of the eigenvalues and eigenfunctions.

Regular Sturm-Liouville problems

After the preparations of the previous sections we can now turn our attention to the generation of complete orthonormal sets via solving boundary value problems. This will be accomplished within the framework of so-called *regular Sturm-Liouville problems*, which are the subject of the following definition.

Definition 2.36 (Regular Sturm-Liouville problem). Let $I = [a, b] \subset \mathbb{R}$ be a bounded interval, assume that $q, r : I \rightarrow \mathbb{R}$ are two continuous functions, and let $p : I \rightarrow \mathbb{R}$ denote a continuously differentiable function. In addition, suppose that

$$p(x) > 0 \quad \text{and} \quad r(x) > 0 \quad \text{for all} \quad x \in I = [a, b], \quad (2.31)$$

i.e., assume that p and r are strictly positive on all of I . Consider the differential operator L given by

$$(Lu)(x) = \frac{d}{dx} (p(x)u'(x)) + q(x)u(x) \quad (2.32)$$

and the set \mathcal{B} of admissible functions given by

$$\mathcal{B} = \{u \in C^2[a, b] : u \text{ satisfies (2.34) below}\}, \quad (2.33)$$

i.e., functions in \mathcal{B} satisfy the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0, \quad (2.34)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$. Then the boundary value problem

$$Lu + \lambda r(x)u = 0 \quad \text{with} \quad u \in \mathcal{B} \quad (2.35)$$

is called a regular Sturm-Liouville problem. A number λ is called an eigenvalue if there exists a nontrivial solution $\varphi \in \mathcal{B}$ of (2.35), i.e., a solution φ satisfying $\varphi \neq 0$. Any such solution is called an eigenfunction corresponding to the eigenvalue λ .

Notice that regular Sturm-Liouville problems are in fact generalized eigenvalue problems associated with second-order linear differential operators in divergence form. As was pointed out at the end of the last section, any eigenvalue problem of the form

$$a_2(x)u'' + a_1(x)u' + a_0(x)u = \mu u \quad \text{with} \quad u \in \mathcal{B}$$

with continuous functions a_0, a_1 , and a_2 , can be converted into a regular Sturm-Liouville problem, provided the coefficient function a_2 is strictly positive on all of I . For this, one only has to consider the functions p and q from Lemma 2.35 and note that the above eigenvalue problem for μ is equivalent to the problem

$$\underbrace{\frac{d}{dx}(p(x)u')}_{=Lu} + q(x)u + \underbrace{(-\mu)}_{=\lambda} \underbrace{\frac{\exp\left(\int \frac{a_1(x)}{a_2(x)} dx\right)}{a_2(x)}}_{=r(x)} u = 0 \quad \text{with} \quad u \in \mathcal{B},$$

which due to $p(x) > 0$ and $r(x) > 0$ on I is a regular Sturm Liouville problem.

The importance of regular Sturm-Liouville problems stems from the fact that considerable information can be obtained on the number of eigenvalues of such problems, the asymptotic behavior of the eigenvalues, as well as on the properties of the associated eigenfunctions. These results can easily be stated, understood, and applied. We collect them in the following central theorem, and prove some of the easier ones right away. Many of the remaining assertions will be established in Section 2.2.3, but some are beyond the scope of this book.

Theorem 2.37 (Solving regular Sturm-Liouville problems). *We consider the regular Sturm-Liouville problem*

$$\underbrace{\frac{d}{dx}(p(x)u')}_{=Lu} + q(x)u + \lambda r(x)u = 0 \quad \text{for} \quad x \in I = [a, b] \subset \mathbb{R} \quad (2.36)$$

subject to the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0, \quad (2.37)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$, the functions q and r are continuous on I , the function p is continuously differentiable on I , and both p and r are strictly positive on I . Then the following hold.

- (a) The Sturm-Liouville problem (2.36), (2.37) has countably infinite many eigenvalues which are all real, and which can be ordered in the form

$$\lambda_1 < \lambda_2 < \lambda_3 < \dots < \lambda_k < \dots$$

Furthermore, we have $\lambda_k \rightarrow \infty$ as $k \rightarrow \infty$.

- (b) For each eigenvalue λ_k , there exists an eigenfunction $\varphi_k(x)$, which is unique up to scalar multiples.

- (c) For each $k \in \mathbb{N}$, the eigenfunction φ_k has exactly $k - 1$ zeros in the open interval (a, b) . Furthermore, between any two consecutive zeros of the eigenfunction φ_{k+1} there is exactly one zero of φ_k . In other words, the zeros of consecutive eigenfunctions are interlaced.

- (d) For $k \neq \ell$, the eigenfunctions $\varphi_k(x)$ and $\varphi_\ell(x)$ are orthogonal with respect to the weight function $r(x)$, i.e., one has

$$\int_a^b \varphi_k(x)\varphi_\ell(x)r(x)dx = 0 \quad \text{whenever} \quad k \neq \ell.$$

- (e) The eigenfunctions φ_k , for $k \in \mathbb{N}$, form a complete orthogonal set with respect to the weighted scalar product $\langle \cdot, \cdot \rangle_r$, see Definition 2.22.

Proof. The proofs of several of the statements of this theorem require a deeper — yet beautiful — analysis of boundary value problems, and the discussion of these will therefore be deferred to Section 2.2.3 below. At this point, we only prove statements (b) and (d). While the former of these follows immediately from Proposition 2.31, the latter requires a little work.

Let φ and ψ denote eigenfunctions of the Sturm-Liouville problem (2.36), (2.37) with associated eigenvalues λ and μ , respectively, where $\lambda \neq \mu$. Then we have both

$$L\varphi + \lambda r(x)\varphi = 0 \quad \text{and} \quad L\psi + \mu r(x)\psi = 0.$$

This implies, in combination with Theorem 2.34, the identity

$$\begin{aligned} \lambda \langle \varphi, \psi \rangle_r &= \int_a^b \lambda r(x)\varphi(x)\psi(x) dx = - \int_a^b (L\varphi)(x)\psi(x) dx \\ &= -\langle L\varphi, \psi \rangle = -\langle \varphi, L\psi \rangle \\ &= -\int_a^b \varphi(x)(L\psi)(x) dx = \int_a^b \varphi(x)\mu r(x)\psi(x) dx = \mu \langle \varphi, \psi \rangle_r, \end{aligned}$$

which furnishes

$$(\lambda - \mu) \langle \varphi, \psi \rangle_r = 0.$$

Together with $\lambda \neq \mu$ this completes the proof of (d). \square

The importance of Theorem 2.37 cannot be emphasized enough, and we will make use of it throughout this book, even in the context of partial differential equations on higher-dimensional domains. In fact, we immediately state the following useful corollary which can be used to show that a set of functions is a complete orthogonal set without ever having to calculate an inner product. The proof of this corollary follows directly from Theorem 2.37.

Corollary 2.38 (Complete orthogonal sets). *Assume $S = \{\varphi_k\}_{k \in \mathbb{N}}$ is a set of eigenfunctions for the regular Sturm-Liouville problem on $I = [a, b]$ given in Theorem 2.37. Assume that for each $k \in \mathbb{N}$, the function φ_k has $k - 1$ zeros in the open interval (a, b) . Then S is a complete orthogonal set in $L^2(I)$.*

Theorem 2.37 requires the coefficient functions p and r to be strictly positive on the complete interval $I = [a, b]$, including the endpoints. If this assumption is violated, then we call (2.36), (2.37) a *singular Sturm-Liouville problem*. While many of the conclusions of Theorem 2.37 remain essentially unchanged in the singular case, some have to be changed slightly in certain situations, and some are no longer true. The precise formulations and validities of statements for singular Sturm-Liouville problems depend on the type of singular Sturm-Liouville problem, and a comprehensive discussion is therefore beyond the scope of this book.

Periodic boundary conditions

We note that our general form of boundary condition in (2.14) does not include periodic boundary conditions. Nevertheless, Sturm-Liouville problems can be formulated

for periodic boundary conditions as well, if we assume that $p(a) = p(b)$. In this case, the boundary conditions are given by $u(a) = u(b)$ and $u'(a) = u'(b)$. Most of the results given above remain valid in this case, with the notable exception that it is possible for up to two linearly independent eigenfunctions to correspond to the same eigenvalue. In particular, for periodic boundary conditions, the eigenvalues for the problem are still all real and can be ordered in the form

$$\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq \lambda_5 \dots ,$$

with $\lambda_k \rightarrow \infty$ as $k \rightarrow \infty$. Moreover, the corresponding eigenfunctions $\varphi_k(x)$ form a complete orthogonal set.

2.2.2 • Some Concrete Sturm-Liouville Problems

After having presented the general theory of regular Sturm-Liouville problems, we now present a number of examples which will be used throughout the book. In addition to regular Sturm-Liouville problems, we also describe two singular ones.

Examples of regular Sturm-Liouville problems

For the remainder of this section we will illustrate through a sequence of examples how one can find the eigenvalues and eigenfunctions for concrete Sturm-Liouville problems. We will draw upon these examples later on in the book. To begin with, we discuss a number of regular Sturm-Liouville problems.

Example 2.39 (Sine family as Sturm-Liouville eigenfunctions). Consider the family of boundary value problems from Example 2.30

$$u'' + \lambda u = 0, \quad \text{subject to} \quad u(0) = u(L) = 0. \quad (2.38)$$

This example corresponds to a regular Sturm-Liouville problem given in Theorem 2.37 with $p(x) = 1$, $q(x) = 0$, and $r(x) = 1$. We have already solved this problem. To summarize our findings in the context of the Sturm-Liouville theory, the increasing sequence of eigenvalues $\lambda_1 < \lambda_2 < \dots$ of the Sturm-Liouville problem (2.38) are given by

$$\lambda_\ell = \frac{\ell^2\pi^2}{L^2} \quad \text{for } \ell \in \mathbb{N},$$

with associated eigenfunctions

$$\varphi_\ell(x) = \sin \frac{\ell\pi x}{L},$$

the members of the sine family.

Note that for each $\ell \in \mathbb{N}$, the function φ_ℓ has exactly $\ell - 1$ zeroes strictly between 0 and 1, at the points $x = jL/\ell$, for $j = 1, 2, \dots, \ell - 1$. Therefore, by Corollary 2.38, the set of eigenfunctions $\{\varphi_\ell\}_{\ell \in \mathbb{N}}$ is a complete orthogonal set. In other words, the Sturm-Liouville theory has allowed us to conclude that the sine family is a complete orthogonal set, by just counting zeros. ■

For regular Sturm-Liouville problems in which the coefficient functions p , q , and r are constant, one can always obtain all eigenvalues and eigenfunctions, by considering the three cases $\lambda < 0$, $\lambda = 0$, and $\lambda > 0$. In the following examples, for the sake of brevity we do not present the detailed steps of the derivations, but rather a sketch of how to get to the final answers — and leave the details to the reader.

Example 2.40 (Cosine family as Sturm-Liouville eigenfunctions). We now show that the cosine family consists of eigenfunctions for a Sturm-Liouville problem. Consider the Sturm-Liouville problem

$$u'' + \lambda u = 0, \quad \text{subject to} \quad u'(0) = u'(L) = 0, \quad (2.39)$$

which is based on the same differential equation as in the previous example, but with another set of boundary conditions. As before, the boundary value problem has no non-trivial solutions for $\lambda < 0$. For $\lambda = 0$, one obtains constant functions as solutions. Furthermore, for $\lambda > 0$ solutions are of the form

$$u(x) = c_1 \cos(px) \quad \text{with} \quad p = \sqrt{\lambda},$$

and we need $pL = \ell\pi$ with $\ell \in \mathbb{N}$ to satisfy the boundary condition at $x = L$. Thus, the eigenvalues $\lambda_0 < \lambda_1 < \lambda_2 < \dots$ of the Sturm-Liouville problem (2.39) are given by

$$\lambda_\ell = \frac{\ell^2\pi^2}{L^2} \quad \text{for} \quad \ell \in \mathbb{N}_0,$$

with associated eigenfunctions

$$\varphi_\ell(x) = \cos \frac{\ell\pi x}{L}. \quad (\text{Note that } \varphi_0(x) \equiv 1.)$$

This is the cosine family. As in the previous example, it is possible to use the Sturm-Liouville context of the cosine family to show that the cosine family is a complete orthogonal set. See Review Question 2.7.12. Even though the eigenvalues appear to be the same as in the previous example, note that this time we gain the additional eigenvalue $\lambda_0 = 0$ with eigenfunction $\varphi_0(x) = 1$. ■

Example 2.41 (Sturm-Liouville problem with mixed boundary conditions). Also in this example we consider the interval $I = [0, L]$, but this time for the Sturm-Liouville problem

$$u'' + \lambda u = 0, \quad \text{subject to} \quad u'(0) = u(L) = 0. \quad (2.40)$$

One can easily see that this boundary value problem has no solutions for $\lambda \leq 0$. For $\lambda > 0$ the general solution to the differential equation without any boundary conditions is of the form

$$u(x) = c_1 \cos px + c_2 \sin px \quad \text{where} \quad p = \sqrt{\lambda}.$$

Based on the boundary conditions, the solution then has to be of the form

$$u(x) = \gamma \cos(px) \quad \text{for} \quad \gamma \in \mathbb{R},$$

where $pL = (2\ell - 1)\pi/2$, i.e., we need

$$p = \frac{(2\ell - 1)\pi}{2L} \quad \text{for} \quad \ell \in \mathbb{N}.$$

In other words, the eigenvalues $\lambda_1 < \lambda_2 < \lambda_3 < \dots$ of the Sturm-Liouville problem (2.40) are given by

$$\lambda_\ell = \frac{(2\ell - 1)^2\pi^2}{4L^2} \quad \text{for} \quad \ell \in \mathbb{N},$$

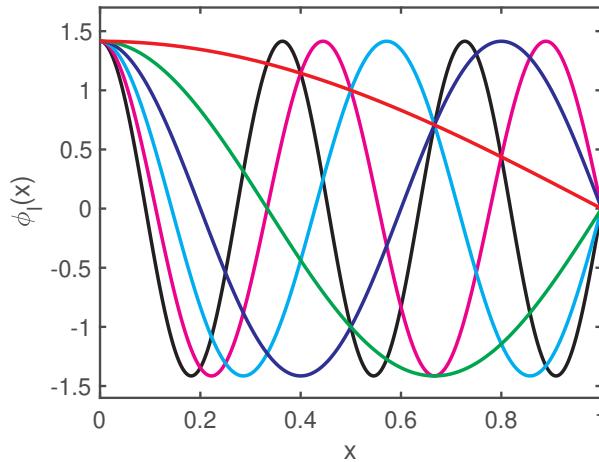


Figure 2.4. Eigenfunctions φ_ℓ for the Sturm-Liouville problem (2.40) in Example 2.41 for the indices $\ell = 1, 2, \dots, 6$ and $L = 1$. The colors of these functions are red, green, blue, cyan, magenta, and black, respectively.

with associated eigenfunctions

$$\varphi_\ell(x) = \cos \frac{(2\ell - 1)\pi x}{2L}.$$

The first six eigenfunctions of this regular Sturm-Liouville problem are shown in Figure 2.4. Notice that the eigenfunction φ_ℓ has exactly $\ell - 1$ zeros in the interval $(0, 1)$, which ensures the completeness of the set. ■

In all of the previous examples, the boundary conditions have been simple enough that we did not need to use the linear-algebraic methods for solving for eigenvalues. The next example shows how to use these methods.

Example 2.42 (Algebraic methods for finding eigenvalues). We consider the Sturm-Liouville problem on the interval $I = [0, 1]$ given by

$$u'' + \lambda u = 0, \quad \text{with} \quad u(0) + u'(0) = 0 \quad \text{and} \quad u(1) + u'(1) = 0. \quad (2.41)$$

In order to find all nontrivial solutions of this boundary value problem, we have to use the methods from Section 2.2.1 and try to solve (2.41) for every $\lambda \in \mathbb{R}$. For most of these λ -values, the problem will only have the trivial solution. However, once we find all λ -values for which nontrivial solutions exist, we obtain the complete set of eigenvalues and associated eigenfunctions. For this, based on the different fundamental sets of solutions of the second-order differential equation in (2.41), we divide our consideration again into the three cases (1) $\lambda < 0$, (2) $\lambda = 0$, and finally (3) $\lambda > 0$.

- (1) If $\lambda < 0$, then the solution to the differential equation without boundary conditions is given by

$$u(x) = c_1 e^{px} + c_2 e^{-px}, \quad \text{where} \quad p = \sqrt{-\lambda}.$$

Thus, arbitrary solutions are a linear combination of the two linearly independent functions $u_1(x) = e^{px}$ and $u_2(x) = e^{-px}$, and we have

$$u'_1(x) = p e^{px}, \quad \text{and} \quad u'_2(x) = -p e^{-px}.$$

Using the matrix in (2.19), there is a nontrivial solution to the Sturm-Liouville problem only if the matrix

$$A = \begin{pmatrix} 1+p & 1-p \\ e^p + pe^p & e^{-p} - pe^{-p} \end{pmatrix}$$

has determinant zero. Since this determinant is given by

$$\det(A) = e^{-p}(1-p^2) - e^p(1-p^2) = (e^{-p} - e^p)(1+\lambda),$$

the only admissible λ -value is $\lambda = -1$, with $p = 1$. Furthermore, the column vector $c = (c_1, c_2)$ has to satisfy the equation $Ac = 0$, where

$$A = \begin{pmatrix} 2 & 0 \\ 2e & 0 \end{pmatrix}.$$

This equation has the solutions $c = (c_1, c_2) = (0, \gamma)$ for $\gamma \in \mathbb{R}$, and therefore solutions are of the form

$$u(x) = \gamma e^{-x}.$$

(2) If $\lambda = 0$, then $u(x) = c_1 x + c_2$. A simple check of the boundary conditions shows that this implies $c_1 = c_2 = 0$. Therefore, $\lambda = 0$ is not an eigenvalue, since it does not yield nontrivial solutions.

(3) If $\lambda > 0$, then the solutions to the differential equation are of the form

$$u(x) = c_1 \cos(px) + c_2 \sin(px), \quad \text{where } p = \sqrt{\lambda}.$$

Thus, solutions are a linear combination of the two linearly independent functions

$$u_1(x) = \cos(px) \quad \text{and} \quad u_2(x) = \sin(px),$$

and therefore

$$u'_1(x) = -p \sin(px) \quad \text{and} \quad u'_2(x) = p \cos(px).$$

Using the matrix in (2.19), there is a nontrivial solution to the problem if and only if the matrix

$$A = \begin{pmatrix} 1 & p \\ \cos p - p \sin p & \sin p + p \cos p \end{pmatrix}$$

has determinant zero. This reduces to

$$\det(A) = \sin p + p \cos p - (p \cos p - p^2 \sin p) = (1 + \lambda) \sin p.$$

This is equal to zero for positive values of λ only when

$$p = \ell\pi \quad \text{and} \quad \lambda = p^2 = \ell^2\pi^2 \quad \text{for} \quad \ell \in \mathbb{N}.$$

Furthermore, $c = (c_1, c_2)$ has to satisfy the equation $Ac = 0$, where

$$A = \begin{pmatrix} 1 & \ell\pi \\ (-1)^\ell & \ell\pi(-1)^\ell \end{pmatrix}.$$

Thus, the components of the solution vector have to satisfy $c_1 + \ell\pi c_2 = 0$, i.e., we have the solutions $c = (c_1, c_2) = \gamma(-\ell\pi, 1)$ for each $\gamma \in \mathbb{R}$. The corresponding solutions of the Sturm-Liouville problem are now of the form

$$u(x) = \gamma(-\ell\pi \cos(\ell\pi x) + \sin(\ell\pi x)), \quad \text{with}$$

$$u'(x) = \gamma(\ell^2\pi^2 \sin(\ell\pi x) + \ell\pi \cos(\ell\pi x)).$$

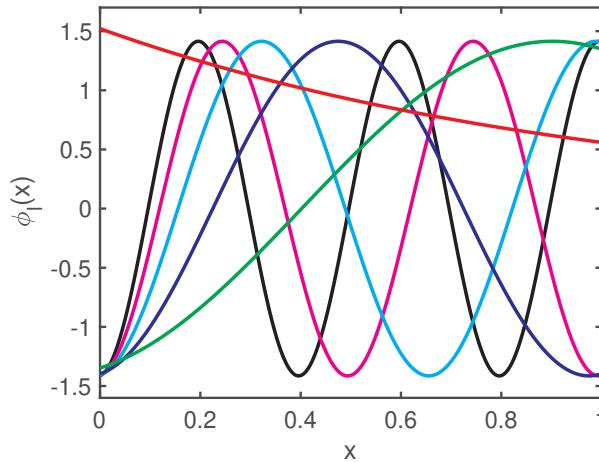


Figure 2.5. Eigenfunctions φ_ℓ for the Sturm-Liouville problem (2.41) in Example 2.42 for the indices $\ell = 0, 1, \dots, 5$. The colors of these functions are red, green, blue, cyan, magenta, and black, respectively.

Altogether, the above considerations show that the eigenvalues¹⁴ $\lambda_0 < \lambda_1 < \lambda_2 < \dots$ of the regular Sturm-Liouville problem (2.41) are given by

$$\lambda_\ell = \begin{cases} -1 & \text{for } \ell = 0, \\ \ell^2\pi^2 & \text{for } \ell \in \mathbb{N}, \end{cases}$$

with associated eigenfunctions

$$\varphi_\ell(x) = \begin{cases} e^{-x} & \text{for } \ell = 0, \\ \sin(\ell\pi x) - \ell\pi \cos(\ell\pi x) & \text{for } \ell \in \mathbb{N}, \end{cases}$$

which form a complete orthogonal set.

The first six eigenfunctions of this regular Sturm-Liouville problem are shown in Figure 2.5. Notice that one can clearly see that the ℓ -th eigenfunction has exactly ℓ zeros in the interval $(0, 1)$. (Recall that we are using indexing for the eigenvalues which starts at zero!) ■

It was mentioned at the end of the previous section that the general form of the boundary condition in (2.14) does not include periodic boundary conditions, but that Sturm-Liouville problems can be formulated in this situation as well. Our last example demonstrates this.

Example 2.43 (Periodic boundary conditions). We now derive the family of eigenfunctions for the periodic Sturm-Liouville problem

$$u'' + \lambda u = 0, \quad \text{subject to} \quad u(0) = u(L) \quad \text{and} \quad u'(0) = u'(L). \quad (2.42)$$

We show that the eigenfunctions for this problem are the Fourier family introduced in Example 2.6. As in the previous example, we distinguish three cases:

¹⁴In Theorem 2.37 the Sturm-Liouville eigenvalues are always denoted by $\lambda_1 < \lambda_2 < \lambda_3 < \dots$. However, in many applications it will be more convenient to start the indexing at 0, or also other values, and we hope that this change in notation will not cause any confusion.

- (1) If $\lambda < 0$, then the solution to the differential equation without boundary conditions is given by

$$u(x) = c_1 e^{px} + c_2 e^{-px}, \quad \text{where } p = \sqrt{-\lambda},$$

and $c_1, c_2 \in \mathbb{R}$. One can easily see that the boundary conditions in (2.42) are satisfied if and only if

$$c_1 + c_2 = c_1 e^{pL} + c_2 e^{-pL} \quad \text{and} \quad pc_1 - pc_2 = c_1 p e^{pL} - c_2 p e^{-pL},$$

and this is equivalent to the system

$$A \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{where } A = \begin{pmatrix} e^{pL} - 1 & e^{-pL} - 1 \\ pe^{pL} - p & -pe^{-pL} + p \end{pmatrix}.$$

Nontrivial solutions only exist if the determinant of A is zero, and this determinant is given by

$$\det(A) = -2p(e^{pL} - 1)(e^{-pL} - 1) > 0,$$

since $p > 0$ and $L > 0$. Thus, there are no nontrivial solutions in this case.

- (2) If $\lambda = 0$, then the general solution is of the form $u(x) = c_1 x + c_2$. A simple check of the boundary condition $u(0) = u(L)$ shows that this implies $c_1 = 0$. Furthermore, one can easily check that for any $c_2 \in \mathbb{R}$ the constant function $u(x) = c_2$ does indeed satisfy the boundary conditions. Therefore, $\lambda = 0$ is an eigenvalue, with respect to constant eigenfunctions.

- (3) If $\lambda > 0$, then the solutions to the differential equation are of the form

$$u(x) = c_1 \cos(px) + c_2 \sin(px), \quad \text{where } p = \sqrt{\lambda},$$

and $c_1, c_2 \in \mathbb{R}$. As before one can see that the boundary conditions in (2.42) are satisfied if and only if

$$c_1 = c_1 \cos(pL) + c_2 \sin(pL) \quad \text{and} \quad pc_2 = -c_1 p \sin(pL) + c_2 p \cos(pL),$$

and this is equivalent to the system

$$A \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{where } A = \begin{pmatrix} -1 + \cos(pL) & \sin(pL) \\ -p \sin(pL) & -p + p \cos(pL) \end{pmatrix}.$$

Nontrivial solutions only exist if the determinant of A is zero, and this determinant is given by

$$\det(A) = p(-1 + \cos(pL))^2 + p \sin^2(pL) = 2p(1 - \cos(pL)).$$

This is equal to zero if and only if

$$p = \frac{2\ell\pi}{L} \quad \text{and} \quad \lambda = \frac{4\ell^2\pi^2}{L^2} \quad \text{for} \quad \ell \in \mathbb{N}.$$

The associated coefficient vectors $c = (c_1, c_2) \in \mathbb{R}^2$ have to satisfy the system of equations $Ac = 0$. Yet, if $\cos(pL) = 1$, then the matrix A equals the zero matrix, and therefore there are no restrictions on c_1 and c_2 . This gives the eigenfunctions

$$u(x) = c_1 \cos \frac{2\ell\pi x}{L} + c_2 \sin \frac{2\ell\pi x}{L}, \quad \text{for all } c_1, c_2 \in \mathbb{R}.$$

In other words, we obtain a two-dimensional space of eigenfunctions, which is spanned by the eigenfunctions $\cos(2\ell\pi x/L)$ and $\sin(2\ell\pi x/L)$.

Combining the above three cases now shows that the periodic Sturm-Liouville problem (2.42) has eigenfunctions and eigenvalues given by

$$\varphi_{2\ell+1}(x) = \cos \frac{2\ell\pi x}{L} \quad \text{and} \quad \lambda_{2\ell+1} = \frac{4\ell^2\pi^2}{L^2}, \quad \text{for } \ell \in \mathbb{N}_0,$$

as well as

$$\varphi_{2\ell}(x) = \sin \frac{2\ell\pi x}{L} \quad \text{and} \quad \lambda_{2\ell} = \frac{4\ell^2\pi^2}{L^2}, \quad \text{for } \ell \in \mathbb{N}.$$

Notice again that in contrast to the previous examples, all of the positive eigenvalues are associated to two linearly independent eigenfunctions. ■

Singular Sturm-Liouville problems

In this section, we consider some singular Sturm-Liouville problems which will be used later in order to solve partial differential equations on certain domains. These examples all have the following singular property: Although $r > 0$ still holds everywhere on the interval I , we now have $p = 0$ at one or both of the endpoints of the interval I . In this case, one can replace the boundary condition at each endpoint where $p = 0$ with an assumption of boundedness of the function and its derivative at that endpoint. Even though our main Theorem 2.37 no longer applies in this situation, these singular Sturm-Liouville problems share many of the properties of their regular counterparts. It would be too cumbersome to state a general result for singular problems, with the exception of the following orthogonality result:

- In the formulation of Theorem 2.37, as long as $r > 0$ for all $a < x < b$, part (d) concerning the orthogonality of eigenfunctions with distinct eigenvalues remains true even if the coefficient function p is zero at one or both of the endpoints, and even if the interval $I = (a, b)$ is unbounded at one or both ends.

This follows by adapting the proofs of Theorems 2.34 and 2.37.

Example 2.44 (Legendre polynomials). The Legendre polynomials are a set of polynomials which form a complete orthogonal set for $L^2(0, 1)$. In this example, we introduce them as eigenfunctions of a Sturm-Liouville problem. We will later show that these functions are useful for solving partial differential equations, as they are one of the angular portions of the solutions for Laplace's equation in spherical coordinates in Section 2.6.3. The Legendre Sturm-Liouville problem is given by

$$\frac{d}{dx} \left((1-x^2) \frac{du}{dx} \right) + \beta u = 0 \quad \text{for } x \in I = [-1, 1].$$

Notice that we have $p(x) = 1 - x^2$, $q(x) = 0$, and $r(x) = 1$. Thus the inequality $r(x) > 0$ holds on the interval I , and $p(x) > 0$ is satisfied for $-1 < x < 1$. Since $p(-1) = p(1) = 0$, this is a singular Sturm-Liouville problem. Thus we prescribe the boundary conditions that u and u' remain bounded at 0 and 1. Since any scalar multiple of a solution is again a solution, we additionally require $u(1) = 1$ for now, even though at the moment this assumption might preclude solutions which vanish at $x = 1$. In fact, we will justify later that this situation cannot occur. For our second boundary condition, we cannot specify a value. We can only require that u and its derivative remain bounded at the remaining endpoint of I given by $x = -1$.

In order to solve this equation, since the coefficient function p depends explicitly on the spatial variable x , we cannot use the same method as in the previous examples — they were all concerned with constant coefficients. Rather, we will now try to construct as many solutions as possible by other means, and then justify afterwards that we have in fact found a complete orthogonal set of functions. Once this has been accomplished, we know that all eigenvalues and eigenfunctions have been determined.

In situations where one has no prior knowledge as to what type of solution should be expected, it is often a good strategy to try to find solutions in terms of power series — this is sometimes referred to as the *method of Frobenius*. That is, we attempt to find power series solutions of the above differential equation which are of the form

$$u(x) = \sum_{k=0}^{\infty} a_k (x-1)^k,$$

where we choose a power series centered at $x = 1$ due to the form of our boundary conditions. Namely, by the boundary condition $u(1) = 1$, we have $a_0 = 1$. Differentiating the series term-by-term, we get

$$\begin{aligned} \frac{d}{dx} \left((1-x^2) \frac{du}{dx} \right) &= \frac{d}{dx} \left((1-x^2) \sum_{k=1}^{\infty} a_k k (x-1)^{k-1} \right) \\ &= \sum_{k=1}^{\infty} a_k k \frac{d}{dx} ((x-1)^{k-1} (1-x^2)) \\ &= \sum_{k=1}^{\infty} a_k k \frac{d}{dx} ((x-1)^{k-1} (x-1+2)) \\ &= -\sum_{k=1}^{\infty} a_k k \left(k(x-1)^{k-1}(x-1+2) + (x-1)^k \right) \\ &= -\sum_{k=1}^{\infty} a_k k \left((k+1)(x-1)^k + 2k(x-1)^{k-1} \right). \end{aligned}$$

Substituting into the differential equation, one obtains

$$\beta a_0 - \sum_{k=1}^{\infty} \left((k^2 + k - \beta) a_k (x-1)^k + 2k^2 a_k (x-1)^{k-1} \right) = 0. \quad (2.43)$$

We can change the summation index to see that

$$\sum_{k=1}^{\infty} 2k^2 a_k (x-1)^{k-1} = \sum_{\ell=0}^{\infty} 2(\ell+1)^2 a_{\ell+1} (x-1)^{\ell}.$$

Thus, by using the indexing variable k instead of ℓ for the right-hand side of this equation, the identity in (2.43) becomes

$$\beta a_0 - 2a_1 - \sum_{k=1}^{\infty} \left((k^2 + k - \beta) a_k (x-1)^k + 2(k+1)^2 a_{k+1} (x-1)^k \right) = 0.$$

By setting $x = 1$, we now conclude $\beta a_0 - 2a_1 = 0$, and this implies $a_1 = \beta a_0 / 2 = \beta / 2$. Therefore

$$\sum_{k=1}^{\infty} c_k (x-1)^k = 0, \quad \text{where} \quad c_k = (k^2 + k - \beta) a_k + 2(k+1)^2 a_{k+1}.$$

We now use a useful fact for working with power series. The only way for a sum of the form $\sum_{k=1}^{\infty} c_k(x-1)^k$ to be zero for every value of $x \in (0, 1)$ is if $c_k = 0$ for all k . That is, for $k \geq 1$, we have

$$(k^2 + k - \beta) a_k + 2(k+1)^2 a_{k+1} = 0,$$

and this leads to the general recursive expression

$$a_{k+1} = -\frac{k^2 + k - \beta}{2(k+1)^2} a_k, \quad (2.44)$$

for all $k \geq 1$. Based on this recursive formula, we now consider a specific set of values for β . Fix an integer $n \in \mathbb{N}_0$, and let $\beta_n = n^2 + n$. Notice that this implies $a_{n+1} = 0$, and in addition that $a_k = 0$ for all $k > n$. In fact, since $k^2 + k < n^2 + n$ for all $k < n$, the recursion (2.44) also shows that the numbers a_0, \dots, a_n are all nonzero. Therefore, for each n , the solution associated with β_n is an n -th degree polynomial, meaning that the solution and its first derivative are bounded at $x = -1$, thereby satisfying the second boundary condition for the problem.

For each fixed $n \in \mathbb{N}_0$, denote the solution of the Legendre problem with $\beta_n = n^2 + n$ by $L_n(x)$. These polynomials are called the *Legendre polynomials*. From the recursive formula for a_k , we can write down the polynomial for any specific n -value. For example, we have

$$L_0(x) = 1,$$

$$L_1(x) = 1 + (x-1) = x,$$

$$L_2(x) = 1 + 3(x-1) + \frac{3}{2}(x-1)^2 = \frac{3x^2 - 1}{2},$$

$$L_3(x) = 1 + 6(x-1) + \frac{15}{2}(x-1)^2 + \frac{5}{2}(x-1)^3 = \frac{5x^3 - 3x}{2},$$

$$L_4(x) = 1 + 10(x-1) + \frac{45}{2}(x-1)^2 + \frac{35}{2}(x-1)^3 + \frac{35}{8}(x-1)^4 = \frac{35x^4 - 30x^2 + 3}{8}.$$

These first five Legendre polynomials are depicted in Figure 2.6.

So far, we have merely assumed that $\beta = n(n+1)$ for $n \in \mathbb{N}_0$, and this does give rise to solutions, but without any indication as to why these β -values are the only possible eigenvalues for the Sturm-Liouville problem. That there are no other β -values follows from the following two facts. First, although this is a singular Sturm-Liouville problem, we still have that $r(x) > 0$ on the interval I , and this implies that all eigenfunctions corresponding to distinct eigenvalues are orthogonal. Second, one can easily see that this family of polynomials has to be complete. According to the Stone-Weierstraß Theorem, a collection of polynomials is dense in $L^2(I)$, if it contains polynomials of every degree — which is clearly satisfied by the Legendre polynomials. Combining these two facts shows that we have already found all possible eigenvalues for this problem, and this also provides a final justification for our normalization assumption $u(1) = 1$.

Despite the singular nature of the boundary value problem, the eigenfunctions of the Legendre Sturm-Liouville problem still satisfy the assertions of Theorem 2.37 concerning their zeros, as can be seen in Figure 2.6. ■

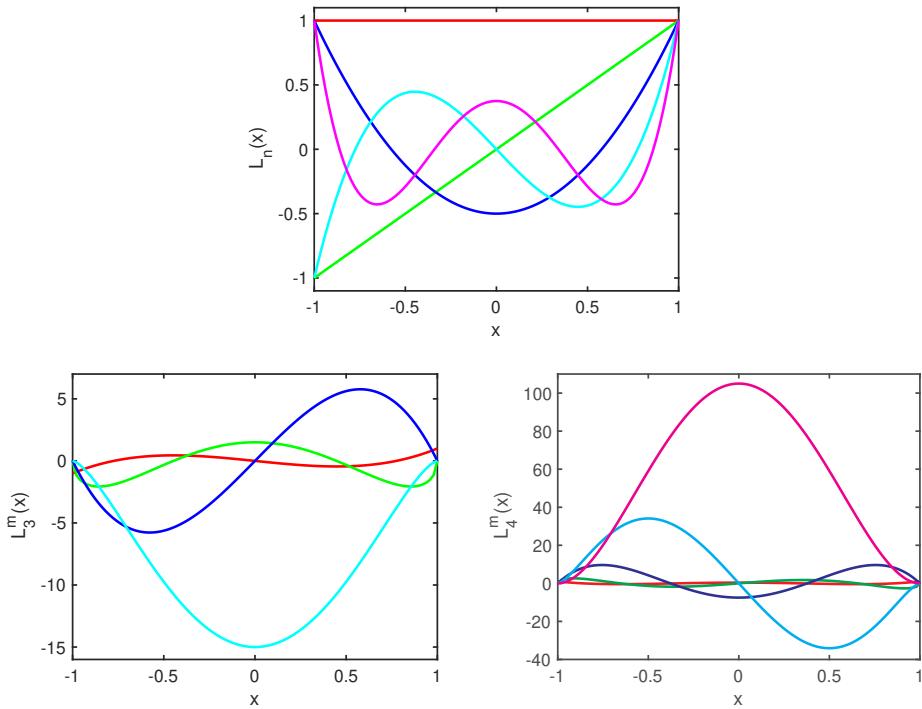


Figure 2.6. The Legendre polynomials and associated Legendre functions. (a) The Legendre polynomials $L_n(x)$ for $n = 0, 1, 2, 3, 4$ in respectively red, green, blue, cyan, and magenta. (b) The associated Legendre functions $L_3^m(x)$ for $m = 0, 1, 2, 3$ in red, green, blue, and cyan, respectively. (c) The associated Legendre functions $L_4^m(x)$ for $m = 0, 1, 2, 3, 4$ in red, green, blue, cyan, and magenta.

Example 2.45 (Associated Legendre functions). A second Sturm-Liouville problem which is closely related to the Legendre problem is given by

$$\frac{d}{dx} \left((1-x^2) \frac{du}{dx} \right) - \frac{m^2}{1-x^2} u + \beta u = 0,$$

where $m \in \mathbb{N}_0$ and $x \in (-1, 1)$. Note that the Legendre polynomial equation in Example 2.44 is the $m = 0$ case of this equation. Again we have $p(x) = 1 - x^2$, and therefore the problem is also a singular Sturm-Liouville problem. We use the same boundary conditions as before, namely $u(1) = 1$ at the right interval endpoint, and the boundedness of u and its derivative at the left endpoint $x = -1$. Using a reparameterization of the previous equation, one can show that nontrivial solutions of this equation occur again for $\beta = n(n+1)$, but this time only for integers $n \geq m$. The solutions $L_n^m(x)$ to this equation are called *associated Legendre functions*. It is possible to show that the associated Legendre functions $L_n^m(x)$ are related to the Legendre polynomials $L_n(x)$ by the equation

$$L_n^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m L_n}{dx^m}(x). \quad (2.45)$$

We can use the Sturm-Liouville orthogonality theorem to show the orthogonality of these functions. Fixing the parameter m and allowing β to vary implies $q(x) = -m^2/(1-x^2)$ and $r(x) = 1 > 0$. Thus we can conclude that the functions L_n^m are mutually orthogonal

in n with constant weight function 1. In addition, it is possible to show using a different choice of q and r that for fixed $n \geq m$, the functions L_n^m are mutually orthogonal in m with respect to the weight function $(1-x^2)^{-1}$. The associated Legendre functions for the indices $n=3$ and $n=4$ are depicted in Figure 2.6(b) and (c). ■

Example 2.46 (Bessel functions). This example introduces Bessel functions as solutions of a Sturm-Liouville problem. We will show in Section 2.6.2 that these functions are the radial piece of the solution of Laplace's equation in cylindrical coordinates. Fix $\ell \in \mathbb{N}_0$. The Bessel Sturm-Liouville problem is given by

$$\frac{d}{dx} \left(x \frac{du_\ell}{dx} \right) - \frac{\ell^2}{x} u_\ell + \lambda x u_\ell = 0 \quad \text{for } x \in I = (0, 1). \quad (2.46)$$

This is a singular Sturm-Liouville problem with $p(x)=x$, $q(x)=-\ell^2/x$, and $r(x)=x$. We have $r(x)>0$ and $p(x)>0$ for all $0 < x < 1$, but at the left interval endpoint $x=0$ both p and r vanish — and q is singular. Since $p(1) \neq 0$ we specify the homogeneous boundary condition $u_\ell(1)=0$ at $x=1$, while we only assume that u and u' are bounded at the left endpoint $x=0$ due to $p(0)=0$.

Under these conditions, the results of Theorem 2.37 turn out to be still valid. Specifically, for each fixed $\ell \in \mathbb{N}_0$, there are infinitely many solutions corresponding to distinct eigenvalues λ . For every fixed ℓ , the set of these solutions form a complete orthogonal set in $L_w^2(0, 1)$ with weight function $w(x)=x$.

For fixed $\ell \in \mathbb{N}_0$, we now solve for the solutions to the Bessel equation by using the Frobenius method. That is, we write the solution as a power series

$$u_\ell(x) = x^n \sum_{k=0}^{\infty} a_k x^k,$$

where n is chosen such that $a_0 \neq 0$, and use the form of the differential equation to solve for the coefficients a_k .

As a first step, we rewrite the Bessel Sturm-Liouville problem. One can easily see that (2.46) is equivalent to the form

$$x^2 \frac{d^2 u_\ell}{dx^2}(x) + x \frac{du_\ell}{dx}(x) + (\lambda x^2 - \ell^2) u_\ell(x) = 0, \quad (2.47)$$

which we will now use for finding its solutions. Evaluating the terms in this differential equation for the series solution, we get

$$x^2 \frac{d^2 u_\ell}{dx^2} = \sum_{k=0}^{\infty} (k+n)(k+n-1) a_k x^{k+n} = \sum_{k=0}^{\infty} ((k+n)^2 - (k+n)) a_k x^{k+n},$$

$$x \frac{du_\ell}{dx} = \sum_{k=0}^{\infty} (k+n) a_k x^{k+n},$$

$$(\lambda x^2 - \ell^2) u_\ell = \sum_{k=0}^{\infty} a_k (\lambda x^2 - \ell^2) x^{k+n} = \sum_{m=2}^{\infty} \lambda a_{m-2} x^{m+n} - \sum_{k=0}^{\infty} \ell^2 a_k x^{k+n}.$$

By using again the indexing variable k instead of m , and then substituting into (2.47), one obtains the identity

$$\sum_{k=0}^{\infty} ((k+n)^2 - \ell^2) a_k x^{k+n} + \sum_{k=2}^{\infty} \lambda a_{k-2} x^{k+n} = 0.$$

This equation is now of the form $\sum_{k=0}^{\infty} c_k x^{k+n} = 0$, and therefore one can deduce $c_k = 0$ for all $k \geq 0$.

To obtain an explicit series representation for the solution, it remains to find the expression for c_k for every $k \in \mathbb{N}_0$, and then use it to recover a_k . For this, recall that we assumed both $a_0 \neq 0$ and $\ell \in \mathbb{N}_0$. Thus, setting $k = 0$ in the above identity yields

$$c_0 = (n^2 - \ell^2)a_0 = 0, \quad \text{which implies } n = \ell.$$

This implies that for $\ell = 0$, $u_\ell(0) = a_0 > 0$, whereas $u_\ell(0) = 0$ for all $\ell > 0$, as can be seen in Figure 2.7. For the next index $k = 1$, we use the fact that $n = \ell > 0$ is an integer, to obtain

$$c_1 = ((1 + \ell)^2 - \ell^2)a_1 = (2\ell + 1)a_1 = 0, \quad \text{which implies } a_1 = 0.$$

For the remaining indices $k \geq 2$, again using the fact that $n = \ell$, one finally obtains

$$c_k = (k^2 + 2k\ell)a_k + \lambda a_{k-2} = 0,$$

and this gives the recursive expression

$$a_k = -\frac{\lambda}{k(2\ell + k)}a_{k-2}. \quad (2.48)$$

Combining this formula with the fact that $a_1 = 0$ shows that the resulting series solution contains only nonzero coefficients a_k for even indices k . Therefore, if ℓ is even then u_ℓ is an even function, and if ℓ is odd, then u_ℓ is an odd function, and in either case it is of the form

$$u_\ell(x) = \sum_{j=0}^{\infty} a_{2j} x^{2j+\ell}.$$

By the recursive formula, the absolute value of the ratio of two successive terms in this series satisfies

$$\frac{|\lambda|x^2}{2j(2\ell+2j)} \rightarrow 0 \quad \text{as } j \rightarrow \infty,$$

and therefore by the ratio test the series converges for all $x \in \mathbb{R}$ and all $\lambda \in \mathbb{R}$. Thus we have satisfied the boundary condition of boundedness at the left interval endpoint $x = 0$. The recursive expression further implies both

$$a_2 = -\frac{\lambda}{2 \cdot (2\ell + 2)}a_0 \quad \text{and} \quad a_4 = \frac{\lambda^2}{4 \cdot 2 \cdot (2\ell + 4)(2\ell + 2)}a_0,$$

and in general one obtains

$$a_{2j} = \frac{(-\lambda)^j}{(2j)(2j-2) \cdot \dots \cdot 4 \cdot 2 \cdot (2\ell + 2j)(2\ell + 2(j-1)) \cdot \dots \cdot (2\ell + 4)(2\ell + 2)} \cdot a_0.$$

After a few simplifications, this finally leads to the formula

$$u_\ell(x) = a_0 \sum_{j=0}^{\infty} \frac{(-\lambda)^j \ell!}{2^{2j} j!(\ell+j)!} x^{2j+\ell}.$$

Notice that since the differential equation is linear, the value of the coefficient a_0 can be chosen arbitrarily. For any $\lambda \in \mathbb{R}$, this series for u_ℓ solves the differential equation (2.46),

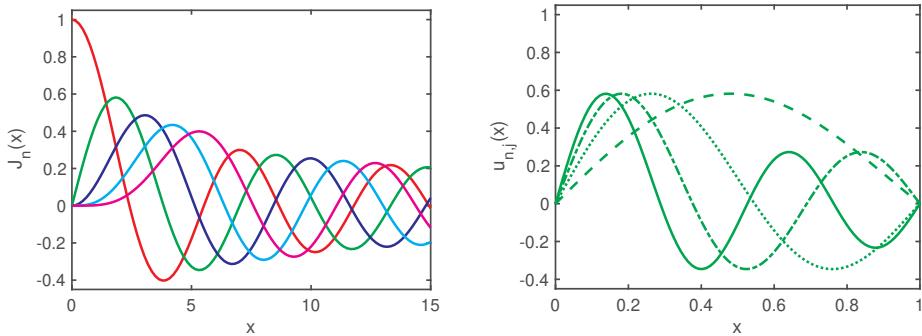


Figure 2.7. (a) The Bessel functions $J_n(x)$ for $n = 0, 1, 2, 3, 4$. The colors of these functions are respectively red, green, blue, cyan, and magenta. (b) The first four solutions $u_{1,j}$ given in (2.50) for the Bessel Sturm-Liouville problem (2.46). These solutions are all rescaled versions of $J_1(x)$, the green function in (a).

but for most λ -values, the resulting function u_ℓ does not satisfy the second boundary condition $u_\ell(1) = 0$.

We now show that there are infinitely many λ -values $\lambda_{\ell,1} < \lambda_{\ell,2} < \dots \rightarrow \infty$ for which the constraint $u_\ell(1) = 0$ is in fact satisfied. For this, we introduce a particular family of functions J_ℓ which are defined on the whole real line, and show that every solution u_ℓ is a rescaled version $J_\ell(\nu_{\ell,j}x)$ on the interval $(0, 1)$. The scaling factors $\nu_{\ell,j}$ are chosen in such a way that $u_\ell(1) = 0$, which is equivalent to $J_\ell(\nu_{\ell,j}) = 0$. At first sight, this rescaling may seem odd, but in fact we performed the same process of rescaling a function on \mathbb{R} in our earlier examples; in those cases, however, the solutions to the Sturm-Liouville problem were rescaled versions of sines and cosines. The difference here is that the zeroes of sines and cosines are evenly spaced, whereas the zeroes of J_ℓ are not. The functions J_ℓ are defined by the series

$$J_\ell(x) = \sum_{j=0}^{\infty} \frac{(-1)^j}{2^{2j+\ell} j!(\ell+j)!} x^{2j+\ell}. \quad (2.49)$$

These functions are called the *Bessel functions of the first kind*. For the indices $\ell = 0, \dots, 4$ they are depicted in Figure 2.7(a).

Bessel functions have been studied extensively over the years, and in certain branches of mathematics their importance is similar to the standard trigonometric functions sine and cosine. In fact, they do share many similarities. One of these is related to their zeros, and while its proof is beyond the scope of this book, it can be stated as follows.

- For any integer $m \in \mathbb{N}$ and any point $m\pi < x < (m + 1/2)\pi$ the Bessel function value $J_0(x)$ is positive for even m and negative for odd values of m . In other words, the Bessel function J_0 has infinitely many zeros on \mathbb{R}_0^+ due to its continuity and the intermediate value theorem, and the distance between any two consecutive zeros lies in the interval $[\pi/2, 3\pi/2]$.

Notice that the definition of the Bessel function J_ℓ given in (2.49) implies the identity

$$\frac{d}{dx} (x^{-\ell} J_\ell(x)) = -x^{-\ell} J_{\ell+1}(x) \quad \text{for all } x \in \mathbb{R}^+ \quad \text{and } \ell \in \mathbb{N}_0.$$

Combined with the above statement about the zeros of J_0 and Rolle's theorem, one then obtains that every Bessel function J_ℓ has infinitely many zeros on \mathbb{R}_0^+ , for all $\ell \in \mathbb{N}_0$. For

Bessel function zeroes $\nu_{\ell,j}$						
j	$J_0(x)$	$J_1(x)$	$J_2(x)$	$J_3(x)$	$J_4(x)$	$J_5(x)$
1	2.4048	3.8317	5.1356	6.3802	7.5883	8.7715
2	5.5201	7.0156	8.4172	9.7610	11.0647	12.3386
3	8.6537	10.1735	11.6198	13.0152	14.3725	15.7002
4	11.7915	13.3237	14.7960	16.2235	17.6160	18.9801
5	14.9309	16.4706	17.9598	19.4094	20.8269	22.2178

Table 2.1. The first five zeroes $\nu_{\ell,j}$ for the Bessel function $J_\ell(x)$ computed in Matlab using the built-in Bessel function.

later reference, we denote the j -th strictly positive zero of J_ℓ by $\nu_{\ell,j}$. These zeros will be used in the remainder of this example to solve the Bessel Sturm-Liouville problem, and some of them are listed in Table 2.1.

We now establish the scaling relationship between the solutions u_ℓ and the Bessel functions J_ℓ . If one defines the arbitrary constant a_0 by $a_0 = \lambda^{\ell/2}/(2^\ell \ell!)$. Then one can immediately verify the identity

$$u_\ell(x) = J_\ell(\sqrt{\lambda}x).$$

For each zero $\nu_{\ell,j}$ of J_ℓ , the function above satisfies the boundary condition $u_\ell(1) = 0$ if we choose

$$\lambda_{\ell,j} = \nu_{\ell,j}^2, \quad u_{\ell,j}(x) = J_\ell(\sqrt{\lambda_{\ell,j}}x) = J_\ell(\nu_{\ell,j}x), \quad \text{for } j \in \mathbb{N}. \quad (2.50)$$

This formula gives the complete orthogonal set of solutions of the Bessel Sturm-Liouville problem, which are orthogonal with respect to the weight function $r(x) = x$. The first five solutions for $\ell = 1$ are shown in Figure 2.7(b). Table 2.1 contains the first five zeroes $\nu_{\ell,j}$ for $\ell = 0, \dots, 5$. Both the figure and the table are calculated in Matlab using the built-in command `besselj` for determining the Bessel functions of the first kind. Observe that these values are ascending as both j and ℓ increase. Therefore, the smallest possible value of the eigenvalue $\lambda_{\ell,j}$ is given by $\lambda_{0,1} \approx 2.4048^2$. ■

We introduce several other examples of Sturm-Liouville problems in the problem sets, each giving rise to a complete orthogonal set of functions on $L_w^2(I)$ for some interval I and weight function w . In particular, Problem 2.7.10 introduces the *Hermite polynomials* as solutions to a Sturm-Liouville problem, Problem 2.7.11 gives the Sturm-Liouville problem with the *Chebyshev polynomials* as solutions, and Problem 2.7.12 introduces the *Laguerre polynomials* as solutions to a Sturm-Liouville problem.

General boundary value problems

Sturm-Liouville problems are not the only boundary value problems associated with second-order linear differential equations. If the boundary conditions are not of the specific form required, it is still possible to obtain explicit solutions. We illustrate this with the following example.

Example 2.47. Consider the boundary value problem

$$u'' = k^2 u, \quad \text{subject to} \quad u(0) = u_0, \quad u(L) = u_L.$$

This is not a regular Sturm-Liouville problem since the boundary values are nonhomogeneous. The differential equation without boundary conditions has the general solution

$$u(x) = Ae^{kx} + Be^{-kx},$$

and due to the boundary conditions one needs to solve

$$A + B = u_0 \quad \text{and} \quad Ae^{kL} + Be^{-kL} = u_L,$$

which can always be achieved for any $k > 0$. Notice that the existence of solutions for every value of k could not happen in the case of a Sturm-Liouville problem with homogeneous boundary conditions. In the homogeneous case we are guaranteed to have a discrete set of eigenvalues. ■

Situations such as the one described in the above example will be discussed more extensively in the following sections, in the context of boundary value problems for partial differential equations.

2.2.3 ■ A Deeper Look: Comparison and Oscillations

In this final section on Sturm-Liouville problems we revisit the central Theorem 2.37 and establish most of its remaining assertions. In the course of this, we introduce two extremely useful concepts. On the one hand, it will be demonstrated that certain second-order boundary value problems can be studied via planar ordinary differential equations through a mechanism called *phase plane analysis*. This approach can in fact be applied successfully to a number of partial differential equations on one-dimensional domains, even certain nonlinear ones. On the other hand, we will see that it is often advantageous to transform a given problem into a *new and not necessarily Cartesian coordinate system*, which capitalizes on central aspects of the problem. A much more extensive treatment of the ideas presented in this section can be found in [15].

Reformulation as a first-order initial value problem

To illustrate our ideas, we return to Example 2.42 from the last section, but this time discuss it from a different perspective.

Example 2.48. Consider the Sturm-Liouville problem on the interval $I = [0, 1]$ given by

$$u'' + \lambda u = 0, \quad \text{with} \quad u(0) + u'(0) = 0 \quad \text{and} \quad u(1) + u'(1) = 0. \quad (2.51)$$

It was shown in Example 2.42 that this problem has infinitely many eigenvalues λ_k , where we have $k \in \mathbb{N}_0$. This was accomplished by using the general form of the solution set of the differential equation part, and then solving a system of two equations in order to satisfy the boundary conditions. In the course of this, basically as a side product, we deduced conditions on the set of λ -values for which nontrivial solutions exist.

In contrast, we now consider the problem (2.51) by only dropping the boundary condition at the right endpoint of the interval I while keeping the one at the left endpoint, i.e., we seek solutions of $u'' + \lambda u = 0$ which satisfy only $u(0) + u'(0) = 0$. Since our interest is in nontrivial solutions only, and any nonzero scalar multiple of a nontrivial solution is again a nontrivial solution of $u'' + \lambda u = 0$ with $u(0) + u'(0) = 0$, we can without loss of generality study the associated *initial value problem*

$$u'' + \lambda u = 0, \quad \text{with} \quad u(0) = 1 \quad \text{and} \quad u'(0) = -1. \quad (2.52)$$

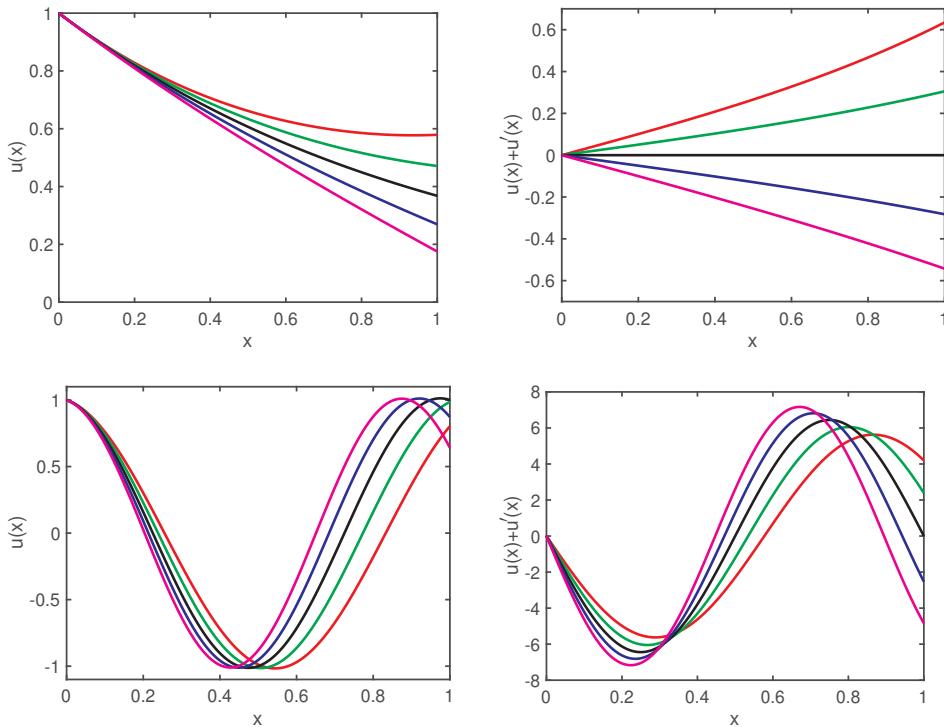


Figure 2.8. Solutions of the initial value problem (2.52) for a variety of different λ -values. The left column shows solutions $u(x)$, and on the right the functions $u(x) + u'(x)$ are shown. The figures in the top row correspond to the λ -values $-1.5, -1.25, -1, -0.75$, and -0.5 which are shown in red, green, black, blue, and magenta, respectively. The bottom row corresponds to the λ -values $3\pi^2, 3.5\pi^2, 4\pi^2, 4.5\pi^2$, and $5\pi^2$, again shown in red, green, black, blue, and magenta, respectively.

This initial value problem has a unique solution for every $\lambda \in \mathbb{R}$, and some of these solutions are shown in the left column of Figure 2.8. In general, however, these solutions do not satisfy the boundary condition $u(1) + u'(1) = 0$. This can be seen in the right column of Figure 2.8, where $u + u'$ is plotted as a function of x , for the solutions shown in the left column. From these images it is clear that only for $\lambda = -1$ and $\lambda = 4\pi^2$ (both in black) one actually obtains solutions for the boundary value problem (2.51). ■

The above example suggests that it might be useful to approach the study of Sturm-Liouville problems by considering a parameter-dependent family of initial value problems. Based on the formulation of the problem, it seems reasonable to expect that the solutions do depend continuously on changes in the parameter λ , and varying this parameter might then force the validity of the second boundary condition for a discrete set of λ -values.

Yet, the graphs of Figure 2.8 have one obvious disadvantage: It is basically impossible to quickly discern from the images of the actual solutions on the left which of these do satisfy the second boundary condition in (2.51). To overcome this complication, we can use a trick from elementary ordinary differential equations — the fact that every higher-order ordinary differential equation can be converted into a system of first-order ordinary differential equation. See also (1.5) and (1.6) in Section 1.1.1.

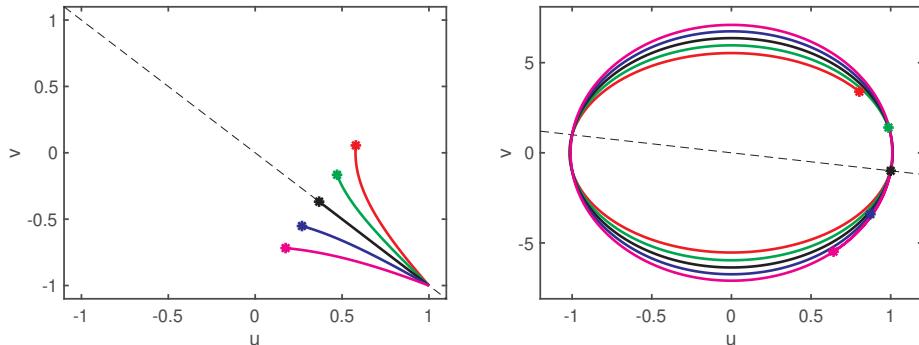


Figure 2.9. A phase diagram for solutions of the initial value problem (2.53) for a variety of different λ -values. The left image shows the solutions corresponding to the λ -values $-1.5, -1.25, -1, -0.75$, and -0.5 , which are shown in red, green, black, blue, and magenta, respectively. The graphs in the right image correspond to the λ -values $3\pi^2, 3.5\pi^2, 4\pi^2, 4.5\pi^2$, and $5\pi^2$, again shown in red, green, black, blue, and magenta, respectively. The colored stars mark the location of the endpoints $(u(1), v(1))$ for each of the solutions.

Example 2.49. We continue our study of Example 2.48. By introducing the new variable $v = u'$, one can equivalently rewrite the initial value problem (2.52) as

$$\begin{aligned} u' &= v, & u(0) &= 1, \\ v' &= -\lambda u, & v(0) &= -1. \end{aligned} \quad (2.53)$$

A variety of solutions of this initial value problem are shown in Figure 2.9. Rather than producing three-dimensional images for the graphs of the functions $x \mapsto (u(x), v(x))$, these images only show the projection of the solutions into the u - v -plane. In other words, we only show the *orbits*

$$\{(u(x), v(x)) \in \mathbb{R}^2 : x \in [0, 1]\}.$$

The resulting image is called a *phase diagram*, and the Euclidean space \mathbb{R}^2 in this context is referred to as the *phase plane*.

The main advantage of the introduction of the phase plane interpretation lies in the fact that satisfying the boundary condition can now be reformulated as a simple geometric condition. Notice that a function $u(x)$ satisfies

$$\begin{aligned} u(\xi) + u'(\xi) &= 0 \quad \text{if and only if} \quad u(\xi) + v(\xi) = 0 \quad \text{if and only if} \\ (u(\xi), v(\xi)) &\in \mathcal{L} = \{(w_1, w_2) \in \mathbb{R}^2 : w_1 + w_2 = 0\}, \end{aligned}$$

i.e., the function u satisfies the boundary value problem 2.51 if and only if the associated solution (u, v) originates and ends exactly on the line \mathcal{L} . Thus, it can easily be seen in the images of Figure 2.9 that only the black curves lead to actual solutions of the boundary value problem.

In order to gain more intuition, Figure 2.10 depicts the dependence of the solutions of (2.53) for a larger number of λ -values. In the left image of this figure, one can see solutions of the initial value problem (2.53) for forty λ -values between -2 and 35 . These values are spaced quadratically in the sense that the ℓ -th value is given by $-2 + 37(\ell/39)^2$ for $\ell = 0, \dots, 39$. For each of these orbits, the endpoint $(u(1), v(1))$ is represented by a blue

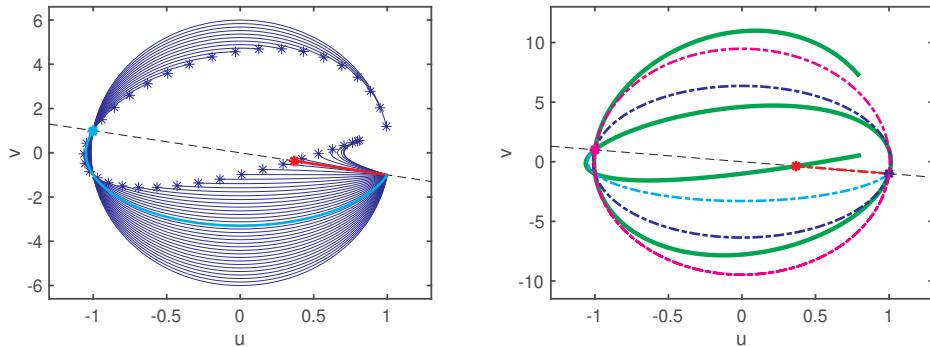


Figure 2.10. The left image shows solutions of the initial value problem (2.53) for forty λ -values between -2 and 35 . In each case, the endpoint $(u(1), v(1))$ is indicated by a blue star. The red and cyan curves correspond to the two λ -values in this range which actually lead to solutions of the boundary value problem (2.51), namely for $\lambda = -1$ and $\lambda = \pi^2$. In the right image, the curve which is traced out by the endpoints $(u(1), v(1))$ for increasing λ is shown in green, where in this case λ -values vary between -2 and 140 . One can see that this curve spirals clockwise around the origin. Every intersection with the line \mathcal{L} gives rise to a solution of the boundary value problem (2.51), and the solutions corresponding to $\lambda = -1, \pi^2, 4\pi^2$, and $9\pi^2$ are shown in red, cyan, blue, and magenta, respectively.

star. The red and cyan curves in the image correspond to the two λ -values in this range which actually furnish solutions of the boundary value problem (2.51), namely $\lambda = -1$ and $\lambda = \pi^2$. In the right image of Figure 2.10, the curve which is traced out by the endpoints $(u(1), v(1))$ for increasing λ is shown in green. This curve spirals clockwise around the origin and seemingly leads to infinitely many intersections with the line \mathcal{L} . Every such intersection in turn leads to a solution of the boundary value problem (2.51). The solutions for the λ -values $\lambda = -1, \pi^2, 4\pi^2$, and $9\pi^2$ are shown in red, cyan, blue, and magenta, respectively. Notice that the orbit for the last of these actually wraps around the origin one and a half times. ■

The above example makes it clear that in order to understand the sequence of solutions of the Sturm-Liouville problem (2.51) one has to be able to accurately describe the behavior of the green curve in the right image of Figure 2.10. And due to its rotational nature, it seems prudent to perform this study in a different coordinate system.

Transformation into polar coordinates

We now turn our attention to the study of general second-order linear boundary value problems of the form

$$\frac{d}{dx} (p(x)u') + f(x)u = 0 \quad \text{for } x \in I = [a, b] \subset \mathbb{R} \quad (2.54)$$

subject to the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0. \quad (2.55)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$, and with a continuous function f and a continuously differentiable function p , which satisfies $p(x) > 0$ for all $x \in I$. In the following, we will demonstrate how both the equation and the boundary conditions can be reformulated using polar coordinates.

We begin by transforming the differential equation (2.54) into polar coordinates. For this, we first have to transform the scalar second-order differential equation into a first-order system, similarly to the above example. Only then can we introduce radial and angular variables. Due to the specific form of the differential equation, namely its divergence form, we write the equation as a system using the new variable $v = p(x)u'$. Thus, the differential equation (2.54) is equivalent to the system

$$\begin{aligned} u' &= \frac{1}{p(x)} v, & u(a) &= u_0, \\ v' &= -f(x) u, & v(a) &= v_0. \end{aligned} \quad (2.56)$$

One can easily see that any solution of this system with initial condition $(u_0, v_0) \neq (0, 0)$ satisfies $(u(x), v(x)) \neq (0, 0)$ for all $x \in [a, b]$.¹⁵ It is therefore possible to write the solution in polar coordinates by defining

$$u = \rho \sin \theta \quad \text{and} \quad v = \rho \cos \theta. \quad (2.57)$$

In order to rewrite the differential equation in terms of the new variables $\rho > 0$ and $\theta \in \mathbb{R}$, we differentiate both equations in (2.57). This furnishes

$$u' = \rho' \sin \theta + \rho \theta' \cos \theta \quad \text{and} \quad v' = \rho' \cos \theta - \rho \theta' \sin \theta,$$

which can be written in matrix form as

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \sin \theta & \rho \cos \theta \\ \cos \theta & -\rho \sin \theta \end{pmatrix} \begin{pmatrix} \rho' \\ \theta' \end{pmatrix}.$$

After inverting the coefficient matrix one then obtains

$$\begin{pmatrix} \rho' \\ \theta' \end{pmatrix} = \begin{pmatrix} \sin \theta & \rho \cos \theta \\ \cos \theta & -\rho \sin \theta \end{pmatrix}^{-1} \begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \sin \theta & \cos \theta \\ \frac{\cos \theta}{\rho} & -\frac{\sin \theta}{\rho} \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix}.$$

Plugging these expressions into the first-order system (2.56) then yields the following system in polar coordinates,

$$\begin{aligned} \rho' &= \left(\frac{1}{p(x)} - f(x) \right) \rho \sin \theta \cos \theta, & \rho(a) &= \rho_0, \\ \theta' &= \frac{1}{p(x)} \cos^2 \theta + f(x) \sin^2 \theta, & \theta(a) &= \theta_0. \end{aligned} \quad (2.58)$$

which is equivalent to (2.54) via (2.57).

We now turn our attention to the boundary conditions (2.55). Since both sides of these identities can be multiplied by a nonzero number without changing the condition, we can adjust $\alpha_1, \alpha_2, \beta_1$, and β_2 such that

$$\alpha_1^2 + \frac{\alpha_2^2}{p(a)^2} = 1 \quad \text{and} \quad \beta_1^2 + \frac{\beta_2^2}{p(b)^2} = 1, \quad \text{as well as} \quad \alpha_2 \leq 0 \quad \text{and} \quad \beta_2 \leq 0. \quad (2.59)$$

¹⁵Note that the right-hand side of (2.56) satisfies the assumptions of the Picard-Lindelöf Theorem 1.10, and that therefore for any $c \in [a, b]$ the trivial solution is the unique solution with $u(c) = v(c) = 0$. Thus, if a solution (u, v) passes through the origin at any time, then it has to be identically zero.

Furthermore, if $\alpha_2 = 0$, then we set $\alpha_1 = 1$, and similarly we set $\beta_1 = 1$ for $\beta_2 = 0$.

The above normalization allows us to rephrase the boundary conditions in terms of the angular variable θ . To see this, note that if σ and τ are arbitrary real numbers which satisfy $\sigma^2 + \tau^2 = 1$ and $\tau \leq 0$, such that $\sigma = 1$ if $\tau = 0$, then the point (σ, τ) lies on the lower half of the unit circle in the plane \mathbb{R}^2 . For any such (σ, τ) , there exists an angle $\gamma \in [0, \pi)$ such that both $\sigma = \cos \gamma$ and $\tau = -\sin \gamma$ are satisfied. Applying this reasoning to each of the two boundary conditions in (2.55) shows that they can be reformulated equivalently as

$$u(a)\cos \alpha - u'(a)p(a)\sin \alpha = 0 \quad \text{and} \quad u(b)\cos \beta - u'(b)p(b)\sin \beta = 0,$$

where $\alpha, \beta \in [0, \pi)$, where

$$\alpha_1 = \cos \alpha \quad \text{and} \quad \alpha_2 = -p(a)\sin \alpha,$$

and a similar relationship for β_1, β_2 , and β . This can be combined into the single equation

$$\tan \alpha = -\frac{\alpha_2}{\alpha_1 p(a)} \quad \text{with} \quad 0 \leq \alpha < \pi,$$

where we assume $\tan(\pi/2) = \infty$. Notice that for this last formula, the normalization assumption made in (2.59) is no longer necessary. An analogous formula can be obtained for the boundary condition at $x = b$. If we now substitute the definitions for the polar variables ρ and θ in terms of u and $v = p(x)u'$ into these equations and use trigonometric angle formulas, we obtain after some simplifications

$$\sin(\theta(a) - \alpha) = 0 \quad \text{and} \quad \sin(\theta(b) - \beta) = 0.$$

Therefore $\theta(a) - \alpha = n\pi$ for some integer n , and $\theta(b) - \beta = m\pi$ for some integer m . We will assume that $n = m = 0$. That is, we set up the initial angles in our polar coordinate representations such that $\theta(a)$ and $\theta(b)$ are contained in $[0, \pi)$. Thus, we have shown that the boundary conditions in (2.55) are equivalent to boundary conditions on the angular variable θ . For easy reference later on, we summarize our findings in the following lemma.

Lemma 2.50 (Boundary value problem in polar coordinates). *Let $I = [a, b] \subset \mathbb{R}$ denote a bounded interval, let $f : I \rightarrow \mathbb{R}$ be continuous, and assume that $p : I \rightarrow \mathbb{R}$ is continuously differentiable with $p(x) > 0$ for all $x \in I$. Consider a second-order linear differential equation of the form*

$$\frac{d}{dx}(p(x)u') + f(x)u = 0 \quad \text{for} \quad x \in I = [a, b] \subset \mathbb{R}, \quad (2.60)$$

and subject to the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0. \quad (2.61)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$. If we then introduce new variables $\rho > 0$ and $\theta \in \mathbb{R}$ such that

$$u(x) = \rho(x)\sin(\theta(x)) \quad \text{and} \quad p(x)u'(x) = \rho(x)\cos(\theta(x)) \quad (2.62)$$

for all $x \in I$, then the boundary value problem (2.60), (2.61) is equivalent to the boundary value problem

$$\begin{aligned} \rho' &= \left(\frac{1}{p(x)} - f(x) \right) \rho \sin \theta \cos \theta, \\ \theta' &= \frac{1}{p(x)} \cos^2 \theta + f(x) \sin^2 \theta, \end{aligned} \quad (2.63)$$

subject to

$$\theta(a) = \alpha \quad \text{and} \quad \theta(b) = \beta, \quad (2.64)$$

where $\alpha, \beta \in [0, \pi)$ are chosen such that

$$\tan \alpha = -\frac{\alpha_2}{\alpha_1 p(a)} \quad \text{and} \quad \tan \beta = -\frac{\beta_2}{\beta_1 p(b)}. \quad (2.65)$$

Here we use the convention $\tan(\pi/2) = \infty$.

In addition, a function $u : I \rightarrow \mathbb{R}$ is a nontrivial solution of the differential equation (2.60) subject to the initial condition $\alpha_1 u(a) + \alpha_2 u'(a) = 0$ if and only if ρ and θ satisfy (2.63) as well as $\theta(a) = \alpha$.

Using this new coordinate system, we can now study Sturm-Liouville problems in more detail. In several examples, we have observed that the solution to a simple second-order linear differential equation is sinusoidal and thus oscillatory. Our first result quantifies this oscillation of solutions in terms of our polar coordinate variable θ defined above. We then present comparison principles, which allow us to use the oscillatory behavior of the simple system to show the oscillatory behavior of all Sturm-Liouville problems.

Lemma 2.51. Let $[a, b] \subset \mathbb{R}$ denote a bounded interval, let $\mu \geq 1$ denote a constant, and let $u_\mu : [a, b] \rightarrow \mathbb{R}$ be a nontrivial solution of

$$u'' + \mu u = 0 \quad \text{on } [a, b] \quad \text{with} \quad u(a) = 0 \quad \text{and} \quad u'(a) > 0.$$

Let θ_μ denote the angular variable associated with u_μ as defined in (2.62) such that $\theta_\mu(a) = 0$. Then we have

$$\theta_\mu(b) > \sqrt{\mu}(b-a) - \pi \quad \text{for all} \quad \mu \geq 1.$$

Proof. We start by stating that every solution to this equation is of the form

$$u_\mu(x) = c \sin(\sqrt{\mu}(x-a)) \quad \text{for all } x \in [a, b], c > 0.$$

Therefore $u_\mu(x) = 0$ if and only if $x = a + k\pi/\sqrt{\mu}$, where k is an integer.

According to Lemma 2.50 with $\alpha_1 = 1, \alpha_2 = 0, p(x) = 1, f(x) = \mu$, the angular variable satisfies the initial value problem

$$\theta' = \cos^2 \theta + \mu \sin^2 \theta = 1 + (\mu - 1) \sin^2 \theta \quad \text{with} \quad \theta(a) = 0.$$

Due to $\mu \geq 1$, the solutions θ_μ satisfy $\theta'_\mu \geq 1$ on $[a, b]$, hence they are strictly increasing. Furthermore, in view of (2.62) the function u_μ has a zero at $x \in [a, b]$ if and only if $\theta_\mu(x) = k\pi$ for some integer k . Now choose the integer K in such a way that

$$0 \leq K \leq \frac{\sqrt{\mu}(b-a)}{\pi} < K+1,$$

and let

$$x_k = a + \frac{k\pi}{\sqrt{\mu}} \quad \text{for } k = 1, \dots, K.$$

Then we have $a < x_k \leq b$ as well as $u_\mu(x_k) = 0$ for all $k = 1, \dots, K$. Thus, the function u_μ has K zeros in the interval $(a, b]$, which implies that the angular variable θ_μ has to intersect the lines $\theta = k\pi$ for all values $k = 1, \dots, K$. This yields

$$\theta_\mu(b) \geq K\pi > \left(\frac{\sqrt{\mu}(b-a)}{\pi} - 1 \right)\pi = \sqrt{\mu}(b-a) - \pi,$$

and the proof is complete. \square

Two comparison theorems

It is clear from our discussion in Examples 2.48 and 2.49 that in order to address the question of eigenfunctions of Sturm-Liouville problems, we have to understand how changes in the function f and p in (2.60) affect the solution u of the associated initial value problem. The central results in this direction are two comparison theorems. The first of these shows that if we increase the function f or if we decrease the function p uniformly over the interval $I = [a, b]$, then the angular variable θ associated with the solution will increase. More precisely, we have the following result.

Theorem 2.52 (First Sturm-Liouville comparison theorem). *Let $I = [a, b] \subset \mathbb{R}$ denote a bounded interval, let $f_1, f_2 : I \rightarrow \mathbb{R}$ be two continuous functions, and assume that the two functions $p_1, p_2 : I \rightarrow \mathbb{R}$ are continuously differentiable with $p_1(x) > 0$ and $p_2(x) > 0$ for all points $x \in I$, as well as*

$$p_1(x) \geq p_2(x) \quad \text{and} \quad f_1(x) \leq f_2(x) \quad \text{for all } a \leq x \leq b. \quad (2.66)$$

In addition, for $k = 1, 2$ let $u_k : I \rightarrow \mathbb{R}$ denote a nontrivial solution of

$$\frac{d}{dx} (p_k(x)u') + f_k(x)u = 0 \quad \text{for } x \in I = [a, b] \subset \mathbb{R}. \quad (2.67)$$

Let θ_1 and θ_2 denote the angular variables associated with u_1 and u_2 , respectively, as defined in (2.62). Finally, assume that

$$\theta_1(a) \leq \theta_2(a). \quad (2.68)$$

Then we have

$$\theta_1(x) \leq \theta_2(x) \quad \text{for all } a \leq x \leq b. \quad (2.69)$$

Furthermore, if we replace the inequality involving f_1 and f_2 in (2.66) by $f_1(x) < f_2(x)$ for all $a < x < b$, then we even have $\theta_1(x) < \theta_2(x)$ for all $a < x \leq b$.

Proof. According to Lemma 2.50 the angular variables θ_1 and θ_2 satisfy the differential equations

$$\theta'_1 = \frac{1}{p_1(x)} \cos^2 \theta_1 + f_1(x) \sin^2 \theta_1 \quad \text{and} \quad \theta'_2 = \frac{1}{p_2(x)} \cos^2 \theta_2 + f_2(x) \sin^2 \theta_2. \quad (2.70)$$

Taking the difference of these two equations, combined with the identity $\cos^2 \theta_k = 1 -$

$\sin^2 \theta_k$, yields the identity

$$\begin{aligned}
\theta'_2 - \theta'_1 &= f_2 \sin^2 \theta_2 - f_1 \sin^2 \theta_1 + \left(\frac{1}{p_2} - \frac{1}{p_1} \right) \cos^2 \theta_2 + \frac{1}{p_1} (\cos^2 \theta_2 - \cos^2 \theta_1) \\
&= f_2 \sin^2 \theta_2 - f_1 \sin^2 \theta_1 + \left(\frac{1}{p_2} - \frac{1}{p_1} \right) \cos^2 \theta_2 + \frac{1}{p_1} (\sin^2 \theta_1 - \sin^2 \theta_2) \\
&= \left(f_2 - \frac{1}{p_1} \right) \sin^2 \theta_2 - \left(f_1 - \frac{1}{p_1} \right) \sin^2 \theta_1 + \left(\frac{1}{p_2} - \frac{1}{p_1} \right) \cos^2 \theta_2 \\
&= \left(f_1 - \frac{1}{p_1} \right) (\sin^2 \theta_2 - \sin^2 \theta_1) + (f_2 - f_1) \sin^2 \theta_2 + \underbrace{\left(\frac{1}{p_2} - \frac{1}{p_1} \right) \cos^2 \theta_2}_{=:v} \\
&= (\theta_2 - \theta_1) \underbrace{\left(f_1 - \frac{1}{p_1} \right) \frac{\sin \theta_2 - \sin \theta_1}{\theta_2 - \theta_1} (\sin \theta_2 + \sin \theta_1)}_{=:u} + v , \tag{2.71}
\end{aligned}$$

where the above definition of μ is only for $x \in [a, b]$ such that $\theta_1(x) \neq \theta_2(x)$, and we set

$$\mu(x) = \left(f_1(x) - \frac{1}{p_1(x)} \right) \cos(\theta_1(x)) (\sin(\theta_1(x)) + \sin(\theta_1(x))) \quad \text{if } \theta_1(x) = \theta_2(x).$$

One can check that the functions μ and v are continuous on $[a, b]$, and that according to (2.66) one has $v \geq 0$ on $[a, b]$. Thus, the difference $w = \theta_2 - \theta_1$ satisfies

$$w'(x) - \mu(x)w(x) = v(x) \geq 0 \quad \text{for all } x \in [a, b].$$

We use the method of integrating factors on this equation. Namely, if M denotes an antiderivative of μ , then one can multiply both sides of this inequality by $e^{-M(x)}$ to obtain

$$\frac{d}{dx} (e^{-M(x)} w(x)) = e^{-M(x)} w'(x) - \mu(x) e^{-M(x)} w(x) \geq 0 \quad \text{for all } x \in [a, b].$$

Let $G(x) = e^{-M(x)} w(x)$. By (2.68), $G(a) \geq 0$, and by the equation above we have $G'(x) \geq 0$ for $a \leq x \leq b$. Therefore if we let $a \leq \alpha \leq \beta \leq b$ and integrate both sides of this inequality from α to β , we get that $G(\beta) \geq G(\alpha)$. Returning to the definition of G and rearranging terms gives

$$\theta_2(\beta) - \theta_1(\beta) \geq e^{M(\beta)-M(\alpha)} (\theta_2(\alpha) - \theta_1(\alpha)) \quad \text{for all } a \leq \alpha \leq \beta \leq b . \tag{2.72}$$

If we choose $\alpha = a$ and $\beta = x$, and use (2.68), then this inequality establishes (2.69).

It remains to show that if the strict inequality $f_1(x) < f_2(x)$ holds for all $x \in (a, b)$, then we have $\theta_1(x) < \theta_2(x)$ for all $x \in (a, b]$. For this, notice that (2.72) implies the following statement: If for some $c \in (a, b)$ one has

$$\theta_2(c) > \theta_1(c), \quad \text{then} \quad \theta_2(x) > \theta_1(x) \quad \text{for all } c \leq x \leq b . \tag{2.73}$$

Now assume that there is a $z \in (a, b)$ such that $\theta_1(z) = \theta_2(z)$. Due to the first part of the proof we have to have $\theta_2 \geq \theta_1$ on $[0, z]$, and according to (2.73) we have to have $\theta_2 \leq \theta_1$ on $[0, z]$. This therefore implies that

$$\theta_2(x) = \theta_1(x) \quad \text{for all } a \leq x \leq z .$$

But then the identity (2.71) can only hold if $v = 0$ on $[a, z]$. Since both terms in v are nonnegative, they have to vanish identically on $[a, z]$. In particular, this implies that

$$(f_2(x) - f_1(x)) \sin^2(\theta_2(x)) = 0 \quad \text{for all } a \leq x \leq z,$$

which together with $f_1 < f_2$ on (a, b) implies that $\theta_1(x) = \theta_2(x) = \ell\pi$ for all $a \leq x \leq z$, for some fixed integer ℓ , i.e., both functions θ_1 and θ_2 are constant on $[a, z]$. However, due to (2.70) this is impossible, since $\sin \theta_1(a) = \ell\pi$ furnishes

$$\theta'_1(a) = \frac{1}{p_1(a)} \cos^2(\theta_1(a)) + f_1(a) \sin^2(\theta_1(a)) = \frac{1}{p_1(a)} > 0.$$

This contradiction completes the proof of Theorem 2.52. \square

Before proceeding to the second comparison result, we illustrate Theorem 2.52 using our standard example.

Example 2.53. We continue our study of Examples 2.48 and 2.49, where we proposed to study a specific Sturm-Liouville boundary value problem by a parameter-dependent initial value problem. We have seen that this initial value problem is given by

$$u'' + \lambda u = 0, \quad \text{with} \quad u(0) = 1 \quad \text{and} \quad u'(0) = -1, \quad (2.74)$$

and we denote its unique solution by u_λ . It is clear that this equation fits into the framework of Theorem 2.52 with $p(x) \equiv 1$ and $f_\lambda(x) \equiv \lambda$. Furthermore, the boundary condition at the left boundary $x = 0$ can be rephrased as an initial condition for the associated angular variable θ_λ . One can easily check that the condition $u_\lambda(0) + u'_\lambda(0) = 0$ is equivalent to

$$\theta_\lambda(0) = \frac{3\pi}{4},$$

since we need to satisfy $\tan \alpha = -1$.

In Figure 2.11 we have shown both the angular and the radial variables for a variety of solutions u_λ of (2.74). The images in the left column are for the angular variable θ_λ , while the images in the right column depict the radial variable $\rho_\lambda > 0$. In the two images in the top row, the five curves correspond to the λ -values $-1.5, -1.25, -1, -0.75$, and -0.5 , which are shown in red, green, black, blue, and magenta, respectively. The five curves in the images in the bottom row are for λ -values $3\pi^2, 3.5\pi^2, 4\pi^2, 4.5\pi^2$, and $5\pi^2$, again shown in red, green, black, blue, and magenta, respectively. Notice that while there is no discernible monotonicity of the radial variable as λ increases, the angular variable increases in a strictly monotonic way. Furthermore, notice that as λ increases, only the solutions shown in black lead to solutions of the boundary value problem (2.51), since only for these functions do we have $\theta_\lambda(b) = 3\pi/4 \bmod \pi$. \blacksquare

We now present the second comparison theorem. In contrast to Theorem 2.52 above, it is concerned with the location of zeros of solutions of a Sturm-Liouville problem. Recall that the zeros of any nontrivial solution u of a second-order linear differential equation necessarily have to be isolated: If we have $u(x_0) = 0$, then $u'(x_0) \neq 0$, since otherwise $u \equiv 0$ on $[a, b]$ ¹⁶.

Theorem 2.54 (Second Sturm-Liouville comparison theorem). *Let $I = [a, b] \subset \mathbb{R}$ denote a bounded interval, let $f_1, f_2 : I \rightarrow \mathbb{R}$ be two continuous functions, and assume that*

¹⁶This is a consequence of the uniqueness theorem for solutions to ordinary differential equations.

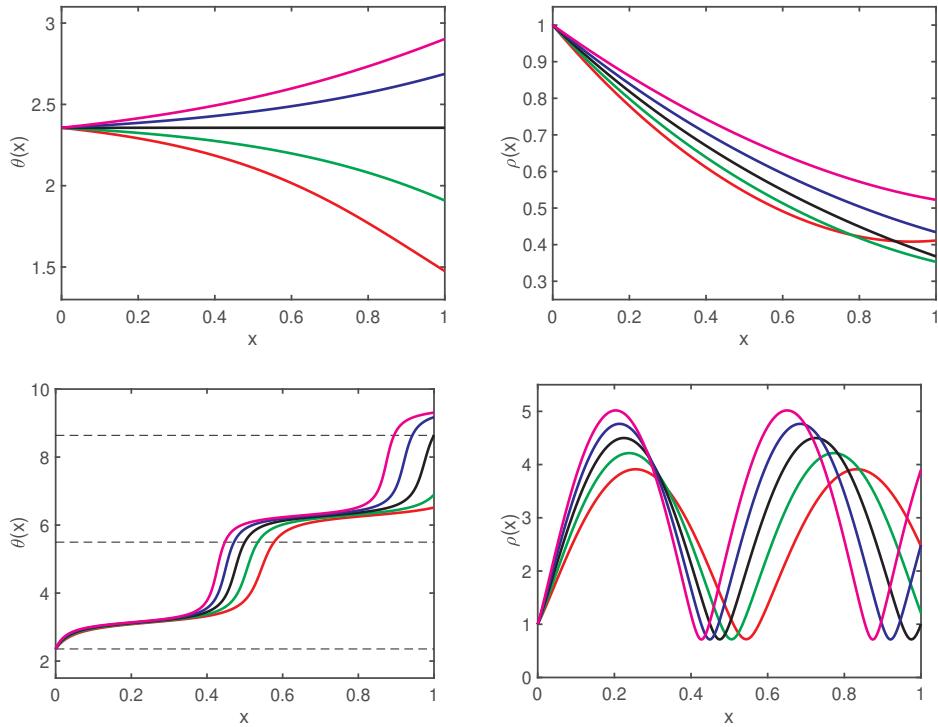


Figure 2.11. Solutions of the initial value problem (2.74) for various λ -values. The left column shows the angular variable θ_λ associated with the solutions, while the right column depicts the radial variable $\rho_\lambda > 0$. In the two images in the top row, the five curves correspond to the λ -values $-1.5, -1.25, -1, -0.75$, and -0.5 , which are shown in red, green, black, blue, and magenta, respectively. The five curves in the images in the bottom row correspond to the λ -values $3\pi^2, 3.5\pi^2, 4\pi^2, 4.5\pi^2$, and $5\pi^2$, again shown in red, green, black, blue, and magenta, respectively. The dashed curves represent the condition $\theta_\lambda = 3\pi/4 \bmod \pi$, which represents both boundary conditions.

the function $p : I \rightarrow \mathbb{R}$ is continuously differentiable with $p(x) > 0$ for all $x \in I$. For $k = 1, 2$, let $u_k : I \rightarrow \mathbb{R}$ denote a nontrivial solution of the second-order linear differential equation

$$\frac{d}{dx} (p(x)u') + f_k(x)u = 0 \quad \text{for } x \in I = [a, b] \subset \mathbb{R}, \quad (2.75)$$

and assume that

$$f_1(x) < f_2(x) \quad \text{for all } a < x < b. \quad (2.76)$$

If $x_1 < x_2$ are the locations of two consecutive zeros of the solution u_1 in the open interval (a, b) , then the function u_2 has to have a zero in the interval (x_1, x_2) .

Proof. We prove this result by contradiction. Assume that the function u_2 has no zero in the interval (x_1, x_2) . Since the differential equation (2.75) is linear, we can then assume without loss of generality that both u_1 and u_2 are strictly positive on (x_1, x_2) — if either of them is negative, just consider its negative instead. If we now multiply the differential equation for u_1 by u_2 , and vice versa, we obtain the two identities

$$\frac{d}{dx} (p u'_1) u_2 + f_1 u_1 u_2 = 0 \quad \text{and} \quad \frac{d}{dx} (p u'_2) u_1 + f_2 u_1 u_2 = 0,$$

and subtracting them implies

$$\frac{d}{dx}(pu'_1)u_2 - \frac{d}{dx}(pu'_2)u_1 = (f_2 - f_1)u_1u_2.$$

Together with

$$\frac{d}{dx}(pu'_1)u_2 - \frac{d}{dx}(pu'_2)u_1 = p'u'_1u_2 + pu''_1u_2 - p'u'_2u_1 - pu''_2u_1 = \frac{d}{dx}(p(u'_1u_2 - u'_2u_1))$$

one then obtains

$$\frac{d}{dx}(p(u'_1u_2 - u'_2u_1)) = (f_2 - f_1)u_1u_2,$$

and integrating both sides from x_1 to x_2 implies

$$p(x_2)u'_1(x_2)u_2(x_2) - p(x_1)u'_1(x_1)u_2(x_1) = \int_{x_1}^{x_2} (f_2(\xi) - f_1(\xi))u_1(\xi)u_2(\xi)d\xi > 0, \quad (2.77)$$

where we have used the fact that $u_1(x_1) = u_1(x_2) = 0$, in combination with the positivity of the integrand on (x_1, x_2) .

According to our assumptions, the function p is positive on $[a, b]$, and we have assumed that $u_2 > 0$ on (x_1, x_2) , which implies $u_2 \geq 0$ on $[x_1, x_2]$. Furthermore, since $u_1(x_1) = 0$ and $u_1 > 0$ on (x_1, x_2) we have $u'_1(x_1) > 0$. Similarly, one can see that the inequality $u'_1(x_2) < 0$ is satisfied. Together, one then obtains

$$p(x_2)u'_1(x_2)u_2(x_2) \leq 0 \quad \text{and} \quad p(x_1)u'_1(x_1)u_2(x_1) \geq 0,$$

and thus $p(x_2)u'_1(x_2)u_2(x_2) - p(x_1)u'_1(x_1)u_2(x_1) \leq 0$. This contradicts (2.77), and the proof of Theorem 2.54 is complete. \square

The above result implies that the zeros of consecutive Sturm-Liouville eigenfunctions have to be interlaced. This can clearly be seen in Figures 2.4 and 2.5.

Existence of Sturm-Liouville eigenvalues

After the preparations above, we can now address most of the remaining assertions of our general Sturm-Liouville Theorem 2.37. For the sake of reference, we restate the parts of the theorem that will be established in this section. We omit parts (b) and (d) in the formulation, since they have already been verified in the proof of Theorem 2.37. Notice that we do not prove part (e) from Theorem 2.37. The verification of this statement involves advanced methods from functional analysis, and lies beyond the scope of this book.

Theorem 2.55 (Sturm-Liouville eigenvalues and eigenfunctions). *We consider the regular Sturm-Liouville problem*

$$\frac{d}{dx}(p(x)u') + q(x)u + \lambda r(x)u = 0 \quad \text{for} \quad x \in I = [a, b] \subset \mathbb{R} \quad (2.78)$$

subject to the boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \quad \text{and} \quad \beta_1 u(b) + \beta_2 u'(b) = 0, \quad (2.79)$$

where $\alpha_1^2 + \alpha_2^2 > 0$ and $\beta_1^2 + \beta_2^2 > 0$, the functions q and r are continuous on I , the function p is continuously differentiable on I , and both p and r are strictly positive on I . Then the following hold.

- (a) The Sturm-Liouville problem (2.78), (2.79) has countably infinitely many eigenvalues which are all real, and which can be ordered in the form

$$\lambda_1 < \lambda_2 < \lambda_3 < \dots < \lambda_k < \dots$$

Furthermore, we have $\lambda_k \rightarrow \infty$ as $k \rightarrow \infty$.

- (c) For each $k \in \mathbb{N}$, the eigenfunction φ_k has exactly $k - 1$ zeros in the open interval (a, b) . Furthermore, between any two consecutive zeros of the eigenfunction φ_{k+1} there is exactly one zero of φ_k . In other words, the zeros of consecutive eigenfunctions are interlaced.

Proof. Our proof is based on considering an initial value problem associated with the Sturm-Liouville problem. More precisely, we consider the differential equation

$$\frac{d}{dx} (p(x)u'_\lambda) + q(x)u_\lambda + \lambda r(x)u_\lambda = 0 \quad \text{for } x \in I = [a, b] \subset \mathbb{R} \quad (2.80)$$

subject to the initial condition

$$u_\lambda(a) = \alpha_2 \quad \text{and} \quad u'_\lambda(a) = -\alpha_1, \quad (2.81)$$

i.e., we have $\alpha_1 u(a) + \alpha_2 u'(a) = 0$. For every $\lambda \in \mathbb{R}$, the initial value problem given by (2.80) and (2.81) has a unique solution $u_\lambda : [a, b] \rightarrow \mathbb{R}$. Furthermore, we can introduce the angular variable θ_λ and the radial component ρ_λ via

$$u_\lambda(x) = \rho_\lambda(x) \sin(\theta_\lambda(x)) \quad \text{and} \quad p(x)u'_\lambda(x) = \rho_\lambda(x) \cos(\theta_\lambda(x)) \quad (2.82)$$

for all $x \in I$. According to Lemma 2.50, these new coordinates satisfy the differential equations

$$\begin{aligned} \rho'_\lambda &= \left(\frac{1}{p(x)} - q(x) - \lambda r(x) \right) \rho_\lambda \sin \theta_\lambda \cos \theta_\lambda, \\ \theta'_\lambda &= \frac{1}{p(x)} \cos^2 \theta_\lambda + (q(x) + \lambda r(x)) \sin^2 \theta_\lambda, \end{aligned} \quad (2.83)$$

subject to

$$\rho_\lambda = \sqrt{p(a)^2 \alpha_1^2 + \alpha_2^2} \quad \text{and} \quad \theta_\lambda(a) = \alpha, \quad (2.84)$$

where $\alpha \in [0, \pi]$ is chosen such that

$$\tan \alpha = -\frac{\alpha_2}{\alpha_1 p(a)}.$$

Here we use the convention $\tan(\pi/2) = \infty$.

Our goal is to determine all values of $\lambda \in \mathbb{R}$ for which the solution u_λ satisfies the second boundary condition in (2.79). This condition can be transformed into a condition on the angular variable. For this, choose $\beta \in (0, \pi]$ (notice the difference in the interval when compared to the choice of α !) in such a way that

$$\tan \beta = -\frac{\beta_2}{\beta_1 p(b)}.$$

Then u_λ satisfies the boundary condition

$$\beta_1 u_\lambda(b) + \beta_2 u'_\lambda(b) = 0 \quad \text{if and only if} \quad \theta_\lambda(b) = \beta + \ell \pi \quad \text{for some } \ell \in \mathbb{Z}. \quad (2.85)$$

We will see later that in fact one has to have $\ell \in \mathbb{N}_0$.

In order to prove the assertions of Theorem 2.55, we make use of the following three facts which will be established at the end of the proof.

- (i) If $\lambda_1 < \lambda_2$, then we have $\theta_{\lambda_1}(x) < \theta_{\lambda_2}(x)$ for all $x \in (a, b]$. Furthermore, for every choice of $x \in [a, b]$ the value $\theta_\lambda(x)$ depends continuously on λ .
- (ii) The function u_λ has a zero at $x \in (a, b)$ if and only if $\theta_\lambda(x) = \ell \pi$ for some $\ell \in \mathbb{Z}$. In addition, if we have $\theta_\lambda(x) = \ell \pi$ for some $\ell \in \mathbb{Z}$, then automatically $\theta'_\lambda(x) > 0$.
- (iii) We have both $\lim_{\lambda \rightarrow -\infty} \theta_\lambda(b) = 0$ and $\lim_{\lambda \rightarrow +\infty} \theta_\lambda(b) = +\infty$.

We delay the verification of these assertions, and begin by showing how their validity implies the theorem.

To begin with, notice that due to (i) the function $\mathbb{R} \ni \lambda \mapsto \theta_\lambda(b) \in \mathbb{R}$ is a continuous and strictly increasing function. Furthermore, (iii) implies that the range of this function is given by \mathbb{R}^+ . Thus, the intermediate value theorem shows that there exists a sequence $\lambda_1 < \lambda_2 < \dots$ such that

$$\theta_\lambda(b) = \beta + (k-1)\pi \quad \text{if and only if} \quad \lambda = \lambda_k, \quad \text{for } k \in \mathbb{N}.$$

Due to $\lim_{\lambda \rightarrow +\infty} \theta_\lambda(b) = +\infty$ one has to have $\lim_{k \rightarrow \infty} \lambda_k = \infty$, and in view of (2.85) this establishes (a).

Next, note that our choice of the boundary value angles $\alpha \in [0, \pi)$ and $\beta \in (0, \pi]$ implies that for every $k \in \mathbb{N} \setminus \{1\}$ one has

$$\alpha < \pi < 2\pi < \dots < (k-1)\pi < \beta + (k-1)\pi \leq k\pi,$$

and that for $k = 1$ there is no integer multiple of π between α and β . In other words, for every $k \in \mathbb{N}$ there are exactly $k-1$ multiples of π which lie strictly between α and $\theta_{\lambda_k}(b)$. Since the function θ_{λ_k} is continuous, it has to achieve each of these values as function value, and according to (ii) this shows that u_{λ_k} has at least $k-1$ zeros. On the other hand, (ii) also implies that if $\theta_{\lambda_k}(c) = \ell \pi$ for some $\ell \in \mathbb{Z}$, then one has to have $\theta_{\lambda_k}(x) > \ell \pi$ for all values $c < x \leq b$. This immediately furnishes that u_{λ_k} has exactly $k-1$ zeros. Together with Theorem 2.54 this establishes (c). Thus, the proof of Theorem 2.55 is complete, once we have verified the three facts (i), (ii), and (iii). This will be done in the remainder of the proof.

Proof of (i): The first assertion follows immediately from Theorem 2.52, since $\lambda_1 < \lambda_2$ implies $q(x) + \lambda_1 r(x) < q(x) + \lambda_2 r(x)$ for all $x \in (a, b)$, due to the assumed strict positivity of r . In addition, we are comparing solutions with the same initial condition, see (2.84). Finally, the continuous dependence of $u_\lambda(x)$ on λ follows from standard ordinary differential equations results, namely the continuous dependence of the solution of an initial value problem on parameters in the equation.

Proof of (ii): In view of (2.82), it is clear that $u_\lambda(x) = 0$ if and only if $\theta_\lambda(x) = \ell \pi$ for some $\ell \in \mathbb{Z}$. Furthermore, if we insert this identity into the differential equation for θ_λ shown in (2.83), one readily obtains

$$\theta'_\lambda(x) = \frac{1}{p(x)} \cos^2(\ell \pi) + (q(x) + \lambda r(x)) \sin^2(\ell \pi) = \frac{1}{p(x)} > 0,$$

due to the assumed strict positivity of p .

Proof of (iii): In order to establish the limit of $\theta_\lambda(b)$ as $\lambda \rightarrow \infty$, we make use of the first comparison theorem, i.e., Theorem 2.52. For this, choose positive constants P , Q , and R such that for all $x \in [a, b]$ we have

$$p(x) \leq P, \quad |q(x)| \leq Q, \quad \text{and} \quad r(x) \geq R > 0.$$

Then for all $\lambda > 0$ we have both

$$P \geq p(x) \quad \text{and} \quad \lambda R - Q \leq q(x) + \lambda r(x) \quad \text{for all } x \in [a, b].$$

Applying Theorem 2.52 with $p_1 = P$, $p_2 = p$, $f_1 = \lambda R - Q$, and $f_2 = q + \lambda r$ then shows that if $\tilde{\theta}_\lambda$ denotes the angular variable associated with the solution \tilde{u}_λ of

$$P u'' + (\lambda R - Q) u = 0 \quad \text{subject to} \quad u(a) = 0,$$

then we have $\theta_\lambda \geq \tilde{\theta}_\lambda$ on $[a, b]$. (Note that $\theta_\lambda(a) = \alpha \geq 0 = \tilde{\theta}_\lambda(a)$.) Since the differential equation for \tilde{u}_λ can be divided by P , one can now apply Lemma 2.51, and this finally yields

$$\theta_\lambda(b) \geq \tilde{\theta}_\lambda(b) > \sqrt{\frac{\lambda R - Q}{P}}(b - a) - \pi \quad \text{for all } \lambda > \frac{Q}{R},$$

which in turn implies $\lim_{\lambda \rightarrow \infty} \theta_\lambda(b) = +\infty$.

Finally, in order to establish the limit of $\theta_\lambda(b)$ as $\lambda \rightarrow -\infty$, let $\varepsilon > 0$ be small enough so that $\alpha < \pi - \varepsilon$ and $\varepsilon < \pi - \varepsilon$, and choose positive constants P , Q , and R such that for all $x \in [a, b]$ we have

$$p(x) \geq P > 0, \quad |q(x)| \leq Q, \quad \text{and} \quad r(x) \geq R > 0.$$

Notice that while Q and R are chosen as above, the constant P is not, since it now gives a positive lower bound on p . Then for all $x \in [a, b]$ for which $\varepsilon \leq \theta_\lambda(x) \leq \pi - \varepsilon$ the second differential equation in (2.83) yields for $\lambda < 0$ the estimate

$$\theta'_\lambda(x) = \frac{1}{p(x)} \cos^2(\theta_\lambda(x)) + (q(x) + \lambda r(x)) \sin^2(\theta_\lambda(x)) \leq \frac{1}{P} + Q - |\lambda| R \sin^2 \varepsilon,$$

and the right-hand side converges to $-\infty$ as $\lambda \rightarrow -\infty$. Combined with the fact that $\theta_\lambda(a) \in [0, \pi]$, this establishes that for every sufficiently large $|\lambda|$ the solution θ_λ has to attain a function value less than ε , somewhere on (a, b) . Note, however, that the above estimate also shows that for large $|\lambda|$ one has $\theta'_\lambda(x) < 0$ whenever $\theta_\lambda(x) = \varepsilon$. Combined with the fact that $\theta'_\lambda(x) > 0$ whenever $\theta_\lambda(x) = 0$, this shows that if $\theta_\lambda(z) \in (0, \varepsilon)$ for some $z \in (a, b)$, then in fact $\theta_\lambda(x) \in (0, \varepsilon)$ for all $z \leq x \leq b$. This implies $0 < \theta_\lambda(b) < \varepsilon$ for all negative λ with sufficiently large $|\lambda|$. Since $\varepsilon > 0$ was arbitrary, this establishes the limit $\lim_{\lambda \rightarrow -\infty} \theta_\lambda(b) = 0$, which completes the proof of the theorem. \square

For the situation of Examples 2.48, 2.49, and 2.53 we illustrate the proof of Theorem 2.55 in Figure 2.12. In this figure, five solutions of the initial value problem (2.74) are considered for various λ -values. The left image shows the angular variable θ_λ associated with the λ -values $-1, \pi^2, 4\pi^2, 9\pi^2$, and $16\pi^2$, shown in cyan, magenta, blue, green, and red, respectively. In other words, these angular variables correspond to the first five eigenfunctions

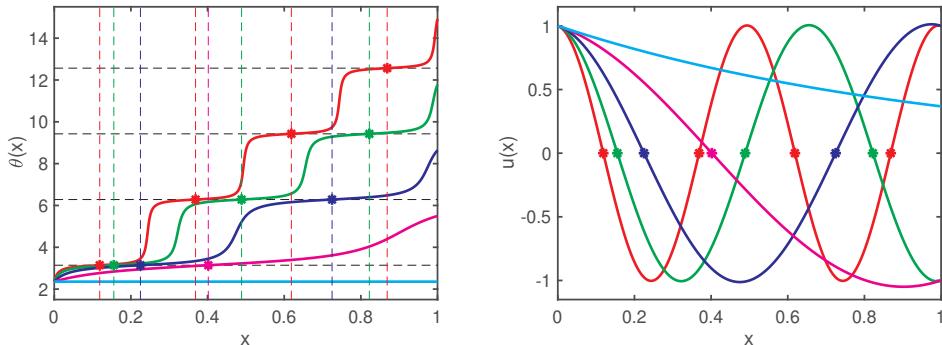


Figure 2.12. Solutions of the initial value problem (2.74) for various λ -values. The left image shows the angular variable θ_λ associated with the λ -values $-1, \pi^2, 4\pi^2, 9\pi^2$, and $16\pi^2$, shown in cyan, magenta, blue, green, and red, respectively. Thus, these solutions correspond to the first five eigenfunctions of the Sturm-Liouville problem (2.51). The image also shows the lines $\theta = \ell\pi$ in dashed black. Intersections with these lines correspond to zeros of the eigenfunctions, which are shown in the right image.

of the Sturm-Liouville problem (2.51). Recall that in this case we have $\alpha = \beta = 3\pi/4$, i.e., the angular curves end at points of the form $\theta = 3\pi/4 + \ell\pi$. The image on the left also shows the lines $\theta = \ell\pi$ as dashed black lines. As we have seen in the above proof, intersections of the angular functions with these lines correspond to zeros of the eigenfunctions, which are shown in the right image.

2.3 ■ A First Taste of Separation of Variables

Separation of variables is a straightforward and powerful analytical method for finding closed form solutions for partial differential equations. It is the single most popular method for finding solutions for partial differential equations. However, separation of variables does come with severe limitations. Before going into details of the method, it has to be pointed out that it does not always yield solutions. In addition, just because one has found all the solutions possible via separation of variables does not mean that all possible solutions of the partial differential equation in question have been determined.

The idea behind the method of separation of variables is to assume that solutions of a given partial differential equation are of a specific form, which is chosen to make it possible to solve the equation directly. The most standard separation of variables assumption is that a solution which depends on several variables can be written as a product of one-variable functions. For example, for a partial differential equation involving functions of two variables, one assumes that the solution is a product of two one-variable functions, such as

$$u(t, x) = T(t)X(x) \quad \text{or} \quad u(x, y) = X(x)Y(y).$$

Substituting this proposed solution form into the partial differential equation gives rise to a series of ordinary differential equations which can in many cases be solved more simply than the original partial differential equation. This method is then combined with the concepts of superposition and orthogonal functions. When successful, the method results in an explicit solution to the original partial differential equation.

In the following, we illustrate the method through a sequence of examples. In the course of this, the reader will undoubtedly recognize solutions to the heat, wave, and

Laplace equations given in the previous chapter, but this time it will be clear how they were constructed.

2.3.1 ■ The Vibrating String of Finite Length

In the first chapter of this book, we introduced the wave equation as a simple and fundamental model for the motion of a vibrating string. Specifically, we presented the one-dimensional wave equation in the form

$$u_{tt} = c^2 u_{xx} \quad \text{for all } t > 0 \quad \text{and} \quad x \in \Omega \subset \mathbb{R}, \quad (2.86)$$

where Ω denotes an open interval. In Section 1.2.4, we were able to present an explicit solution formula for all solutions of the wave equation if the underlying interval is the whole real line $\Omega = \mathbb{R}$, see Theorem 1.52. Yet, for the bounded domain $\Omega = (0, L)$, which actually models a vibrating string of finite length, we satisfied ourselves with a few sporadic solution examples.

In the present section, we demonstrate that the method of separation of variables, in combination with generalized Fourier series, can be used to derive the general solution formula for the finite string case as well. To fix our discussion, we consider the wave equation (2.86) on the finite domain $\Omega = (0, L)$, where $L > 0$, and with homogeneous Dirichlet boundary conditions

$$u(t, 0) = u(t, L) = 0 \quad \text{for all } t \geq 0. \quad (2.87)$$

As was pointed out in Section 1.2.4, a unique solution can only be expected we additionally give initial conditions of the form

$$u(0, x) = f(x) \quad \text{and} \quad u_t(0, x) = g(x), \quad \text{for all } 0 < x < L, \quad (2.88)$$

where f and g are given functions which specify the initial displacement and the initial velocity of the string, respectively.

Reduction to two ordinary differential equations

In order to solve the above problem using separation of variables, we now make the assumption that it is possible to find a solution u of the problem (2.86), (2.87), and (2.88) in the specific form

$$u(t, x) = T(t)X(x),$$

which separates the independent variables t and x . At this point, the functions T and X are completely unknown, and have to be determined from the problem formulation.

As a first step, we now substitute the above solution form into the partial differential equation (2.86). Using the fact that

$$u_{tt}(t, x) = T''(t)X(x) \quad \text{and} \quad u_{xx}(t, x) = T(t)X''(x),$$

this easily implies the identity

$$T''(t)X(x) = c^2 T(t)X''(x) \quad \text{for all } t > 0 \quad \text{and} \quad 0 < x < L.$$

In general, we are interested in finding nontrivial solutions of the wave equation, and therefore we assume that the unknown functions T and X are not identically zero. Dividing both sides of the previous identity by the product $c^2 T(t)X(x)$ then yields

$$\frac{T''(t)}{c^2 T(t)} = \frac{X''(x)}{X(x)} \quad \text{for all } t > 0 \quad \text{and} \quad 0 < x < L,$$

as long as $T(t) \neq 0$ and $X(x) \neq 0$.

We now make a key observation which will allow us to solve this last identity. Notice that the left-hand side of this equation is a function only of t , and the right-hand side is a function which only depends on x . Since we can vary t and x completely independently of each other, the equality of both sides of the equations can only be true if each side is constant, i.e., if both sides equal a constant which is independent of x or t . If we now denote this unknown constant by α , then the validity of the partial differential equation for $u(t, x) = T(t)X(x)$ is equivalent to the two equations

$$X''(x) = \alpha X(x) \quad \text{for all } 0 < x < L, \quad (2.89)$$

as well as

$$T''(t) = \alpha c^2 T(t) \quad \text{for all } t > 0. \quad (2.90)$$

That is, we have separated the partial differential equation into two *linear ordinary differential equations*, one for the unknown function $X(x)$ and one for $T(t)$. These equations are connected via the unknown parameter $\alpha \in \mathbb{R}$.

The boundary value problem for the spatial component

The above discussion showed that assuming the product form of the solution u allowed us to reduce the original partial differential equation (2.86) to two independent ordinary differential equations (2.89) and (2.90). Since the latter equations still allow for an infinity of solutions, we now need to see how the conditions (2.87) and (2.88) impact the functions T and X . Substituting the product $u = TX$ into the boundary conditions implies

$$u(t, 0) = T(t)X(0) = 0 \quad \text{and} \quad u(t, L) = T(t)X(L) = 0 \quad \text{for all } t > 0.$$

This shows that either the function $T(t)$ is identically zero, in which case the solution u is the uninteresting trivial solution, or we need to have both

$$X(0) = 0 \quad \text{and} \quad X(L) = 0.$$

Together with (2.89) we have shown that the function $X(x)$ has to be a solution of the boundary value problem

$$X'' = \alpha X \quad \text{with} \quad X(0) = X(L) = 0, \quad (2.91)$$

i.e., it has to solve the Sturm-Liouville problem discussed in Example 2.39. In this example, we showed that the above boundary value problem has nontrivial solutions only for the parameter values

$$\alpha = -\frac{n^2 \pi^2}{L^2}, \quad \text{where} \quad n \in \mathbb{N}.$$

For any fixed natural number n and $\alpha = \alpha_n = -n^2 \pi^2 / L^2$, the solutions of the boundary value problem (2.91) are precisely the functions

$$X_n(x) = \gamma \sin \frac{n \pi x}{L},$$

where $\gamma \in \mathbb{R}$ is an arbitrary constant, see also Proposition 2.31. Altogether, the Sturm-Liouville problem (2.91) not only provided us explicit formulas for the possible functions $X(x)$, it also narrowed down the admissible values α_n for the constant α in both (2.89) and (2.90).

The general solution of the temporal component

What about the second part of the proposed solution u , i.e., the function $T(t)$? For this we have to consider ordinary differential equation (2.90), but now only for the parameter values $\alpha = \alpha_n$, where $n \in \mathbb{N}$. Then the differential equation can be rewritten as

$$T''(t) = -\frac{c^2 n^2 \pi^2}{L^2} T(t),$$

which is a constant coefficient second-order linear differential equation, whose characteristic polynomial has the two complex imaginary roots $\pm cn\pi i/L$. Thus, the solution $T(t)$ has to be of the form

$$T_n(t) = \hat{F}_n \cos \frac{cn\pi t}{L} + \hat{G}_n \sin \frac{cn\pi t}{L},$$

for arbitrary real constants \hat{F}_n and \hat{G}_n . If we now let $F_n = \hat{F}_n \gamma$ and $G_n = \hat{G}_n \gamma$, we finally obtain that the product

$$u_n(t, x) = T_n(t)X_n(t) = \left(F_n \cos \frac{cn\pi t}{L} + G_n \sin \frac{cn\pi t}{L} \right) \sin \frac{n\pi x}{L} \quad (2.92)$$

is a solution to the wave equation (2.86) subject to the boundary condition (2.87), for arbitrary integers $n \in \mathbb{N}$ and arbitrary coefficients $F_n, G_n \in \mathbb{R}$.

Satisfying special initial conditions

Thus far, the method of separation of variables has produced infinitely many solutions to the wave equation subject to homogeneous Dirichlet boundary conditions, but we have not yet addressed the initial condition (2.88), where f and g are arbitrary given functions. Before we turn to this last remaining issue, note that the above-derived function u_n satisfies the specific initial condition

$$u_n(0, x) = F_n \sin \frac{n\pi x}{L} \quad \text{and} \quad \frac{\partial u_n}{\partial t}(0, x) = \frac{G_n c n \pi}{L} \sin \frac{n\pi x}{L},$$

and this allows us at least to solve very specific initial-boundary value problems for the wave equation.

Example 2.56 (Wave equation, carefully constructed initial conditions). As a first simple example, we now solve the wave equation on the domain $\Omega = (0, L)$ and subject to homogeneous Dirichlet boundary conditions, for the initial conditions in (2.88) with the two functions

$$f(x) = 3 \sin \frac{2\pi x}{L} \quad \text{and} \quad g(x) = \frac{-10c\pi}{L} \sin \frac{2\pi x}{L}.$$

By matching up the terms with the family of wave equation solutions u_n derived above, we see that for $n = 2$, $F_2 = 3$, and $G_2 = -5$, the function $u_2(t, x)$ as in (2.92) yields a solution to the wave equation with homogeneous Dirichlet boundary conditions and the prescribed initial conditions. ■

The above example is an easy victory for the separation of variables method, but in general, unless we are exceptionally lucky with the initial conditions f and g , it is not

possible to choose the integer n , and the coefficients F_n and G_n in such a way that the corresponding u_n satisfies (2.88).

In order to be able to solve for more general initial conditions, we use the superposition principle to form new solutions for the wave equation. Recall that since the wave equation is linear, Lemma 1.63 implies that any linear combination of solutions of (2.86) yields again a solution of the wave equation. In particular, arbitrary linear combinations of the solutions u_n give rise to solutions of the boundary value problem (2.86), (2.87). Specifically, for any real coefficients F_n and G_n and any integer $N \in \mathbb{N}$, the sum

$$u^{(N)}(t, x) = \sum_{n=1}^N \left(F_n \cos \frac{cn\pi t}{L} + G_n \sin \frac{cn\pi t}{L} \right) \sin \frac{n\pi x}{L} \quad (2.93)$$

solves the wave equation. Since each of the terms in the sum in addition satisfies the homogeneous Dirichlet boundary conditions, the sum also satisfies the boundary conditions (2.87). For this solution, the initial conditions are given by

$$f(x) = \sum_{n=1}^N F_n \sin \frac{n\pi x}{L} \quad \text{and} \quad g(x) = \sum_{n=1}^N \frac{G_n c n \pi}{L} \sin \frac{n\pi x}{L}.$$

Therefore, if we stick to finite sums of terms, it is possible to solve the equation as long as the initial conditions are of this form. We demonstrate this in the following example.

Example 2.57 (Wave equation, finite linear combination initial conditions). In this example, our goal is to solve the wave equation with homogeneous Dirichlet boundary conditions and initial conditions

$$f(x) = 18 \sin \frac{2\pi x}{L} + 12 \sin \frac{8\pi x}{L} \quad \text{and} \quad g(x) = \frac{-10c\pi}{L} \sin \frac{2\pi x}{L} + \frac{6c\pi}{L} \sin \frac{3\pi x}{L}.$$

As in the previous example, by matching terms with the general superposition solution u above, we see that for $N = 8$, $F_2 = 18$, $G_2 = -5$, $G_3 = 2$, $F_8 = 12$, and all other coefficients equal to zero, we get a solution with the stated initial conditions. That is, the function

$$\begin{aligned} u(t, x) = & \left(18 \cos \frac{2c\pi t}{L} - 5 \sin \frac{2c\pi t}{L} \right) \sin \frac{2\pi x}{L} \\ & + 2 \sin \frac{3c\pi t}{L} \sin \frac{3\pi x}{L} + 12 \cos \frac{8c\pi t}{L} \sin \frac{8\pi x}{L}. \end{aligned}$$

is a solution with the prescribed initial conditions. ■

General initial conditions and generalized Fourier series

In both examples of initial conditions, we have successfully solved the wave equation, but the examples demonstrate the severe limitations of the solutions we have presented. While the second example of an initial condition consisting of a finite linear combination of sine functions is an improvement over the first example, in which the initial condition which consists of a single multiple of a sine function, a finite linear combinations of sine functions still represents a restrictive choice for initial conditions. However, if we allow for infinite series — that is, if we consider the limit $N \rightarrow \infty$ in our solution sum — then the possible initial conditions f and g are any functions which can be represented by a

generalized Fourier series of sine functions, as illustrated in Example 2.15, as long as the resulting series solution is convergent and differentiable.

Before considering questions of convergence and differentiability, we start by writing down a *formal series solution*. More precisely, for a given pair of initial conditions, we write down an infinite series which converges in the space $L^2(I)$, regardless of whether this series actually converges to a differentiable function. We then subsequently address properties of convergence. The solution will be of the form

$$u(t, x) = \sum_{n=1}^{\infty} \left(F_n \cos \frac{cn\pi t}{L} + G_n \sin \frac{cn\pi t}{L} \right) \sin \frac{n\pi x}{L}. \quad (2.94)$$

If initial profiles for both the displacement f and the velocity g are given, then the coefficients of best approximation in the formal series solution

$$u^{(N)}(t, x) = \lim_{N \rightarrow \infty} u_N(t, x),$$

where $u^{(N)}$ was defined in (2.93), can be obtained by the methods of orthogonal functions. Specifically,

$$u(0, x) = f(x) = \sum_{n=1}^{\infty} F_n \sin \frac{n\pi x}{L},$$

and

$$u_t(0, x) = g(x) = \sum_{n=1}^{\infty} \frac{cn\pi}{L} G_n \sin \frac{n\pi x}{L}.$$

Therefore, by applying Theorem 2.14 on generalized Fourier series, and in particular the identity in (2.7), the coefficients F_n and G_n can be computed as

$$F_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L} dx, \quad (2.95)$$

and

$$G_n = \frac{2}{n\pi c} \int_0^L g(x) \sin \frac{n\pi x}{L} dx. \quad (2.96)$$

We illustrate the explicit derivation of such a series solution in the following two examples, both of which consider polynomial initial conditions.

Example 2.58 (General initial condition for the wave equation I). In the first example, our goal is to solve the wave equation on the interval $\Omega = (0, 1)$ and subject to homogeneous Dirichlet boundary conditions, but for the polynomial initial conditions

$$f(x) = 64x^3(1-x)^3 \quad \text{and} \quad g(x) = 0,$$

which are shown in the top left image in Figure 2.13. For these functions, one can easily determine explicit formulas for the coefficients F_n and G_n by computing the integrals in (2.95) and (2.96), respectively, and this implies that

$$F_n = 18432 \frac{10 - n^2\pi^2}{n^7\pi^7} \quad \text{for all odd integers} \quad n \in 2\mathbb{N}_0 + 1,$$

and that all remaining coefficients F_n and G_n are zero. Therefore, the formal series solution for the wave equation with the above initial conditions is given by

$$u(t, x) = \sum_{k=0}^{\infty} \frac{18432(10 - (2k+1)^2\pi^2)}{(2k+1)^7\pi^7} \cos(c(2k+1)\pi t) \sin((2k+1)\pi x).$$

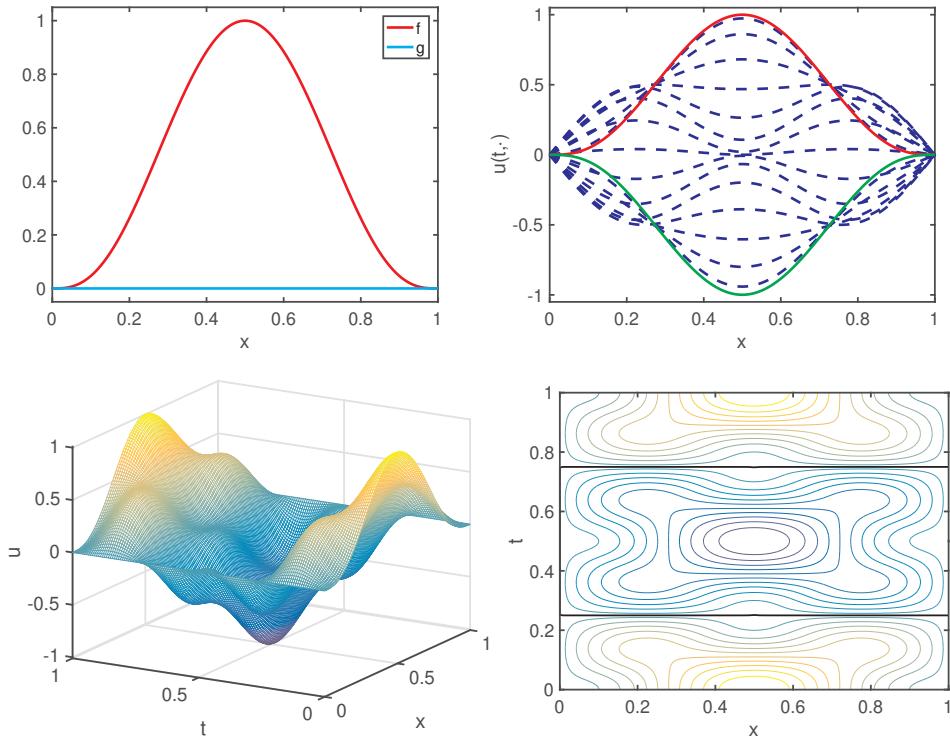


Figure 2.13. Initial value problem for the wave equation on the domain $\Omega = (0, 1)$, as discussed in Example 2.58. The initial conditions $f(x) = 64x^3(1-x)^3$ and $g(x) = 0$ are shown in the upper left panel, while the upper right image contains solution snapshots $u(t, \cdot)$ at times $t = 0$ (red curve), $t = 1/2$ (green curve), as well as fifteen times t between these two values (dashed blue curves). Since we consider $c = 2$, the solution is periodic in time with period 1, and the two lower panels show the solution $u : [0, 1]^2 \rightarrow \mathbb{R}$, both as a surface and as a level curve plot.

In this representation, we already omitted the terms for even n , and only consider the odd indices $n = 2k + 1$ for $k \in \mathbb{N}_0$.

If we further assume that the constant c in the wave equation (2.86) is chosen as $c = 2$, then the above solution becomes periodic with respect to t with minimal period 1. In this case solution snapshots $u(t, \cdot)$ at times $t = j/32$ for $j = 0, \dots, 16$ are shown in the upper right image of Figure 2.13, and the restriction $u : [0, 1]^2 \rightarrow \mathbb{R}$ of the t -periodic solution to its minimal fundamental domain is depicted in the two lower panels as a surface and an intensity plot. ■

Example 2.59 (General initial condition for the wave equation II). As our second example, we consider the wave equation on the interval $\Omega = (0, 1)$ and subject to homogeneous Dirichlet boundary conditions, but now for the initial conditions

$$f(x) = 4x(1-x) \quad \text{and} \quad g(x) = 0.$$

Observe that by Example 2.15, we already know that

$$4x(1-x) = \sum_{n=1}^{\infty} F_n \sin(n\pi x),$$

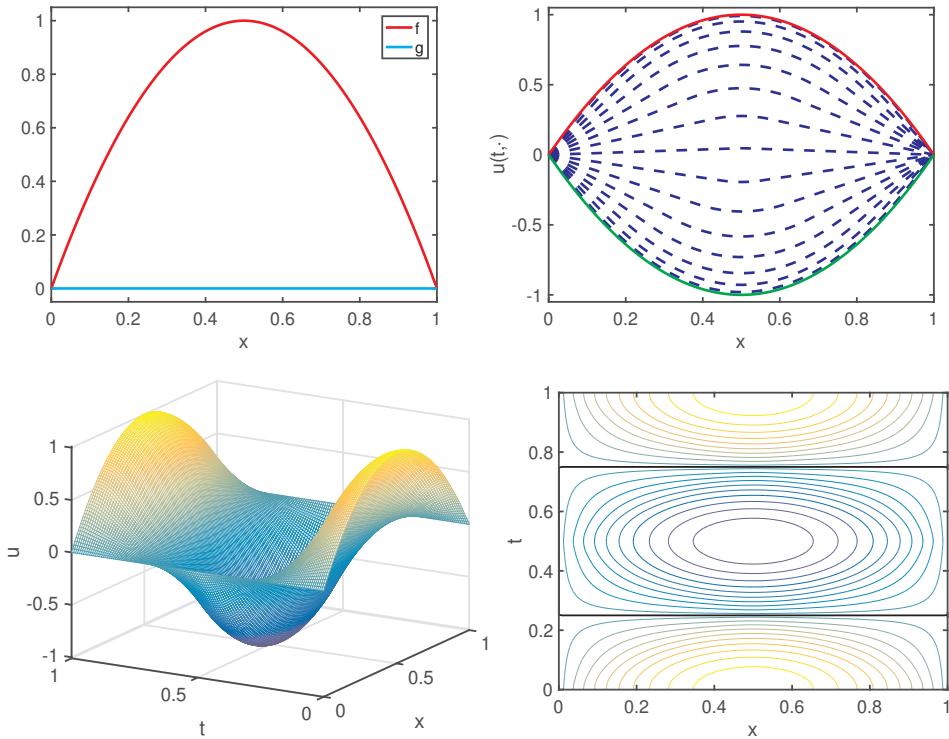


Figure 2.14. Initial value problem for the wave equation on the domain $\Omega = (0, 1)$, as discussed in Example 2.59. The initial conditions $f(x) = 4x(1-x)$ and $g(x) = 0$ are shown in the upper left panel, while the upper right image contains solution snapshots $u(t, \cdot)$ at times $t = 0$ (red curve), $t = 1/2$ (green curve), as well as fifteen times t between these two values (dashed blue curves). Since we consider $c = 2$, the solution is periodic in time with period 1, and the two lower panels show the solution $u : [0, 1]^2 \rightarrow \mathbb{R}$, both as a surface and as a level curve plot.

where $F_n = 32/(n^3\pi^3)$ for odd integers $n \in \mathbb{N}$, and $F_n = 0$ for even n . In addition, notice that the function $g(x) = 0$ can be written as a series $\sum_{n=1}^{\infty} G_n \sin(n\pi x)$ if we set $G_n = 0$ for all $n \in \mathbb{N}$. Therefore, the formal series solution for these initial conditions is given by

$$u(t, x) = \sum_{k=0}^{\infty} \frac{32}{(2k+1)^3\pi^3} \cos((2k+1)\pi t) \sin((2k+1)\pi x).$$

While the initial conditions can be found in the upper left panel of Figure 2.14, the remaining three images contain solution snapshots $u(t, \cdot)$, as well as surface and intensity plots for u , again for the case $c = 2$. ■

Establishing the smoothness of the solution

In the previous examples we have seen concrete evidence that the concept of generalized Fourier series, combined with solution families which are found via separation of variables, can lead to formal series representations of initial/boundary value problems for the wave equation. In fact, another glance at Figures 2.13 and 2.14 seems to confirm that the so-called *formal solutions* are in fact classical solutions of the partial differential equation.

But is this really the case? In this last part of the section we now address the question of differentiability of the formal series. In fact, we start even more elementary and ask whether the series converges for all arguments t and x . For this, we will make use of two results which were mentioned earlier. On the one hand, the uniform convergence of the series can be studied using Theorem 2.19, which in turn is based on Theorem 2.20, the Weierstraß M -test. On the other hand, we can establish the differentiability of the series by applying Theorem 2.21. We begin by revisiting our earlier examples.

Example 2.60 (Smoothness of the series solution I). In our first Example 2.58, the method of separation of variables and its combination with generalized Fourier series produced a formal series solution to the wave equation of the form

$$u(t, x) = \sum_{n \in 2\mathbb{N}-1}^{\infty} \frac{18432(10 - n^2\pi^2)}{n^7\pi^7} \cos(cn\pi t) \sin(n\pi x). \quad (2.97)$$

In order to show that the series on the right-hand side actually converges, consider arbitrary values of the arguments t and x . Since the sine and cosine functions only take function values in the interval $[-1, 1]$, one then obtains the estimate

$$\begin{aligned} & \left| \frac{18432(10 - n^2\pi^2)}{n^7\pi^7} \cos(cn\pi t) \sin(n\pi x) \right| \\ & \leq \frac{18432 |10 - n^2\pi^2|}{n^7\pi^7} \leq \frac{18432(n^2\pi^2 + 10)}{n^7\pi^7} \leq \frac{18432\left(1 + \frac{10}{n^2\pi^2}\right)}{n^5\pi^5} \\ & \leq \frac{18432\left(1 + \frac{10}{\pi^2}\right)}{n^5\pi^5} \leq \frac{121.26}{n^5} \quad \text{for all } n \in 2\mathbb{N}-1 \quad \text{and } t, x \in \mathbb{R}. \end{aligned}$$

Since the series $\sum_{n=1}^{\infty} n^{-5}$ converges, so does the series $\sum_{n \in 2\mathbb{N}-1} 121.26n^{-5}$. Furthermore, each series term in (2.97) is a continuous function with respect to (t, x) . Then Theorem 2.19 implies that the series converges uniformly on all of \mathbb{R}^2 , and that the limit function u is continuous. In other words, the equality in (2.97) is justified.

In order to verify the differentiability of u , we begin by formally differentiating the series with respect to x , followed by showing that the formal series converges to a continuous function which is indeed the desired derivative of u . Differentiating formally, one obtains

$$u_x(t, x) = \sum_{n \in 2\mathbb{N}-1}^{\infty} \frac{18432(10n\pi - n^3\pi^3)}{n^7\pi^7} \cos(cn\pi t) \cos(n\pi x), \quad (2.98)$$

as well as

$$u_{xx}(t, x) = - \sum_{n \in 2\mathbb{N}-1}^{\infty} \frac{18432(10n^2\pi^2 - n^4\pi^4)}{n^7\pi^7} \cos(cn\pi t) \sin(n\pi x). \quad (2.99)$$

Arguing as above, one can now easily see that the absolute value of the n -th series term in (2.98) and (2.99) are bounded by $121.26n^{-4}$ and $121.26n^{-3}$, respectively. Since both of these terms still give rise to convergent series, another application of Theorem 2.19 implies that the series in (2.98) and (2.99) converge uniformly to continuous functions. Even more is true — according to Theorem 2.21 the limiting functions have to be the partial derivatives u_x and u_{xx} .

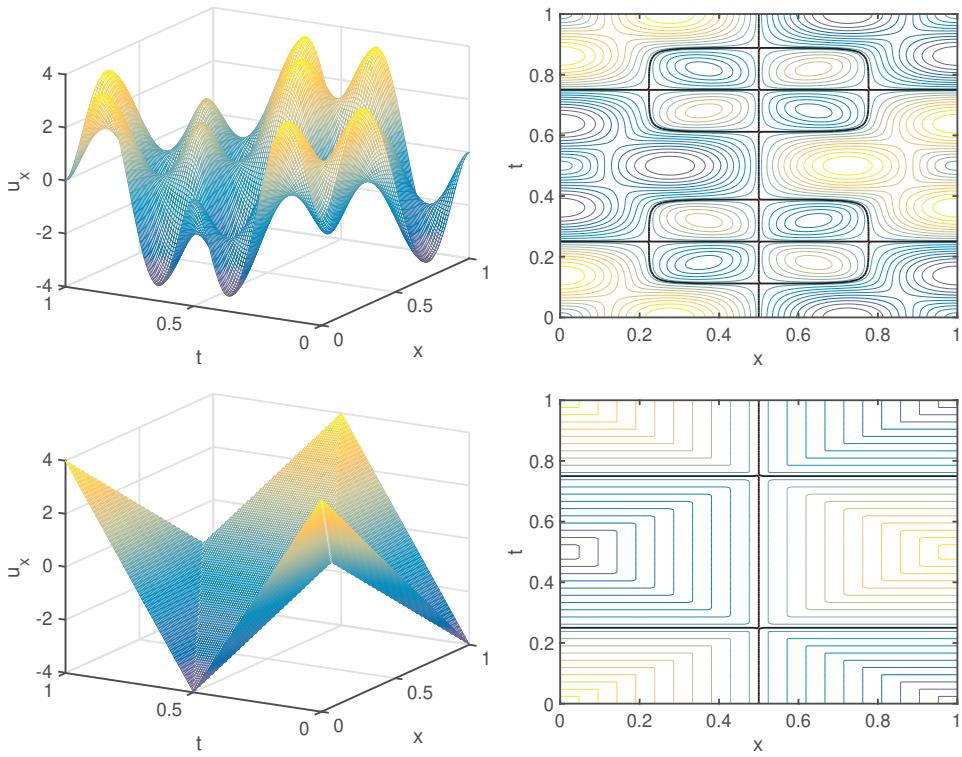


Figure 2.15. Smoothness of the series solutions for the wave equation. The first row shows surface and level curve plots of the first partial derivative u_x for the series solution u from Example 2.58, while the second row contains analogous images for the series in Example 2.59.

In a similar way one can formally differentiate and subsequently show that the derivatives u_t , u_{tt} , and u_{tx} exist and are continuous. We leave these verifications to the reader. Altogether, we have shown that the series in (2.97) defines a twice continuously differentiable function u , which therefore is indeed a solution to the wave equation subject to homogeneous Dirichlet boundary conditions. For illustration purposes, the first partial derivative u_x is shown in the first row of Figure 2.15 as a surface and an intensity plot. ■

Example 2.61 (Smoothness of the series solution II). We now turn our attention to the second Example 2.59. In this situation, we derived the formal series solution

$$u(t, x) = \sum_{n \in 2\mathbb{N}-1}^{\infty} \frac{32}{n^3 \pi^3} \cos(cn\pi t) \sin(n\pi x). \quad (2.100)$$

Proceeding as before, one can easily show that the series converges uniformly on all of \mathbb{R}^2 , and therefore the above equality is justified. Moreover, one can also show that the function u is continuously differentiable and that the partial derivatives u_t and u_x are obtained by term-wise differentiation. Yet, when studying the second partial derivative u_{xx} , formal differentiation implies

$$u_{xx}(t, x) \stackrel{?}{=} - \sum_{n \in 2\mathbb{N}-1}^{\infty} \frac{32}{n\pi} \cos(cn\pi t) \sin(n\pi x),$$

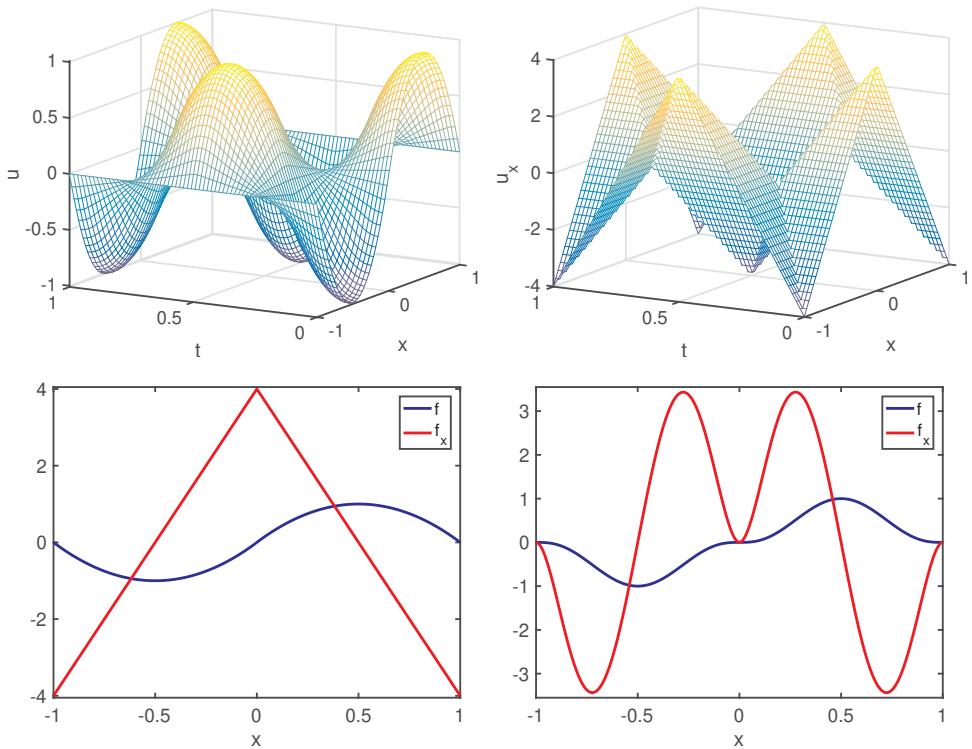


Figure 2.16. Explanation for the lack of higher smoothness in Example 2.59. In the top left panel, the function u defined in (2.100) is shown for the arguments $(t, x) \in [0, 1] \times [-1, 1]$, its partial derivative u_x is shown in the upper right. The function u is a formal series solution of the wave equation for the initial condition shown in the lower left image, together with its first derivative. The function f is obtained from the one in Example 2.59 through an odd extension. Notice that this extended function is only once continuously differentiable. In contrast, the lower right image shows the odd extension of the initial condition in Example 2.58, which gives a twice continuously differentiable function.

and since the harmonic series $\sum_{n=1}^{\infty} 1/n$ diverges, we can no longer establish the uniform convergence of this series. Altogether, using our approach based on the Weierstraß M -test, one can only show that the function u defined in (2.100) is once continuously differentiable. An illustration of the partial derivative u_x can be found in the second row of Figure 2.15.

Needless to say, being unable to use a certain proof technique does not automatically mean that the sought-after result is not true. It is therefore natural to wonder whether the function u defined in (2.100) is actually twice continuously differentiable, and one only needs to use a different argument to show it. We will now demonstrate that in this specific situation, u is not twice differentiable.

To see this, recall that we proved more than required when we established the uniform convergence of the series in (2.100) and of its first partial derivatives. Rather than giving a function on the domain $\mathbb{R}_0^+ \times \overline{\Omega} = \mathbb{R}_0^+ \times [0, 1]$, the function u is in fact defined on all of \mathbb{R}^2 , and it is continuously differentiable there. Consider the function u now on the domain $\mathbb{R}_0^+ \times [-1, 1]$, as shown in the first row of Figure 2.16. According to our derivation, one would still expect the function to be a solution of the wave equation, since it is the limit of linear combinations of the solutions (2.92). Moreover, one can easily see

that the function u satisfies both

$$u(t, -1) = 0 \quad \text{and} \quad u(t, 1) = 0 \quad \text{for all } t > 0,$$

i.e., it satisfies homogeneous Dirichlet boundary conditions on the domain $\tilde{\Omega} = (-1, 1)$. Finally, since the functions $\sin(n\pi x)$ are all odd functions, the limit function u will also be odd, and we therefore have

$$u(0, x) = \begin{cases} 4x(1-x) & \text{for } x \in (0, 1), \\ 4x(1+x) & \text{for } x \in (-1, 0). \end{cases}$$

In other words, the function u satisfies an initial condition which is obtained from the extending the function $f(x) = 4x(1-x)$ on $(0, 1)$ via an odd reflection at $x = 0$ onto the interval $(-1, 0)$. This extended initial condition, together with its first derivative, is shown in the lower left panel of Figure 2.16. One can immediately see that it is not differentiable at $x = 0$. Moreover, a glance at the upper right panel of the same figure shows that the partial derivative u_x is also not differentiable along the line segment $t > 0$ and $x = 0$. Thus, the function u is certainly not twice continuously differentiable.

In contrast, the lower right image in Figure 2.16 shows the initial condition obtained via odd reflection from the function $f(x) = 4x^3(1-x)^3$ in Example 2.60, together with its first derivative. This time the extended function is indeed twice continuously differentiable. ■

The above two examples demonstrate a characteristic property of hyperbolic partial differential equations. For such problems, the smoothness of the initial conditions affects the smoothness of the formal series solution in a profound way. In general, one cannot expect the solution to be any smoother than the initial conditions. In addition, since according to our discussion after Definition 1.53 the Dirichlet boundary value problem on a finite domain can be thought of as a whole-line problem through odd extensions, it is actually the smoothness of the odd extension of the initial conditions that determines the smoothness of the series.

Based on the above discussion, it is natural to wonder whether the method of separation of variables has failed us in the second example. After all, the goal was to obtain a series representation of the solution of an initial/boundary value problem, and this series solution turned out not to be twice differentiable. In fact, it was one of the crucial insights of the theory of partial differential equations that in order to guarantee the existence of solutions for large classes of problems, one has to extend the notion of differentiability itself. This leads to the concept of what are known as *weak derivatives*. One can show that the function in (2.100) is twice differentiable in the weak sense and therefore gives rise to a *weak solution* of the wave equation. Since a more detailed discussion of this topic lies beyond the scope of this book, we will just think of u as a solution, despite the fact that it is not everywhere twice differentiable. We summarize our findings as follows:

- **Smoothness properties of hyperbolic problems:** *Generally, hyperbolic problems such as the wave equation only have smooth solutions if the given initial and boundary data is sufficiently smooth. If this is not the case, a formal series representation for a potential solution might not be differentiable in the interior of the domain $\mathbb{R}^+ \times \Omega$, and in fact, a classical solution might not exist. In order to address this issue, one can broaden the concept of solution to weak solutions.*

The statement shows that when it comes to general hyperbolic equations, there is no such thing as a free smoothness lunch. This will be different for elliptic and parabolic problems.

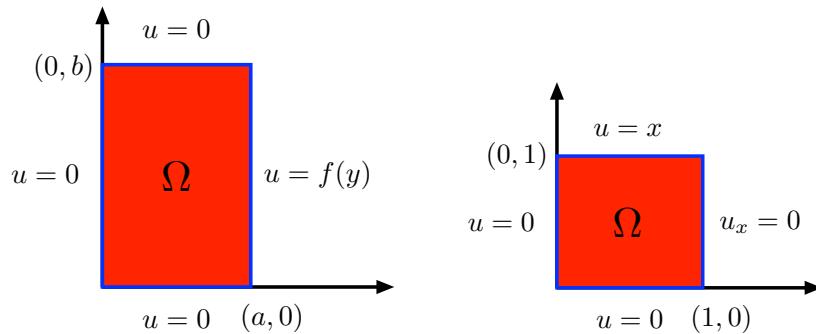


Figure 2.17. On left, an illustration of the Dirichlet boundary conditions for Laplace's equation in Example 2.62. On the top, left, and bottom edges we assume homogeneous conditions, whereas the profile on the right edge follows a prescribed function f . On right, an illustration of the mixed Dirichlet and Neumann boundary conditions for the Helmholtz equation in Example 2.63. On the left and bottom edges, we assume homogeneous Dirichlet conditions, whereas on the right edge we assume a homogeneous Neumann condition. The profile on the top edge follows a prescribed function.

2.3.2 ▪ Two Elliptic Problems on a Rectangle

We now turn to using separation of variables to find explicit solutions for two examples of elliptic partial differential equations, namely Laplace's equation $\Delta u = 0$ and a Helmholtz equation $\Delta u + u = 0$.

Separation of variables for the Laplace equation

Our discussion in the introduction uncovered a range of properties for solutions to Laplace's equation, such as the mean-value property and the maximum principle. In addition, we have presented a detailed proof that solutions to Dirichlet boundary value problems for the Laplace equation are always unique. However, we have not yet given a method for finding solution formulas. In the present section find a solution to the Laplace equation on a two-dimensional rectangular domain with prescribed boundary conditions.

Example 2.62 (Laplace's equation on a rectangular domain). Consider the Laplace equation $\Delta u = 0$ on the rectangle $\Omega = (0, a) \times (0, b)$, equipped with a specific set of Dirichlet boundary conditions. More precisely, assume $u = 0$ on each side of the rectangle except on the vertical right-hand side edge $\{a\} \times (0, b)$. On this edge, we impose the condition $u = f(y)$ for a given function $f : (0, b) \rightarrow \mathbb{R}$. The situation is illustrated in the left-hand image of Figure 2.17.

We proceed as in the last section and try to find solutions of the Laplace equation which are of product form. That is, assume that the solution u is of the form $u(x, y) = X(x)Y(y)$. Substituting this formula into the partial differential equation yields

$$X''(x)Y(y) + X(x)Y''(y) = 0 \quad \text{for all } (x, y) \in (0, a) \times (0, b).$$

Collecting terms with the same variable, and realizing that the two sides have to be equal to a constant, gives rise to the equation

$$\frac{X''(x)}{X(x)} = \frac{-Y''(y)}{Y(y)} = \mu \quad \text{for some constant } \mu \in \mathbb{R}.$$

As before, we made the implicit assumption that we are only interested in nontrivial solutions u , and therefore the division by $X(x)$ and $Y(y)$ is admissible. Separating the above identity into two then gives the two ordinary differential equations

$$X'' = \mu X \quad \text{and} \quad Y'' = -\mu Y.$$

We now see how the boundary conditions translate into conditions on the functions X and Y . Assuming again that the solution u is nontrivial, the homogeneous parts of the boundary conditions imply

$$u(0, y) = X(0)Y(y) = 0, \quad \text{which implies} \quad X(0) = 0,$$

$$u(x, 0) = X(x)Y(0) = 0, \quad \text{which implies} \quad Y(0) = 0,$$

$$u(x, b) = X(x)Y(b) = 0, \quad \text{which implies} \quad Y(b) = 0.$$

Since the final nonhomogeneous Dirichlet boundary condition involves an as yet unspecified function f , we ignore this condition for now. Rather, we try to determine to what extent the above three conditions specify X and Y .

According to the form of the boundary conditions, we have obtained two conditions on Y , and only one on X . We therefore start by determining the possible solution factors $Y(y)$. They have to solve the Sturm-Liouville problem

$$Y'' = -\mu Y \quad \text{subject to} \quad Y(0) = Y(b) = 0,$$

which has nontrivial solutions Y_n only for

$$\mu = \mu_n = \frac{n^2\pi^2}{b^2} \quad \text{for} \quad n \in \mathbb{N},$$

and in this case the function Y_n is of the form

$$Y_n(y) = B_n \sin \frac{n\pi y}{b} \quad \text{for some} \quad B_n \in \mathbb{R}.$$

We now let $n \in \mathbb{N}$ be arbitrary and set $\mu = \mu_n$. Then the differential equation for X , together with the one constraint on X induced by the homogeneous Dirichlet boundary conditions, can be rewritten as

$$X_n'' = \frac{n^2\pi^2}{b^2} X_n \quad \text{subject to} \quad X_n(0) = 0.$$

Since the general solution of the differential equation is

$$X_n(x) = C_n e^{n\pi x/b} + D_n e^{-n\pi x/b},$$

the condition $X_n(0) = 0$ implies the identity $D_n = -C_n$. Thus, the solution X_n has to be of the form $X_n(x) = 2C_n \sinh(n\pi x/b)$, and using the abbreviation $A_n = 2C_n B_n$ we finally obtain the solution

$$u_n(x, y) = X_n(x)Y_n(y) = A_n \sinh \frac{n\pi x}{b} \sin \frac{n\pi y}{b} \quad \text{for some} \quad A_n \in \mathbb{R}.$$

According to our construction, each of the functions u_n satisfies homogeneous Dirichlet boundary conditions on three sides of the rectangular domain, and therefore the same is true if we form the formal series solution

$$u(x, y) = \sum_{n=1}^{\infty} A_n \sinh \frac{n\pi x}{b} \sin \frac{n\pi y}{b} \quad \text{for arbitrary} \quad A_n \in \mathbb{R} \quad (2.101)$$

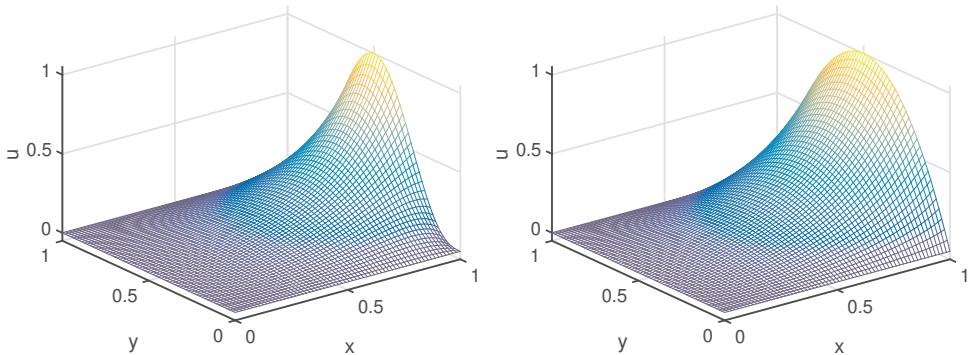


Figure 2.18. Two solutions for the elliptic problem considered in Example 2.62. Both images show solutions of Laplace’s equation subject to Dirichlet boundary conditions, constructed as a series via separation of variables. While the solutions are zero on three sides of the rectangular domain, along the fourth side the function shown in the left image satisfies $u(1,y) = 64y^3(1-y)^3$, while the panel on the right is for the boundary condition $u(1,y) = 4y(1-y)$.

via superposition. Notice that the value of this series at the remaining part of the domain Ω is given by $u(a,y) = \sum_{n=1}^{\infty} A_n \sinh(n\pi a/b) \sin(n\pi y/b)$, i.e., it takes the form of a generalized Fourier series involving the sine family. In order to satisfy the final Dirichlet boundary condition, we now only have to determine the constants A_n in such a way that

$$f(y) = \sum_{n=1}^{\infty} A_n \sinh \frac{n\pi a}{b} \sin \frac{n\pi y}{b},$$

and according to Theorem 2.14 this is the case for

$$A_n = \frac{2}{b \sinh \frac{n\pi a}{b}} \int_0^b f(y) \sin \frac{n\pi y}{b} dy. \quad (2.102)$$

This provides a formal series representation of the solution for general f . ■

In order to illustrate the above example, we consider the square domain $\Omega = (0,1)^2$ and the two specific boundary conditions

$$f(y) = 64y^3(1-y)^3 \quad \text{and} \quad f(y) = 4y(1-y).$$

Using the explicit formulas (2.101) and (2.102), one can then easily verify that the corresponding formal solution series are given by

$$u(x,y) = \sum_{n \in 2\mathbb{N}-1} \frac{18432(10-n^2\pi^2)}{n^7\pi^7 \sinh(n\pi)} \sinh(n\pi x) \sin(n\pi y) \quad (2.103)$$

and

$$u(x,y) = \sum_{n \in 2\mathbb{N}-1} \frac{32}{n^3\pi^3 \sinh(n\pi)} \sinh(n\pi x) \sin(n\pi y), \quad (2.104)$$

respectively. These solutions are shown in Figure 2.18. We would like to point out that at this point, these are only *formal* series solutions, as we have not yet discussed their convergence and smoothness. Before addressing such questions in more detail, we present another example in the next section, thereby allowing us to address the smoothness of both examples simultaneously.

Separation of variables for the Helmholtz equation

In this section, we present the elliptic Helmholtz equation $\Delta u + u = 0$, this time with mixed Dirichlet and Neumann boundary conditions.

Example 2.63 (Helmholtz equation with mixed boundary conditions). As our second example we consider a variation on Laplace's equation, which is usually referred to as the Helmholtz equation. In two space dimensions, this elliptic equation is given by

$$\Delta u + u = u_{xx} + u_{yy} + u = 0 \quad \text{for an unknown function } u = u(x, y), \quad (2.105)$$

and we try to solve it for the square domain $\Omega = (0, 1)^2$ using the method of separation of variables. To illustrate a new aspect of the method, we consider the mixed boundary conditions

$$u(0, y) = 0, \quad u_x(1, y) = 0, \quad u(x, 0) = 0, \quad \text{and} \quad u(x, 1) = x, \quad (2.106)$$

for all $x, y \in (0, 1)$. These boundary conditions are depicted on the right-hand side of Figure 2.17. In Section 2.4, we will see that this problem is in fact a special case of a more general type of problem known as an *eigenvalue problem for the Laplacian*.

To solve this problem using separation of variables, the reader will no longer be surprised to see the assumption that u can be written in product form as $u(x, y) = X(x)Y(y)$. Since the boundary conditions will undoubtedly produce a nontrivial solution, we can again substitute the form of u into the differential equation, and then divide by XY . This implies first the identity $X''Y + XY'' + XY = 0$, and then

$$\frac{X''}{X} + \frac{Y''}{Y} + 1 = 0, \quad \text{leading to} \quad \frac{Y''}{Y} = -\left(\frac{X''}{X} + 1\right).$$

Since both sides of the last equality depend on different independent variables, they have to be constant. Thus, there exists a constant $\mu \in \mathbb{R}$ such that

$$X'' + (\mu + 1)X = 0 \quad \text{and} \quad Y'' = \mu Y.$$

Moreover, the three homogeneous parts of the boundary conditions show that

$$\begin{aligned} u(0, y) &= X(0)Y(y) = 0, & \text{which implies} & \quad X(0) = 0, \\ u_x(1, y) &= X'(1)Y(y) = 0, & \text{which implies} & \quad X'(1) = 0, \\ u(x, 0) &= X(x)Y(0) = 0, & \text{which implies} & \quad Y(0) = 0. \end{aligned}$$

As before, we ignore the nonzero boundary condition for now, since it depends on a nontrivial function and therefore has to be obtained from the form of the resulting series solution.

For the unknown function X there are two boundary conditions, so we will solve for it first. One can show that nontrivial solutions for X only exist if $\mu + 1 > 0$. Then one can define the constant $v > 0$ in such a way that $\mu + 1 = v^2$. This in turn implies that the function X satisfies the boundary value problem

$$X'' + v^2 X = 0 \quad \text{subject to} \quad X(0) = 0 \quad \text{and} \quad X'(1) = 0.$$

The general solution to the differential equation is given by

$$X(x) = A \cos(vx) + B \sin(vx) \quad \text{for arbitrary } A, B \in \mathbb{R}.$$

In order to satisfy the condition $X(0) = 0$ we need to have $A = 0$. Due to the second boundary condition, one needs to satisfy the identity $Bv \cos v = 0$, and therefore nontrivial solutions only occur for

$$v = v_n = \frac{(2n+1)\pi}{2}, \quad \text{where } n \in \mathbb{N}_0.$$

The associated value of the constant μ is then given by

$$\mu = \mu_n = \frac{(2n+1)^2\pi^2}{4} - 1 = \frac{(2n+1)^2\pi^2 - 4}{4},$$

and the solution X_n has to be of the form

$$X_n(x) = B_n \sin \frac{(2n+1)\pi x}{2} \quad \text{for arbitrary } B_n \in \mathbb{R} \quad \text{and } n \in \mathbb{N}_0.$$

Notice that the functions X_n are in fact the solutions of a standard Sturm-Liouville problem, and therefore they form a complete orthogonal set in $L^2(0,1)$.

We now turn to the unknown function Y , and fix $n \in \mathbb{N}_0$ for this arbitrarily. Then the function Y_n has to satisfy the differential equation

$$Y_n'' = \frac{(2n+1)^2\pi^2 - 4}{4} Y_n, \quad \text{as well as } Y_n(0) = 0.$$

One can easily see that because of $n \geq 0$, the multiplicative coefficient on the right-hand side of the differential equation is strictly positive. This implies that the solution Y_n has to be of the form

$$Y_n(y) = C_n \exp \frac{y\sqrt{(2n+1)^2\pi^2 - 4}}{2} + D_n \exp \frac{-y\sqrt{(2n+1)^2\pi^2 - 4}}{2}$$

for some $C_n, D_n \in \mathbb{R}$, and the condition $Y_n(0) = 0$ then immediately implies $C_n + D_n = 0$. Therefore the function Y_n is given by

$$Y_n(y) = 2C_n \sinh \frac{y\sqrt{(2n+1)^2\pi^2 - 4}}{2} \quad \text{for some } C_n \in \mathbb{R},$$

and the product $u_n(x, y) = X_n(x)Y_n(y)$ yields a solution of the Helmholtz equation for all $n \in \mathbb{N}_0$. Since this solution satisfies homogeneous boundary conditions on three parts of the boundary, and since the Helmholtz equation is linear, we can then let $F_n = 2B_n C_n$ and finally derive the formal series solution

$$u(x, y) = \sum_{n=0}^{\infty} F_n \sin \frac{(2n+1)\pi x}{2} \sinh \frac{y\sqrt{(2n+1)^2\pi^2 - 4}}{2}.$$

The constants F_n can be determined from the fourth and last boundary condition, which requires us to satisfy

$$u(x, 1) = \sum_{n=0}^{\infty} F_n \sin \frac{(2n+1)\pi x}{2} \sinh \frac{\sqrt{(2n+1)^2\pi^2 - 4}}{2} = x \quad \text{for all } x \in (0, 1).$$

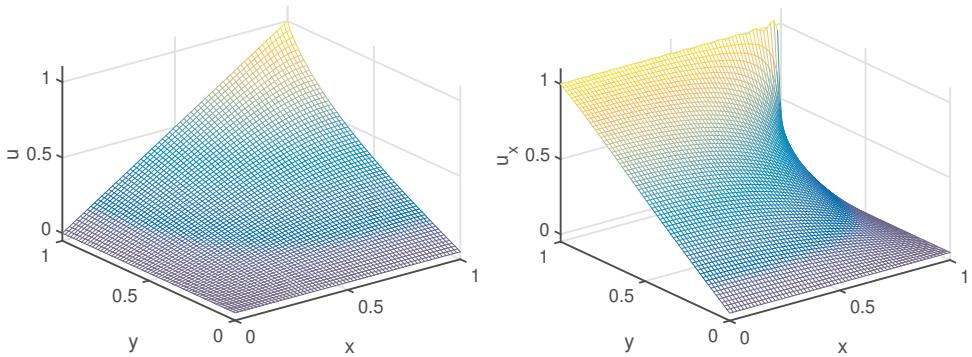


Figure 2.19. The solution for the elliptic Helmholtz equation boundary value problem considered in Example 2.63. The left image shows the solution u , while the right panel contains the partial derivative u_x . Notice that this derivative appears to be discontinuous at the point $(1,1)$, and that the series representation gives rise to sharp oscillations in the graph. This is due to the Gibbs phenomenon described in Example 2.17.

Theorem 2.14 on generalized Fourier series then implies

$$F_n \sinh \frac{\sqrt{(2n+1)^2\pi^2 - 4}}{2} = 2 \int_0^1 x \sin \frac{(2n+1)\pi x}{2} dx = \frac{(-1)^n 8}{(2n+1)^2\pi^2},$$

and from this we derive the formal series representation

$$u(x,y) = \sum_{n=0}^{\infty} \frac{(-1)^n 8 \sin \frac{(2n+1)\pi x}{2} \sinh \frac{y\sqrt{(2n+1)^2\pi^2 - 4}}{2}}{(2n+1)^2\pi^2 \sinh \frac{\sqrt{(2n+1)^2\pi^2 - 4}}{2}}, \quad (2.107)$$

for the solution of the elliptic problem (2.105) subject to (2.106). The solution is shown in the left panel of Figure 2.19, its partial derivative u_x is shown on the right. ■

Smoothness of the formal series representations

The above two examples demonstrate that the method of separation of variables can successfully be used to derive formal series representations of solutions to elliptic boundary value problems. But do these series really define differentiable functions? At first glance, the situation appears to be completely analogous to the case of the wave equation which was discussed in the previous section, as the following example shows.

Example 2.64 (Smoothness of the series solution I). We begin by revisiting the elliptic boundary value problem from Example 2.62, which considered Laplace's equation on a rectangle and led to the formal series representation given in (2.101) with associated generalized Fourier coefficients as in (2.102). Rather than studying the smoothness of this general solution, we concentrate on the specific boundary function $f(y) = 64y^3(1-y)^3$, which gives rise to the series in (2.103).

In order to apply Theorem 2.19 and Theorem 2.21, we need to bound the n -th term in this series uniformly for all $(x,y) \in \Omega = (0,1)^2$. For this, let $n \in 2\mathbb{N} - 1$ be arbitrary

and rewrite the absolute value of the n -th term in (2.103) as

$$\underbrace{\frac{18432|10-n^2\pi^2|}{n^7\pi^7}}_{=\beta_1} \quad \underbrace{\left|\frac{\sinh(n\pi x)}{\sinh(n\pi)}\right|}_{=\beta_2} \quad \underbrace{|\sin(n\pi y)|}_{=\beta_3}.$$

One can easily see that for $(x, y) \in (0, 1)^2$ and for any natural number n these terms can be bounded as

$$\beta_1 \leq \frac{C}{n^5}, \quad \beta_2 \leq 1, \quad \text{and} \quad \beta_3 \leq 1,$$

for some constant $C > 0$, and therefore Theorem 2.19 implies that the above series converges uniformly to a continuous function u .

To determine the differentiability properties of the series, we now differentiate the terms in the series ℓ -times, and denote the analogous factors as above by $\beta_{1,\ell}$, $\beta_{2,\ell}$, and $\beta_{3,\ell}$. Notice that every partial derivative of the n -th term in the series adds a factor $n\pi$ to $\beta_{1,\ell}$, and that the expressions $\sinh(n\pi x)$ and $\sin(n\pi y)$ in $\beta_{2,\ell}$ and $\beta_{3,\ell}$ might have to be replaced by $\cosh(n\pi x)$ and $\cos(n\pi y)$, respectively. We will see soon that for arbitrary $x \in (0, 1)$ the estimate $\cosh(n\pi x) \leq \cosh(n\pi) \leq \coth(\pi) \sinh(n\pi)$ holds. Thus, for an ℓ -th partial derivative of the terms in the series the analogous factors as above are bounded by

$$\beta_{1,\ell} \leq \frac{C}{n^{5-\ell}}, \quad \beta_{2,\ell} \leq \coth(\pi), \quad \text{and} \quad \beta_{3,\ell} \leq 1.$$

Theorem 2.21 now implies that the function u defined in (2.103) is in fact three-times continuously differentiable. ■

The same method can be applied to the remaining two explicit series representations derived above in the elliptic setting. We leave it as an exercise to the reader to show that the series solution u defined in (2.104) for the boundary function $f(y) = 4y(1-y)$ in Example 2.62 is once continuously differentiable. Furthermore, for the series constructed in the Helmholtz Example 2.63 one can merely establish its continuity. Note, however, that in both of these cases, the above method cannot be used to establish any higher degree of smoothness.

So far, the situation seems to be very much comparable to the hyperbolic setting. In order to show that this assessment could not be further from the truth, we need the following simple result about hyperbolic functions.

Lemma 2.65 (Two useful estimates for the hyperbolic functions \sinh and \cosh). *The standard hyperbolic functions \sinh and \cosh satisfy the estimate*

$$\frac{\sinh \sigma}{\sinh \tau} \leq \frac{p!}{(\tau - \sigma)^p} \quad \text{for all } 0 \leq \sigma < \tau \quad \text{and any integer } p \in \mathbb{N}, \quad (2.108)$$

as well as

$$\frac{\cosh \tau}{\sinh \tau} \leq \frac{\cosh \sigma}{\sinh \sigma} \quad \text{for all } 0 < \sigma \leq \tau. \quad (2.109)$$

In both estimates, the variables σ and τ are real numbers.

Proof. We begin by establishing (2.108). Notice that

$$\frac{\sinh \tau}{\sinh \sigma} = \frac{e^\tau - e^{-\tau}}{e^\sigma - e^{-\sigma}} = \frac{e^{\tau-\sigma} - e^{-\tau+\sigma}}{1 - e^{-2\sigma}} \geq e^{\tau-\sigma} - e^{-\tau+\sigma} \geq e^{\tau-\sigma} - 1 = \sum_{k=1}^{\infty} \frac{(\tau-\sigma)^k}{k!}.$$

Every term in the sum is non-negative, and therefore it can be bounded below by the p -th term. Taking the reciprocal gives the claimed estimate. The estimate in (2.109) is just expressing the monotonicity of the hyperbolic cotangent. Thus the proof of the lemma is complete. \square

Equipped with this simple lemma, we now delve deeper into the smoothness of series representations obtained via separation of variables for elliptic problems.

Example 2.66 (Smoothness of the series solution II). We again revisit the elliptic boundary value problem from Example 2.62, but this time for the general rectangular domain $\Omega = (0, a) \times (0, b)$ and for an arbitrary boundary value function $f \in L^2(0, b)$. In other words, we do not even assume that the function f is continuous. We have seen that due to the completeness of the sine family there are coefficients F_n such that

$$f(y) \stackrel{L^2}{=} \sum_{n=1}^{\infty} F_n \sin \frac{n\pi y}{b},$$

and according to (2.9) one can easily see that $F_n \rightarrow 0$ as $n \rightarrow \infty$. We would like to point out, however, that this convergence might be extremely slow. Using the Fourier coefficients F_n , the formal solution formulas (2.101) and (2.102) can be combined into the series representation

$$u(x, y) = \sum_{n=1}^{\infty} \frac{F_n}{\sinh \frac{n\pi a}{b}} \sinh \frac{n\pi x}{b} \sin \frac{n\pi y}{b} \quad (2.110)$$

for the solution to the elliptic problem in Example 2.62.

Now let $0 < \alpha < a$ denote a number, and let $\ell \in \mathbb{N}_0$ be arbitrary. Then any ℓ -th order partial derivative of the n -th term in the above series has an absolute value of the form

$$\underbrace{|F_n| n^\ell \pi^\ell}_{=\beta_{1,\ell}} \quad \underbrace{\left| \frac{\sinh \frac{n\pi x}{b}}{\sinh \frac{n\pi a}{b}} \right|}_{=\beta_{2,\ell}} \quad \underbrace{\left| \sin \frac{n\pi y}{b} \right|}_{=\beta_{3,\ell}} \quad \underbrace{\left| \frac{\sinh \frac{n\pi \alpha}{b}}{\sinh \frac{n\pi a}{b}} \right|}_{=\beta_{4,\ell}}, \quad (2.111)$$

where the expressions $\sinh(n\pi x/b)$ and $\sin(n\pi y/b)$ in $\beta_{2,\ell}$ and $\beta_{3,\ell}$ might have to be replaced by $\cosh(n\pi x/b)$ or $\cos(n\pi y/b)$, respectively. Based on (2.109) and the fact that the F_n converge to zero, one can then see that for some constant $C > 0$ we have

$$\beta_{1,\ell} \leq C n^\ell, \quad \beta_{2,\ell} \leq C, \quad \text{and} \quad \beta_{3,\ell} \leq 1,$$

as long as $(x, y) \in [0, \alpha] \times [0, b]$. Notice that for the $\beta_{2,\ell}$ -estimate it is essential that we choose x only from the interval $[0, \alpha]$, and not from the interval $(0, a)$ which is part of the definition of the underlying domain Ω . Finally, according to (2.108), the last term can be bounded as

$$\beta_{4,\ell} \leq \frac{b^p p!}{(\alpha - \alpha)^p \pi^p n^p}, \quad \text{for any integer } p \in \mathbb{N}.$$

If we now choose $p = \ell + 2$, then the product in (2.111) can be bounded by a constant times n^{-2} , and therefore the series obtained via ℓ -times formal differentiation converges uniformly to a continuous limit function on $[0, \alpha] \times [0, b]$ due to Theorem 2.19. In addition, Theorem 2.21 then shows that the function u defined in (2.110) is infinitely many

times differentiable in $[0, \alpha] \times [0, b]$. Finally, since $0 < \alpha < a$ was arbitrary, we obtain that u is infinitely many times differentiable in the interior of the domain Ω . ■

The implications of this example are truly remarkable. For the Laplace equation on a rectangular domain subject to Dirichlet boundary conditions, separation of variables can be used to construct a formal series solution u which turns out to be infinitely many times differentiable in the interior of the rectangle — regardless of the smoothness properties of the boundary function $f \in L^2(0, b)$. Note in particular that for discontinuous f , the function u is not even continuous on the closure $\overline{\Omega}$. A similar result can also be derived for the solution of the Helmholtz problem in Example 2.63.

This type of behavior, clearly in stark contrast to the situation of hyperbolic problems in the last section, is typical of elliptic boundary value problems. This can be summarized as follows:

- **Smoothness properties of elliptic problems:** *In general, elliptic problems such as the Laplace and the Helmholtz equation lead to solutions which are infinitely many times differentiable in the interior of the domain Ω . This high degree of smoothness can be observed even for nonsmooth boundary conditions. Thus, elliptic problems usually give rise to classical solutions of the underlying partial differential equation in the interior of the domain, despite possibly discontinuous boundary behavior.*

The above statement is usually abbreviated by saying that elliptic equations have a smoothing property.

2.3.3 ■ The Diffusion of Heat in One Dimension

As our last application, we use the method of separation of variables in the context of the heat equation in one space dimension, i.e., we seek solutions $u = u(t, x)$ for the parabolic problem

$$u_t = c^2 u_{xx} .$$

The illustrious history of this specific example is particularly appropriate in the current section. The equation originally appeared in a paper submitted by Fourier in 1807, a paper from which both the technique of separation of variables and the subject of Fourier series stem. Although the paper was considered to contain striking originality, it was never published due to its lack of rigor. Indeed it did lack rigor, as Fourier claimed that all functions could be represented by an infinite sum of trigonometric functions, whereas the conditions under which this is true are quite subtle; see also our discussion in Section 2.1.2. In fact, it took most of the 19-th century to sort out the necessary rigor. A revised version of these ideas was eventually published, leading to the development of Fourier series.

A heat equation problem with homogeneous Dirichlet boundary conditions

We consider the temperature distribution of a rigid length L bar whose ends are kept at the constant temperature zero. If we denote the distance from one end of the bar to the current position by x , then we can think of the interval $\Omega = (0, L)$ as a model for the bar. Furthermore, once we specify an initial temperature distribution $f : \Omega \rightarrow \mathbb{R}$, the evolution of the temperature $u(t, x)$ inside the bar is described by the parabolic partial differential equation

$$u_t = c^2 u_{xx} \quad \text{on the time-space domain } \mathbb{R}^+ \times \Omega = \mathbb{R}^+ \times (0, L) ,$$

subject to the homogeneous Dirichlet boundary conditions

$$u(t, 0) = u(t, L) = 0 \quad \text{for all } t > 0,$$

and the initial condition

$$u(0, x) = f(x) \quad \text{for all } x \in \Omega = (0, L).$$

See Section 1.2.1 for a detailed discussion of the derivation of this equation. Note that we do not specify an initial condition of the form $u_t(0, x) = g(x)$ like we did for the wave equation, since the heat equation is only first order with respect to the time variable t . That is, the initial conditions of the heat and wave equations are respectively analogous to there only being one initial condition for a first-order ordinary differential equation, and two initial conditions for a second-order ordinary differential equation.

For the above equation, we now derive a series representation for the solution through the method of separation of variables. We assume that the solution can be written as a product of the form $u(t, x) = T(t)X(x)$. After substitution into the partial differential equation one obtains

$$T'(t)X(x) = c^2 T(t)X''(x), \quad \text{leading to} \quad \frac{T'(t)}{c^2 T(t)} = \frac{X''(x)}{X(x)} = \mu,$$

where we assume that u is a nontrivial solution, and μ is a real constant. Thus, the validity of the partial differential equation is equivalent to the two ordinary differential equations

$$T' = \mu c^2 T \quad \text{and} \quad X'' = \mu X.$$

For nontrivial solutions u the boundary conditions reduce to $X(0) = X(L) = 0$. This can be used to determine the admissible values of μ and the solutions X . More precisely, if $n \in \mathbb{N}$ is arbitrary, and if we let

$$\mu = \mu_n = -\frac{n^2 \pi^2}{L^2},$$

then the boundary value problem $X'' = \mu X$ subject to $X(0) = X(L) = 0$ has the solutions

$$X_n(x) = A_n \sin \frac{n \pi x}{L} \quad \text{for arbitrary } A_n \in \mathbb{R}.$$

Inserting $\mu = \mu_n$ into the equation for the unknown function T then gives the first-order linear differential equation $T' = -c^2 n^2 \pi^2 T / L^2$, which has the general solution

$$T_n(t) = B_n e^{-c^2 n^2 \pi^2 t / L^2} \quad \text{for arbitrary } B_n \in \mathbb{R}.$$

If we finally set $F_n = A_n B_n$, and use the principle of superposition as in the previous two sections, then we obtain the formal series representation

$$u(t, x) = \sum_{n=1}^{\infty} F_n e^{-c^2 n^2 \pi^2 t / L^2} \sin \frac{n \pi x}{L} \quad \text{for arbitrary } F_n \in \mathbb{R}. \quad (2.112)$$

In order to satisfy the initial condition $u(0, x) = f(x)$, the coefficients F_n need to be determined in such a way that

$$f(x) = \sum_{n=1}^{\infty} F_n \sin \frac{n \pi x}{L}, \quad \text{and therefore} \quad F_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n \pi x}{L} dx. \quad (2.113)$$

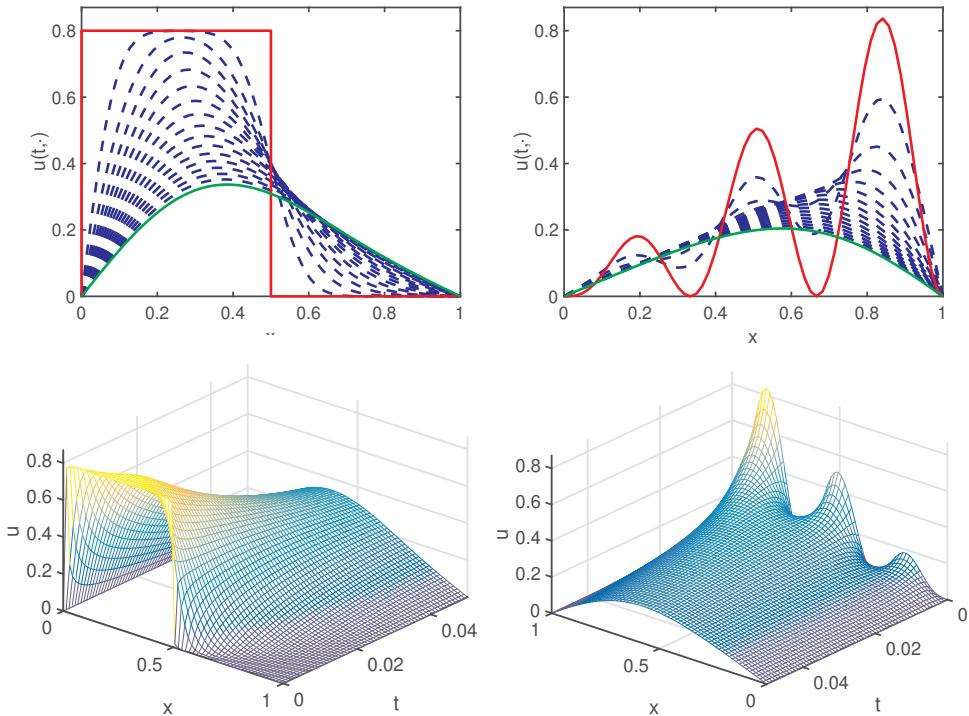


Figure 2.20. Initial value problem for the heat equation on the domain $\Omega = (0, 1)$ subject to homogeneous Dirichlet boundary conditions, and with heat coefficient $c = 1$. The left column is for an initial condition f which has the constant value $4/5$ on the interval $(0, 1/2)$, and is zero otherwise. The right column is for the function $f(x) = x \sin^2(3\pi x)$. The images in the top row show solution snapshots $u(t, \cdot)$ at times $t = 0$ (red curve), $t = 1/20$ (green curve), as well as fifteen times t between these two values (dashed blue curves). The panels in the second row show the solutions. Notice that the image in the lower left shows evidence of the Gibbs phenomenon near $t = 0$.

Thus, the method of separation of variables produces an explicit formal series representation for the solution u .

In the next example, we specify the function f and thus obtain concrete solution formulas for (2.112) and (2.113).

Example 2.67 (Two specific examples of heat diffusion). Consider the unit length domain $\Omega = (0, 1)$ and let $c = 1$. Then Figure 2.20 shows the solutions of two initial value problems for the heat equation subject to homogeneous Dirichlet boundary conditions. The left column is for the initial condition f which has the constant value $4/5$ on the interval $(0, 1/2)$, and is zero otherwise. In this case, formula (2.113) implies

$$f(x) = \begin{cases} 4/5 & \text{for } x \in (0, 1/2), \\ 0 & \text{for } x \in [1/2, 1], \end{cases} \quad \text{which implies} \quad F_n = \frac{16 \sin^2(n\pi/4)}{5n\pi}.$$

In contrast, the right column is for the function

$$f(x) = x \sin^2(3\pi x), \quad \text{which gives} \quad F_n = \begin{cases} \frac{(-1)^n 36}{n\pi(n^2 - 36)} & \text{for } n \neq 6, \\ -\frac{1}{8\pi} & \text{for } n = 6. \end{cases}$$

In both cases, we have $n \in \mathbb{N}$. In Figure 2.20, the images in the top row show solution snapshots $u(t, \cdot)$ at times $t = 0$ (red curve), $t = 1/20$ (green curve), as well as fifteen times t between these two values (dashed blue curves). The panels in the second row show the solutions. Notice that the image in the lower left exhibits significant oscillations near the line $t = 0$, i.e., near the initial condition. This is due to the fact that in this example, the function f is not continuous. The oscillations are a result of the Gibbs phenomenon described in Example 2.17. ■

In the above discussion we considered the case of homogeneous Dirichlet boundary conditions. However, other boundary conditions are also possible using the method of separation of variables. For example, if one would like to solve the heat equation for homogeneous Neumann boundary conditions, only the resulting Sturm-Liouville problem for the function X in the product $u(t, x) = T(t)X(x)$ changes. We leave it as a simple exercise for the reader to show that the resulting series solution then has the same form as in (2.112), but with \sin replaced by \cos , and with the series summation starting at $n = 0$.

Smoothness of the formal series representations

To close this section, we return to the series representation for u given in (2.112). Since the Fourier coefficients F_n satisfy (2.113), the theorem on generalized Fourier series again implies $F_n \rightarrow 0$ as $n \rightarrow \infty$, i.e., the sequence (F_n) is certainly bounded. Now let $\tau > 0$ be arbitrary, but fixed. If we take partial derivatives of the terms in the series, one can easily see that the absolute value of the n -th term in the formal derivative series is bounded by an expression of the form

$$C n^\ell e^{-c^2 n^2 \pi^2 \tau / L^2} \quad \text{for all } (t, x) \in [\tau, \infty) \times [0, L],$$

where $C > 0$ is a constant, and $\ell \in \mathbb{N}$ depends on the order of the considered partial derivative of u . For the exponential term, note that the definition of the natural exponential function implies the estimate

$$e^{-c^2 n^2 \pi^2 \tau / L^2} = \left(\sum_{k=0}^{\infty} \frac{c^{2k} n^{2k} \pi^{2k} \tau^k}{L^{2k}} \right)^{-1} \leq \frac{L^{2p}}{c^{2p} \pi^{2p} \tau^p} \frac{1}{n^{2p}} \quad \text{for any } p \in \mathbb{N},$$

since the series can be bounded below by the term for $k = p$. This, however, shows that if we choose p in such a way that $2p - \ell \geq 2$, then Theorems 2.19 and 2.21 can be applied as in Example 2.66 to imply that all partial derivatives of the series u in (2.112) exist on the domain $[\tau, \infty) \times [0, L]$. This in turn demonstrates that u is infinitely differentiable on this restricted domain. Finally, since $\tau > 0$ was arbitrary, we see that u is in fact infinitely many times differentiable in $\mathbb{R}^+ \times \bar{\Omega} = (0, \infty) \times [0, L]$.

The above derivation is maybe even more remarkable than the smoothness of solutions in the elliptic case. For the heat equation subject to homogeneous Dirichlet boundary conditions, one obtains an infinitely many times differentiable solution, which is smooth up to and including the boundary $\partial\Omega$, as long as $t > 0$. This remains true even if the initial condition f is a discontinuous function in the space $L^2(\Omega)$. This can be summarized as follows:

- **Smoothness properties of parabolic problems:** *Generally, parabolic problems such as the heat equation have the smoothing property. Even for discontinuous initial conditions $f : \Omega \rightarrow \mathbb{R}$, the formal series solution u usually is infinitely many times differentiable on the set $\mathbb{R}^+ \times \bar{\Omega}$.*

Note again that the smoothing property for parabolic problems is somewhat stronger than the one for elliptic equations, as it usually includes the boundary of Ω for $t > 0$.

2.4 ■ Eigenvalue Problems via Separation of Variables

In this section we turn our attention to solving eigenvalue problems for partial differential equations using the method of separation of variables. Such eigenvalue problems are the natural generalization of Sturm-Liouville problems to higher-dimensional domains. We will see later that problems of this type arise in a variety of settings, such as for example in the study of the acoustics of musical instruments. Moreover, we will demonstrate that these problems can even be used to devise powerful numerical methods for solving partial differential equations. Finally, these problems are of significant theoretical interest, since they often make it possible to construct explicit series representations of solutions to time-dependent partial differential equations.

2.4.1 ■ Eigenfunctions and Their Geometry

We begin our discussion with the study of the Helmholtz equation on general domains $\Omega \subset \mathbb{R}^d$. This equation is given by

$$\Delta u + \lambda u = 0 \quad \text{in } \Omega \subset \mathbb{R}^d, \quad (2.114)$$

and has to be studied subject to suitable boundary conditions. Solving the Helmholtz equation is equivalent to studying the eigenvalue problem for the negative Laplacian $-\Delta$. Namely, for which values of λ does the equation

$$-\Delta u = \lambda u \quad \text{in } \Omega \subset \mathbb{R}^d$$

have nontrivial solutions u that satisfy the chosen boundary conditions? Notice that this question is a natural generalization of a Sturm-Liouville eigenvalue problem to higher-dimensional domains. Thus, we will adopt the same notation as in this one-dimensional setting and call every value of λ for which nontrivial solutions of (2.114) exist an *eigenvalue* of (2.114), even though technically it is an “eigenvalue” of the negative Laplacian. A corresponding solution to the eigenvalue equation is called an *eigenfunction*.

For specific geometrically simple domains $\Omega \subset \mathbb{R}^d$, the method of separation of variables can be used to determine the eigenvalues and eigenfunctions for the Helmholtz equation (2.114). We have already found an eigenfunction in Example 2.63 where $\lambda = 1$ was given. We now demonstrate examples for which both eigenfunction u and the eigenvalue λ are unknown. These examples can be solved directly using separation of variables. For comparison, we refer the reader to Examples 2.39, 2.40, and 2.43, which cover the case of a one-dimensional domain $\Omega = (0, L)$ subject to homogeneous Dirichlet, homogeneous Neumann, and periodic boundary conditions, respectively.

Example 2.68 (Neumann boundary conditions in two dimensions). Our first example is concerned with two space dimensions and considers the Helmholtz equation on the rectangular domain $\Omega = (0, a) \times (0, b)$ with homogeneous Neumann boundary conditions. Since the outer unit normal vector at points on the boundary is either $(\pm 1, 0)$ or $(0, \pm 1)$, we have to solve the boundary value problem

$$\begin{aligned} \Delta u + \lambda u &= 0 && \text{in } \Omega, \\ u_x(0, y) &= u_x(a, y) = 0 && \text{for } y \in (0, b), \\ u_y(x, 0) &= u_y(x, b) = 0 && \text{for } x \in (0, a). \end{aligned}$$

Since this is a two-dimensional problem, the one-dimensional Sturm-Liouville problem techniques do not work directly. Instead, we seek solutions using separation of variables. That is, we assume that the solution u is separable, i.e., of the form $u(x, y) = X(x)Y(y)$ where $X : (0, a) \rightarrow \mathbb{R}$ and $Y : (0, b) \rightarrow \mathbb{R}$. Inserting this into the Helmholtz equation, one obtains the equation

$$X''Y + XY'' + \lambda XY = 0.$$

Furthermore, assuming that X and Y are nontrivial, the boundary conditions translate into the conditions

$$X'(0) = X'(a) = Y'(0) = Y'(b) = 0.$$

Collecting like terms, we get from the first equation the identity

$$\frac{X''}{X} = -\frac{Y''}{Y} - \lambda.$$

Since the left-hand side of this equation depends only on x , and the right-hand side depends only on y , both have to be constant. If we call this constant μ , then the left-hand side of this equation, together with the boundary conditions for X , shows that all solutions are of the form

$$X(x) = A_m \cos \frac{m\pi x}{a} \quad \text{with} \quad \mu = -\frac{m^2\pi^2}{a^2}$$

for some $m \in \mathbb{N}_0$. Likewise, the right-hand side of the equation implies

$$\frac{Y''}{Y} = -\lambda - \mu,$$

and in combination with the boundary conditions for Y this yields

$$Y(y) = B_n \cos \frac{n\pi y}{b} \quad \text{with} \quad -\lambda - \mu = -\frac{n^2\pi^2}{b^2}$$

for some $n \in \mathbb{N}_0$. Combining these two equations for μ produces eigenvalues of the form

$$\lambda_{m,n} = \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right) \pi^2 \quad \text{where} \quad m, n \in \mathbb{N}_0,$$

and the corresponding eigenfunctions are given by multiples of the functions

$$\varphi_{m,n}(x, y) = \cos \frac{m\pi x}{a} \cos \frac{n\pi y}{b}.$$

To summarize, using separation of variables, we have found an infinite set of eigenvalues and corresponding eigenfunctions. However, at this point it is not clear whether this is a complete list of all the possible eigenvalues and eigenfunctions. ■

Example 2.69 (Dirichlet boundary conditions in two dimensions). If we consider the two-dimensional rectangular domain $\Omega = (0, a) \times (0, b)$ in \mathbb{R}^2 subject to homogeneous Dirichlet boundary conditions, separation of variables shows that the multiples of

$$\varphi_{m,n}(x, y) = \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b}$$

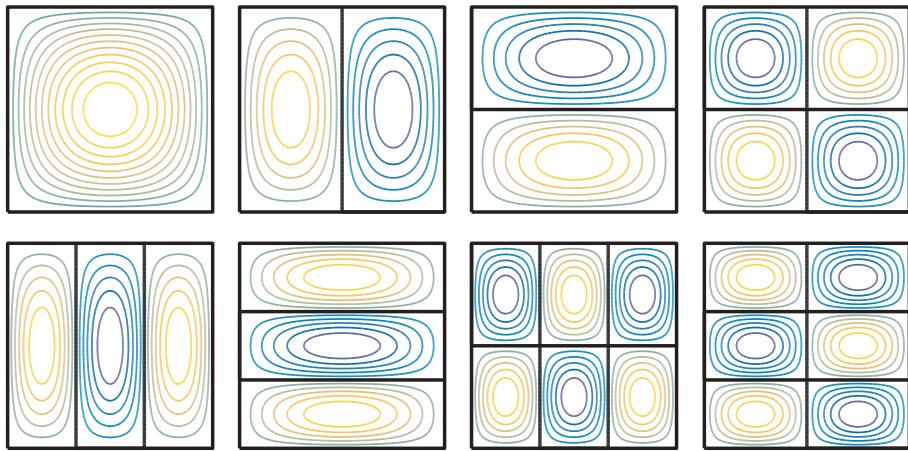


Figure 2.21. Eigenfunctions of the Laplacian on the unit square, subject to homogeneous Dirichlet boundary conditions. The eigenfunctions correspond to the first eight eigenvalues (counting multiplicities) which are given by $2\pi^2, 5\pi^2, 5\pi^2, 8\pi^2, 10\pi^2, 10\pi^2, 13\pi^2$, and $13\pi^2$, from top left to bottom right, and are given by $\varphi_{1,1}, \varphi_{2,1}, \varphi_{1,2}, \varphi_{2,2}, \varphi_{3,1}, \varphi_{1,3}, \varphi_{3,2}$, and $\varphi_{2,3}$, respectively. In the images, yellow corresponds to function values close to 1, and blue corresponds to values close to -1 . The zero set of the eigenfunctions is shown as a union of black contour lines.

are eigenfunctions, and the corresponding eigenvalues are given by

$$\lambda_{m,n} = \left(\frac{m^2}{a^2} + \frac{n^2}{b^2} \right) \pi^2, \quad \text{for } m, n \in \mathbb{N}.$$

Note that the structure of the eigenfunctions is very much analogous to the ones for the Neumann boundary value case. Yet, in some sense there are fewer eigenvalues, since the indices m and n cannot be zero. ■

One could easily extend the above list to other types of boundary conditions, and some of this will be done in the exercises for this chapter. We would like to emphasize the fact that the method of separation of variables can so far only be applied to rectangular domains, but we will relax this requirement soon.

Though it does not follow from what has been presented so far, the set of eigenvalues and eigenfunctions listed in each of the above two examples is complete. However, the criteria for completeness of eigenfunctions for two-dimensional domains are by no means as simple as the ones for a one-dimensional domain, where all we had to do was count zeros. In fact in dimension two or higher, the zero set does not tell us all that much about the eigenfunctions.

In order to illustrate this fact, we close this section with a few comments about the geometry of the eigenfunctions in the case of homogeneous Dirichlet boundary conditions. In Figure 2.21 we show contour plots of the eigenfunctions corresponding to the first eight eigenvalues, counting multiplicities. Notice that among these eight eigenfunctions, all but two correspond to eigenvalues with a two-dimensional eigenspace. Only the eigenvalues $2\pi^2$ and $8\pi^2$ have a one-dimensional eigenspace. The associated eigenfunctions are shown as surface plots in Figure 2.22.

At first glance, the images of the first eigenfunctions seem to imply that the zero set of the eigenfunctions always divides the unit square into a regular checkerboard pattern.

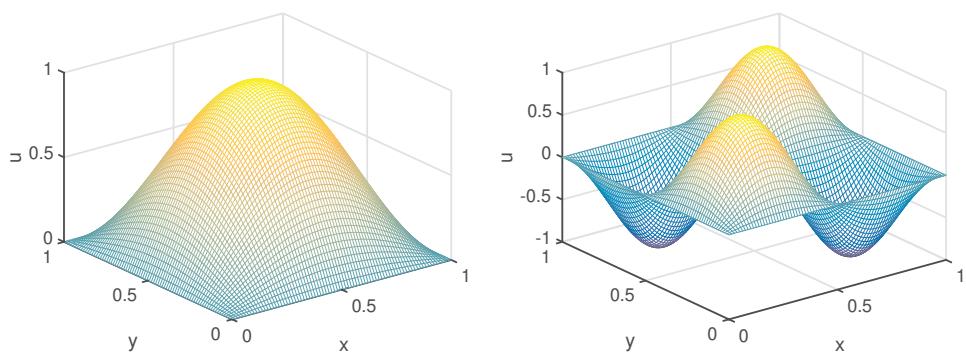


Figure 2.22. Surface plots of specific eigenfunctions of the Laplacian on the unit square, subject to homogeneous Dirichlet boundary conditions. The left panel shows the eigenfunction $\varphi_{1,1}$ corresponding to the eigenvalue $2\pi^2$, the right panel shows the eigenfunction $\varphi_{2,2}$, which corresponds to the eigenvalue $8\pi^2$. Notice that among the eigenfunctions shown in Figure 2.21, these are the only two eigenfunctions which correspond to simple eigenvalues.

However, nothing could be further from the truth. For the eigenvalues with multiplicity two, any linear combination of two eigenfunctions is again an eigenfunction for the same eigenvalue, and these superpositions can have complicated zero sets. This is illustrated in Figures 2.23 and Figures 2.24, where we present a variety of eigenfunctions for the eigenvalues $29\pi^2$ and $145\pi^2$, respectively. Figure 2.25 shows surface plots for two of them.

Despite the complexity seen in the geometries of the eigenfunctions of the Laplacian on a rectangular domain, it is actually possible to obtain significant theoretical insight into eigenfunctions and their geometry. This question will be revisited later on in Section 4.4. More precisely, in Section 4.4.2 we will show that Sturm-Liouville problems can be generalized to higher dimensions, that they still give rise to complete orthogonal sets, and that the associated eigenvalues can be determined through an associated optimization problem. The latter will then be used in Section 4.4.3 to derive a wide variety of surprisingly general results about eigenvalues, their asymptotic growth rate, and the geometry of the eigenfunctions.

2.4.2 • Completeness of the Set of Eigenfunctions

In the previous section, we established an infinite set of eigenvalues and corresponding eigenfunctions for the Helmholtz equation on a two-dimensional domain for two different sets of boundary conditions. However, showing that eigenfunctions form a complete orthogonal set is much more difficult when we move beyond one-dimensional regular Sturm-Liouville problems as in Theorem 2.37. Furthermore, there is a breakdown of the nice structure of eigenvalues and eigenfunctions once we move beyond this context. We have already seen this breakdown in Example 2.43 in the case of boundary value problems with periodic boundary conditions. In particular, for a complete set of eigenfunctions in the context of regular Sturm-Liouville boundary problems for one-dimensional domains, every eigenfunction corresponds to one eigenvalue uniquely up to multiplicative factors, and the k -th eigenfunction has exactly $k - 1$ internal zeros, where $k \in \mathbb{N}$. In contrast, in the case of periodic boundary conditions, there is usually more than one linearly independent eigenfunction associated with a given eigenvalue, and in addition more than one linearly independent eigenfunction has the same number of internal zeros. Furthermore, we could not directly establish the completeness of the Fourier family, and instead had to

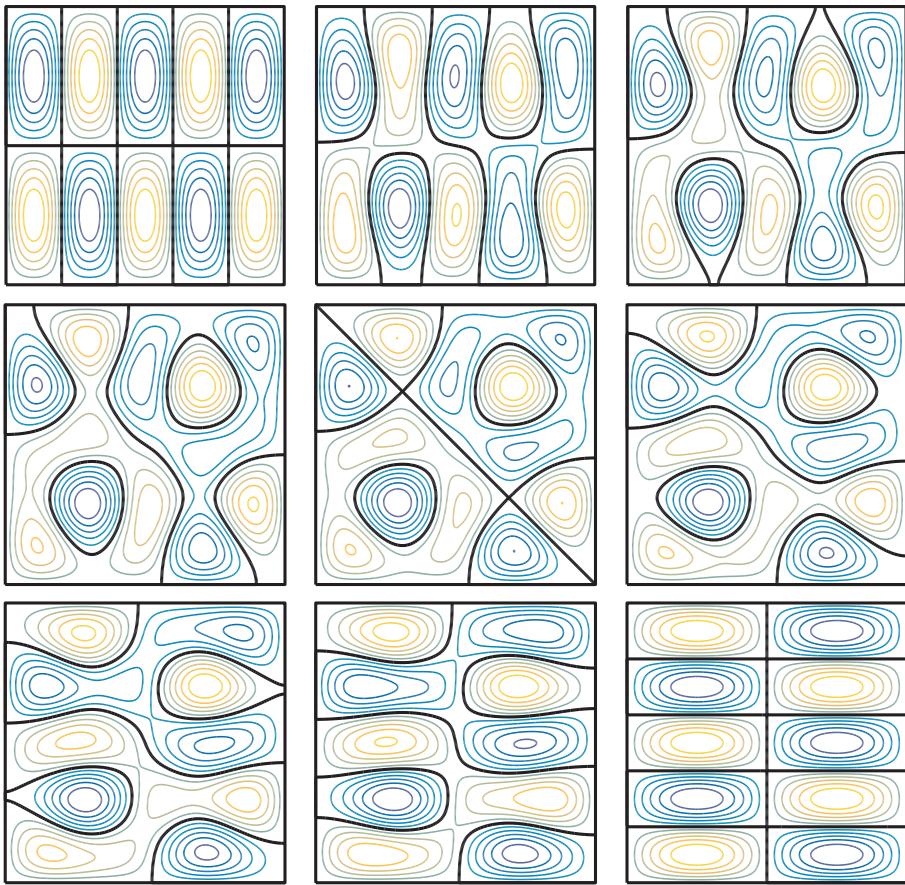


Figure 2.23. Variety of eigenfunctions of the Laplacian on the unit square, subject to homogeneous Dirichlet boundary conditions. The eigenfunctions are all for the eigenvalue $29\pi^2$ and are of the form $c \sin(5\pi x) \sin(2\pi y) + \sqrt{1-c^2} \sin(2\pi x) \sin(5\pi y)$ for nine c -values between 0 and 1. Note that the first and the last shown eigenfunctions are $\varphi_{5,2}$ and $\varphi_{2,5}$, respectively, which are the 18-th and 19-th eigenfunctions when ordered linearly according to increasing eigenvalues. In the images, yellow corresponds to function values close to the global maximum of the eigenfunction, and blue corresponds to values close to the global minimum. The zero set of the eigenfunctions is shown as a union of black contour lines.

quote classical results in order to do so.

Once we move into higher-dimensional domains Ω , understanding eigenvalue problems and the structure of eigenfunctions become even more difficult, and no simple characterization for the completeness of a set of eigenfunctions exists. Furthermore, as the images from the last section have demonstrated, there is no simple discernible pattern that would indicate that the zero set of the eigenfunctions is related to the eigenvalue. Thus in this more general setting, we must keep our goals much more modest. In this section, we restrict our focus to establishing the following two simple facts.

- If φ and ψ are two eigenfunctions of the Helmholtz equation (2.114) on an arbitrary domain $\Omega \subset \mathbb{R}^d$ which correspond to two distinct eigenvalues λ and μ respectively with $\lambda \neq \mu$, and if both eigenfunctions satisfy either homogeneous Neumann or

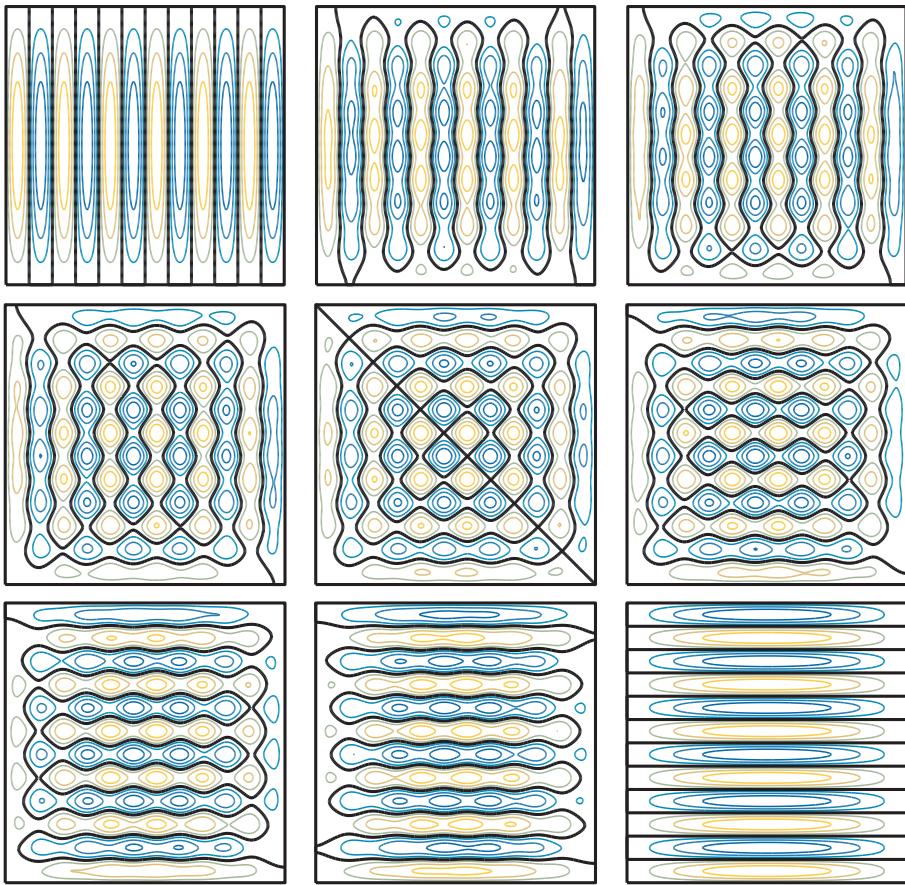


Figure 2.24. Eigenfunctions of the Laplacian on the unit square, with homogeneous Dirichlet boundary conditions. The eigenfunctions are for the eigenvalue $145\pi^2$ and of the form $c \sin(12\pi x) \sin(\pi y) + \sqrt{1-c^2} \sin(\pi x) \sin(12\pi y)$ for nine c -values between 0 and 1. The first and the last shown eigenfunctions are $\varphi_{12,1}$ and $\varphi_{1,12}$, respectively, which are the 99-th, the 100-th, the 101-st, or the 102-nd eigenfunctions when ordered linearly according to increasing eigenvalues. (The eigenfunctions $\varphi_{9,8}$ and $\varphi_{8,9}$ also correspond to the eigenvalue $145\pi^2$.) In the images, yellow corresponds to function values close to the global maximum of the eigenfunction, and blue corresponds values close to the global minimum. The zero set of the eigenfunctions is shown as a union of black contour lines.

homogeneous Dirichlet boundary conditions, then the functions φ and ψ are orthogonal with respect to the standard $L^2(\Omega)$ -scalar product.

- For the rectangular domain $\Omega = (0, a) \times (0, b)$ in \mathbb{R}^2 , the eigenfunctions of the Helmholtz equation (2.114) constructed in the previous two examples form complete orthogonal sets in $L^2(\Omega)$.

We begin by establishing the first claim. For this, note that the inner product of φ and ψ can be rewritten using the Helmholtz equation, which leads to the identity

$$\lambda \langle \varphi, \psi \rangle_{L^2(\Omega)} = \int_{\Omega} \lambda \varphi(x) \psi(x) dx = - \int_{\Omega} \Delta \varphi(x) \psi(x) dx .$$

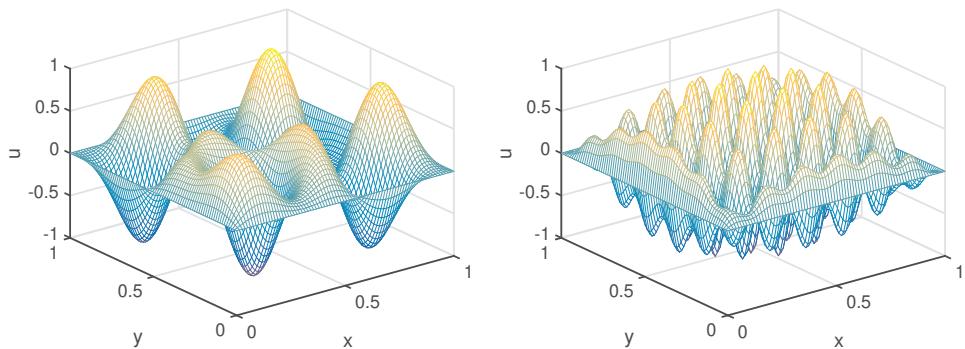


Figure 2.25. Surface plots of specific eigenfunctions of the Laplacian on the unit square, subject to homogeneous Dirichlet boundary conditions. The left panel is for the eigenvalue $29\pi^2$ and corresponds to the first plot in the second row of Figure 2.23. The right panel is for the eigenvalue $145\pi^2$ and corresponds to the first plot in the second row of Figure 2.24.

If we now apply the integration by parts formula of Theorem 1.40 twice, we obtain

$$\int_{\Omega} \Delta \varphi(x) \psi(x) dx = \int_{\Omega} \varphi(x) \Delta \psi(x) dx - \underbrace{\int_{\partial\Omega} (\varphi \nabla \psi - \psi \nabla \varphi) \cdot n ds}_{=0},$$

and using the fact that ψ satisfies the Helmholtz equation for μ finally gives

$$\lambda \langle \varphi, \psi \rangle_{L^2(\Omega)} = \int_{\Omega} \mu \varphi(x) \psi(x) dx = \mu \langle \varphi, \psi \rangle_{L^2(\Omega)}.$$

Since we have assumed that $\lambda \neq \mu$, this shows that the inner product of φ and ψ has to vanish. We would like to point out that the above argument is almost identical to the one used in Section 2.2 — one only has to replace integration by parts by the divergence theorem. In fact, the validity of the eigenfunction orthogonality is the consequence of a deeper underlying principle. One can show (using basically a computation based on the divergence theorem) that the left-hand side of the Helmholtz equation is a symmetric operator in the sense of Definition 2.32, but generalized in the obvious way to higher-dimensional domains. Thus, the orthogonality is a consequence of the symmetry of the underlying differential operator, as was demonstrated in the one-dimensional setting in Theorem 2.37.

We now turn our attention to the second claim from above. Rather than presenting the argument for both types of boundary conditions, we only consider the case of homogeneous Neumann boundary conditions discussed in Example 2.68. For this, we define the functions

$$\chi_m(x) = \cos \frac{m\pi x}{a} \quad \text{and} \quad \psi_n(y) = \cos \frac{n\pi y}{b}$$

for $m, n \in \mathbb{N}_0$. It was shown in Section 2.2 that the family $\{\chi_m\}_{m \in \mathbb{N}_0}$ is a complete orthogonal set in $L^2(0, a)$, and that $\{\psi_n\}_{n \in \mathbb{N}_0}$ is a complete orthogonal set in $L^2(0, b)$. Now let $u : \overline{\Omega} \rightarrow \mathbb{R}$ denote an arbitrary continuous function, where $\Omega = (0, a) \times (0, b)$. Then for each $x \in [0, a]$ the function $u(x, \cdot) : [0, b] \rightarrow \mathbb{R}$ is continuous, hence in $L^2(0, b)$, and therefore Theorem 2.14 implies

$$u(x, y) = \sum_{n=0}^{\infty} \beta_n(x) \psi_n(y), \tag{2.115}$$

where

$$\beta_n(x) = \frac{1}{\int_0^b \psi_n(y)^2 dy} \int_0^b u(x,y) \psi_n(y) dy.$$

Based on the definition of $\beta_n : [0,a] \rightarrow \mathbb{R}$ one can easily see that β_n is continuous, hence in $L^2(0,a)$, and we can apply Theorem 2.14 a second time to obtain

$$\beta_n(x) = \sum_{m=0}^{\infty} \alpha_{m,n} \chi_m(x), \quad (2.116)$$

where

$$\alpha_{m,n} = \frac{1}{\int_0^a \chi_m(x)^2 dx} \int_0^a \beta_n(x) \chi_m(x) dx.$$

By combining (2.115) and (2.116), one can then show that

$$u(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \alpha_{m,n} \chi_m(x) \psi_n(y),$$

in the sense of $L^2(\Omega)$ -convergence, even though we skip the details of this argument. Since the functions $\varphi_{m,n}$ from Example 2.68 are in fact given by the products $\chi_m(x) \psi_n(y)$, this shows that the family $\{\varphi_{m,n}\}_{m,n \in \mathbb{N}_0}$ can be used to approximate arbitrary continuous functions arbitrarily well. Again without proof, we can use the fact that every integrable function in $L^2(\Omega)$ can be approximated by continuous functions to within any prescribed tolerance, and this finally implies that the functions $\{\varphi_{m,n}\}_{m,n \in \mathbb{N}_0}$ form a complete orthogonal set in $L^2(\Omega)$. In addition, one obtains

$$\alpha_{m,n} = \frac{1}{\int_0^a \chi_m(x)^2 dx} \frac{1}{\int_0^b \psi_n(y)^2 dy} \int_0^a \int_0^b u(x,y) \chi_m(x) \psi_n(y) dy dx,$$

which is equivalent to

$$\alpha_{m,n} = \frac{1}{\int_{\Omega} \varphi_{m,n}(x,y)^2 d(x,y)} \int_{\Omega} u(x,y) \varphi_{m,n}(x,y) d(x,y),$$

i.e., we recover the formula for the generalized Fourier coefficients given in (2.7).

While the above derivation was only made for the case of Neumann boundary conditions, the arguments convey a larger underlying principle. Namely, it is possible in many situations to reduce the higher-dimensional situation to a sequence of one-dimensional ones. For example, for two-dimensional rectangular domains, the following result holds.

Theorem 2.70 (Complete orthogonal sets on planar rectangles). *Consider a two-dimensional rectangular domain $\Omega = (0,a) \times (0,b)$. Suppose that $\{\chi_m\}_{m \in \mathbb{N}}$ denotes a complete orthogonal set in $L^2(0,a)$, and let $\{\psi_n\}_{n \in \mathbb{N}}$ be a complete orthogonal set in $L^2(0,b)$. Define the products*

$$\varphi_{m,n}(x,y) = \chi_m(x) \psi_n(y).$$

Then the functions $\{\varphi_{m,n}\}_{m,n \in \mathbb{N}}$ form a complete orthogonal set in $L^2(\Omega)$.

The verification of this statement follows almost verbatim from our above discussion. Only the proof of orthogonality requires a new idea, and it can be established by a suitable application of Fubini's theorem. We leave it to the reader to formulate analogous results for higher-dimensional rectangular domains.

2.4.3 • Solving Time-Dependent Problems via Eigenfunctions

In many ways, eigenvalue problems such as the ones studied in the last two sections are an essential first step in understanding the behavior of specific partial differential equations or the associated operators. But they also serve as a powerful tool for solving evolution equations, i.e., partial differential equations which are time-dependent.

A general method for solving evolution equations based on eigenfunctions

In this section, we develop an alternative to separation of variables for finding solutions to partial differential equations, called the *method of eigenfunctions*. In particular, the knowledge of a complete orthogonal set of eigenfunctions for an elliptic differential operator L allows one to easily construct series solutions to associated time-dependent equations such as

$$u_t = Lu, \quad u_{tt} = Lu, \quad \text{or} \quad u_{tt} + u_t = Lu. \quad (2.117)$$

To illustrate how this can be done, let us assume that we are given an elliptic operator L , such as for example $L = \Delta$, and that the associated eigenvalue equation

$$Lu + \lambda u = 0 \quad \text{in a domain} \quad \Omega \subset \mathbb{R}^d \quad (2.118)$$

subject to suitable homogeneous boundary conditions has a complete set of eigenfunctions φ_n for $n \in \mathbb{N}$ with associated eigenvalues $\lambda_n \in \mathbb{R}$.

For the sake of definiteness, suppose now that we would like to find a series representation for the general solution $u(t, x)$, where $(t, x) \in \mathbb{R}_0^+ \times \Omega$, of the parabolic partial differential equation

$$u_t = Lu \quad \text{in the parabolic domain} \quad \mathbb{R}_0^+ \times \Omega, \quad (2.119)$$

subject to the same homogeneous boundary conditions as the ones used in the eigenvalue problem (2.118). Since we are only interested in formal series representations, let us assume that for every fixed $t \geq 0$ the function $u(t, \cdot) : \Omega \rightarrow \mathbb{R}$ is contained in the space $L^2(\Omega)$. Then since we have assumed that the functions $\{\varphi_n\}$ are complete, we can write the function in terms of a generalized Fourier series with respect to the functions $\{\varphi_n\}$. Notice the boundary conditions are preserved due to the choice of *homogeneous* boundary conditions. Moreover, the coefficients in this series representation will still depend on time t . Thus, we can express any solution u in the form

$$u(t, x) = \sum_{n=1}^{\infty} \alpha_n(t) \varphi_n(x) \quad \text{for all} \quad (t, x) \in \mathbb{R}_0^+ \times \Omega, \quad (2.120)$$

where the series is considered in a formal sense, and the coefficient functions $\alpha_n : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ are unknown. To determine these coefficients, one can substitute the series representation (2.120) into the evolution equation (2.119) to obtain

$$\sum_{n=1}^{\infty} \alpha'_n(t) \varphi_n(x) = u_t(t, x) = Lu(t, x) = \sum_{n=1}^{\infty} \alpha_n(t) L \varphi_n(x) = - \sum_{n=1}^{\infty} \lambda_n \alpha_n(t) \varphi_n(x).$$

Fix an arbitrary $\ell \in \mathbb{N}$, take the inner product of the left-most and right-most series with the function φ_ℓ , and then divide by $\langle \varphi_\ell, \varphi_\ell \rangle \neq 0$. Then the orthogonality of the eigenfunctions implies

$$\alpha'_\ell(t) = -\lambda_\ell \alpha_\ell(t) \quad \text{for all} \quad \ell \in \mathbb{N}.$$

This is a standard linear first-order differential equation for the coefficient function α_ℓ , which has the general solution $\alpha_\ell(t) = A_\ell e^{-\lambda_\ell t}$, where $A_\ell \in \mathbb{R}$ is an arbitrary constant. Altogether, this shows that the solution u of (2.119) can be written as

$$u(t, x) = \sum_{n=1}^{\infty} A_n e^{-\lambda_n t} \varphi_n(x) \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \Omega. \quad (2.121)$$

In fact, the same method also works for the other types of evolution equations listed in (2.117). In the second and the third case, the left-hand side in the differential equation for α_ℓ merely has to be replaced by $\alpha''_\ell(t)$ and $\alpha''_\ell(t) + \alpha'_\ell(t)$, respectively. In other words, rather than solving a first-order linear equation to obtain $\alpha_\ell(t)$, one now has to solve a second-order linear ordinary differential equation — and one only has to replace the term $A_n e^{-\lambda_n t}$ by the resulting general solution.

Solving parabolic equations via eigenfunctions

We now demonstrate in a few examples how this formal theoretical approach can be used in practice. As we will see, this leads to the easy solution of a variety of linear parabolic partial differential equations, both on one- and on higher-dimensional domains.

Example 2.71 (The heat equation in one dimension). As our first example we consider the heat equation

$$u_t = c^2 u_{xx} \quad \text{on } \mathbb{R}_0^+ \times \Omega \quad \text{for } \Omega = (0, a) \subset \mathbb{R},$$

subject to homogeneous Dirichlet boundary conditions. In this situation, we have to determine the eigenfunctions and eigenvalues of the elliptic operator $Lu = c^2 u_{xx}$. Clearly, this operator is just a constant multiple of the standard Laplacian which was studied in Example 2.39. This implies that the complete orthogonal set derived in the example is a complete orthogonal set for L , only the eigenvalues have to be multiplied by c^2 . Thus, the eigenfunctions and eigenvalues of L are given by

$$\varphi_n(x) = \sin \frac{n\pi x}{a} \quad \text{and} \quad \lambda_n = \frac{c^2 n^2 \pi^2}{a^2} \quad \text{for all } n \in \mathbb{N}.$$

We can now immediately substitute these expressions into the solution representation (2.121), and this yields

$$u(t, x) = \sum_{n=1}^{\infty} A_n e^{-c^2 n^2 \pi^2 t / a^2} \sin \frac{n\pi x}{a} \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \Omega.$$

Notice that this formula agrees with the one obtained directly through separation of variables as stated in (2.112). This time, however, we could avoid separation of variables altogether, since we relied exclusively on eigenfunction and eigenvalue information.

If instead of homogeneous Dirichlet boundary conditions we want to study the heat equation subject to Neumann conditions, all we have to do is to exchange the underlying eigenfunctions and eigenvalues. This time, we can use Example 2.40, which shows that the operator L has eigenfunctions and eigenvalues given by

$$\varphi_n(x) = \cos \frac{n\pi x}{a} \quad \text{and} \quad \lambda_n = \frac{c^2 n^2 \pi^2}{a^2} \quad \text{for all } n \in \mathbb{N}_0.$$

Note that in the formulas (2.120) and (2.121) we can always change the indexing, as long as the sums extend over a complete set of eigenfunctions for the problem (2.118). Thus, in the Neumann case the solution of the heat equation is given by

$$u(t, x) = \sum_{n=0}^{\infty} A_n e^{-c^2 n^2 \pi^2 t / a^2} \cos \frac{n \pi x}{a} \quad \text{for all } (t, x) \in \mathbb{R}_0^+ \times \Omega.$$

Similarly, one can easily derive solutions for other types of homogeneous boundary conditions. ■

Example 2.72 (The heat equation in two dimensions). The above approach is not limited to one-dimensional base domains. Consider for example the heat equation on the unit square given by

$$u_t = \Delta u \quad \text{on } \mathbb{R}_0^+ \times \Omega \quad \text{for } \Omega = (0, 1)^2 \subset \mathbb{R}^2,$$

subject to the boundary conditions

$$u(t, 0, y) = u(t, 1, y) = 0 \quad \text{and} \quad u_y(t, x, 0) = u_y(t, x, 1) = 0 \quad (2.122)$$

for all $t > 0$ and $x, y \in (0, 1)$. In other words, we assume homogenous Dirichlet boundary conditions on the left and right edges of the square, and homogeneous Neumann conditions along the top and bottom edges.

As the first and really only step we have to find a complete orthogonal set of eigenfunctions for the Laplacian on the unit square, subject to the above boundary conditions. This can be accomplished using the method of separation of variables, by trying to find eigenfunctions of the form $\varphi(x, y) = X(x)Y(y)$. As we have done many times already, plugging this identity into the eigenvalue partial differential equation $\Delta\varphi + \lambda\varphi = 0$ immediately implies that X and Y have to satisfy the ordinary differential equations

$$X'' + \mu X = 0 \quad \text{and} \quad Y'' + \nu Y = 0$$

for two constants μ and ν , which give the associated eigenvalue via $\lambda = \mu + \nu$. The above boundary conditions further yield the identities

$$X(0)Y(y) = X(1)Y(y) = 0 \quad \text{and} \quad X(x)Y'(0) = X(x)Y'(1) = 0.$$

In other words, the X -factor of the eigenfunction φ has to satisfy homogeneous Dirichlet, and the Y -factor has to satisfy homogeneous Neumann boundary conditions. Combining Examples 2.39 and 2.40 now shows that the functions

$$\varphi_{m,n}(x, y) = \sin(m \pi x) \cos(n \pi y) \quad \text{with eigenvalues} \quad \lambda_{m,n} = (m^2 + n^2) \pi^2$$

are eigenfunctions for the Laplacian subject to the boundary conditions (2.122), where the indices are chosen according to $(m, n) \in \mathbb{N} \times \mathbb{N}_0$. Finally, we can apply Theorem 2.70, which implies that these eigenfunctions form a complete orthogonal set in $L^2(\Omega)$.

With the eigenfunctions in hand, finding the solution to the heat equation is now easy. Substituting the above formulas into (2.121) immediately implies that the sought-after solution is given by

$$u(t, x, y) = \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} A_{m,n} e^{-(m^2 + n^2) \pi^2 t} \sin(m \pi x) \cos(n \pi y) \quad (2.123)$$

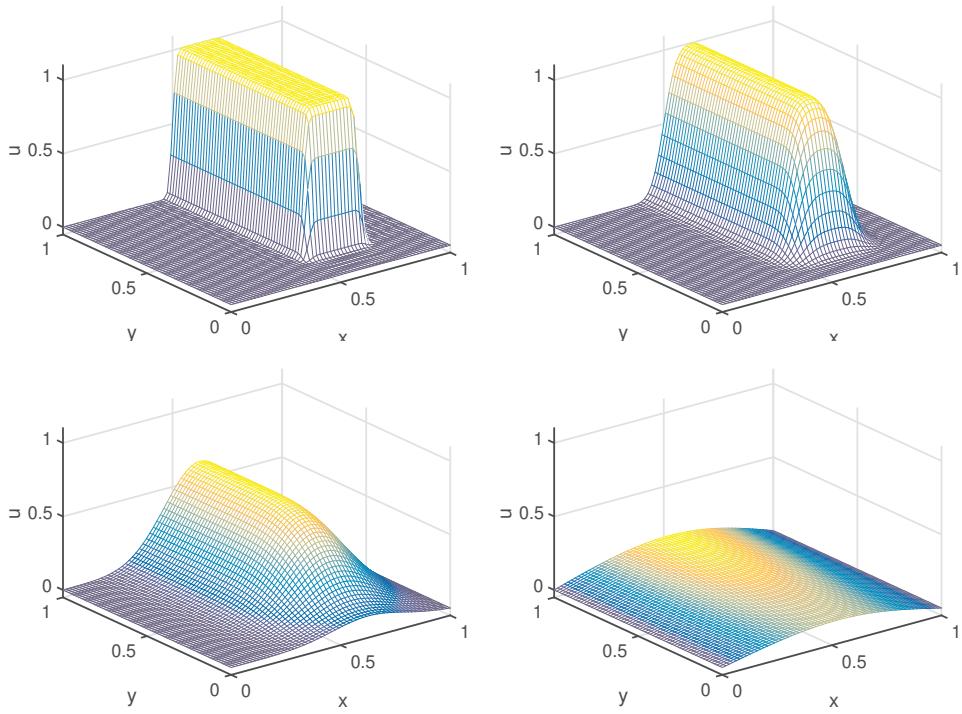


Figure 2.26. Surface plots of four solution snapshots for the heat equation considered in Example 2.72. The images show the functions $u(t, \cdot, \cdot)$ defined in (2.123) which correspond to the discontinuous initial condition f given in (2.124). From top left to bottom right the panels are for times $t = 10^{k-5}$, where $k = 1, 2, 3, 4$.

for all $t \geq 0$ and $(x, y) \in \Omega$.

To illustrate the above series solution, we now consider the heat equation subject to the specific initial condition

$$u(0, x, y) = f(x, y) = \begin{cases} 1 & \text{for } (x, y) \in [1/2, 3/4] \times [1/5, 1], \\ 0 & \text{otherwise.} \end{cases} \quad (2.124)$$

Using the results on generalized Fourier series one can readily determine the coefficients $A_{m,n}$ in the series (2.123), and they are given by

$$A_{m,n} = \begin{cases} \frac{8}{5} \frac{\cos(m\pi/2) - \cos(3m\pi/4)}{m\pi} & \text{for } n=0, m \in \mathbb{N}, \\ -\frac{4\sin(n\pi/5)}{n\pi} \frac{\cos(m\pi/2) - \cos(3m\pi/4)}{m\pi} & \text{for } n \neq 0, m \in \mathbb{N}. \end{cases} \quad (2.125)$$

Four solution snapshots of the solution u are shown in Figure 2.26. One can further use the methods from the previous sections to show that the solution u is actually infinitely many times differentiable in $\mathbb{R}^+ \times \overline{\Omega}$, despite the discontinuous initial condition f . ■

The last two examples considered our tried and tested standard parabolic partial differential equation. However, the power of the eigenfunction method lies in the fact that it can easily be used to find formal series solutions for other linear parabolic equations

as well. We demonstrate this now for the linearized version of a fourth-order parabolic equation which will be discussed in considerable more detail later on.

Example 2.73 (Linearized Cahn-Hilliard equation). We consider the time-dependent fourth-order linear partial differential equation

$$u_t = -\Delta(\varepsilon^2 \Delta u + vu) \quad \text{on} \quad \mathbb{R}_0^+ \times \Omega \quad \text{for} \quad \Omega = (0, 1)^2 \subset \mathbb{R}^2,$$

which depends on two positive parameters ε and v . According to our classification in Section 1.3.3, this equation is of parabolic type, since the leading-order term on the right-hand side is the elliptic operator $-\varepsilon^2 \Delta^2$. In this example, we want to derive an explicit series representation for the solution of this equation on the unit square, subject to homogeneous Dirichlet boundary conditions on the left and right edges of Ω , and homogeneous Neumann boundary conditions along the top and bottom edges, i.e., we assume that the solution satisfies (2.122).

One can easily see that if we define $Lu = -\varepsilon^2 \Delta^2 u - v \Delta u$, the differential equation is of the abstract form $u_t = Lu$ discussed earlier. Thus, in order to find the general solution by the eigenfunction method, we need to find a complete orthogonal set of eigenfunctions of L , and their associated eigenvalues. For this, let

$$\varphi_{m,n}(x, y) = \sin(m\pi x) \cos(n\pi y) \quad \text{and define} \quad \chi_{m,n} = (m^2 + n^2)\pi^2 \quad (2.126)$$

for $(m, n) \in \mathbb{N} \times \mathbb{N}_0$. It was shown in the previous example that the functions $\varphi_{m,n}$ form a complete orthogonal set. In fact, the numbers $\chi_{m,n}$ are the associated eigenvalues for the Laplacian Δ . The crucial reason why these functions are exactly what we need in the Cahn-Hilliard situation is the fact that the operator L is a polynomial in Δ . This implies that for any $(m, n) \in \mathbb{N} \times \mathbb{N}_0$ the identity $\Delta \varphi_{m,n} = -\chi_{m,n} \varphi_{m,n}$ results in

$$\begin{aligned} L\varphi_{m,n} &= -\varepsilon^2 \Delta^2 \varphi_{m,n} - v \Delta \varphi_{m,n} = -\varepsilon^2(-\chi_{m,n})^2 \varphi_{m,n} - v(-\chi_{m,n})\varphi_{m,n} \\ &= (-\varepsilon^2 \chi_{m,n}^2 + v \chi_{m,n}) \varphi_{m,n}, \end{aligned}$$

and therefore the function $\varphi_{m,n}$ is an eigenfunction for L with associated eigenvalue

$$\lambda_{m,n} = \varepsilon^2 \chi_{m,n}^2 - v \chi_{m,n},$$

since eigenvalues are defined via the equation $L\varphi + \lambda\varphi = 0$. Since we have already established the completeness of the eigenfunctions $\varphi_{m,n}$, we are exactly in the situation discussed in the beginning of this section. This implies that the general solution to the linearized Cahn-Hilliard equation subject to the boundary conditions (2.122) is given by

$$u(t, x) = \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} A_{m,n} e^{(\varepsilon^2 \chi_{m,n}^2 - v \chi_{m,n})t} \sin(m\pi x) \cos(n\pi y) \quad (2.127)$$

for all $(t, x, y) \in \mathbb{R}_0^+ \times \Omega$, where $\chi_{m,n}$ was defined in (2.126).

This formal series can now easily be used to solve specific initial value problems of the form $u(0, x, y) = f(x, y)$. For example, if we consider the piecewise constant initial condition f defined in (2.124), the formula for the coefficients $A_{m,n}$ remains unchanged. For this solution, and the parameter values $\varepsilon = 0.02$ and $v = 0.4$, four solution snapshots are shown in Figure 2.27 at times $t = 10^{k-5}$, where $k = 1, 2, 3, 4$. Notice that in contrast to the heat equation, the solution seems to grow exponentially, and in the course of this develops a complicated geometry. We will see later that this is to be expected from this equation,

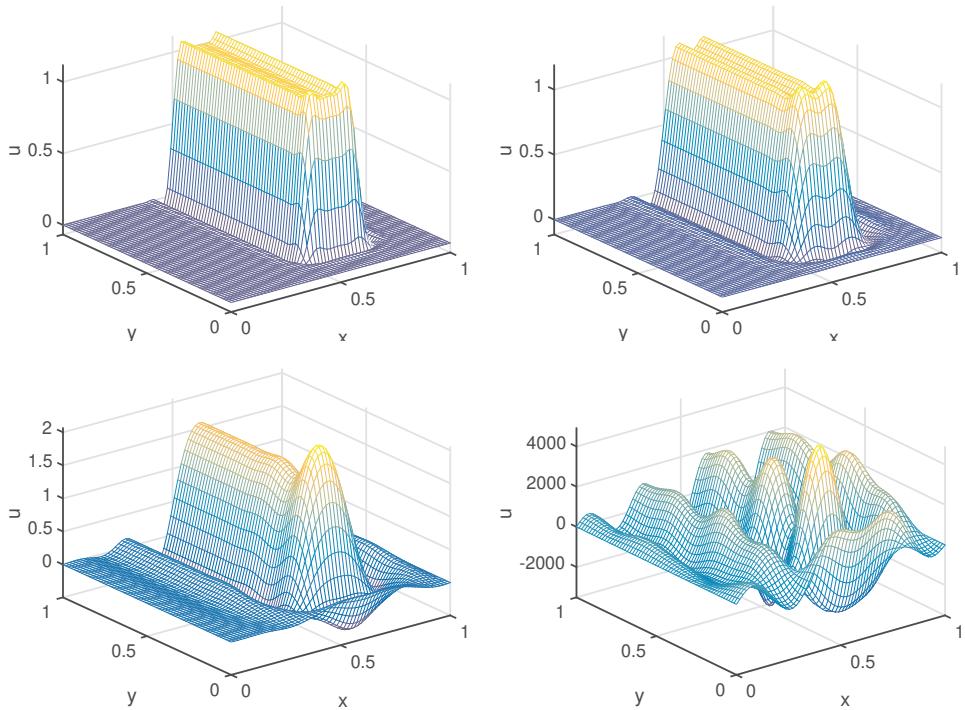


Figure 2.27. Surface plots of four solution snapshots for the linearized Cahn-Hilliard equation for $\varepsilon = 0.02$ and $\nu = 0.4$ considered in Example 2.73. The images show the functions $u(t, \cdot, \cdot)$ defined in (2.127) which correspond to the discontinuous initial condition f given in (2.124). From top left to bottom right the panels are for times $t = 10^{k-5}$, where $k = 1, 2, 3, 4$. Note the change in the vertical scale.

as it is the linearized version of a nonlinear parabolic partial differential equation which models phase separation in metal alloys.

To close this example, we briefly comment on the smoothness of the solution u defined in (2.127). Unsurprisingly, the situation is very similar to the one in the heat equation. Notice that the exponential term becomes exceedingly small as $\|(m, n)\| \rightarrow \infty$, since the dominant term is basically of the order $-\varepsilon^2 \pi^4 (m^2 + n^2)^2 t = -\varepsilon^2 \pi^4 \|(m, n)\|^4 t$. This implies that the exponential term is smaller than the reciprocal of any positive power of $\|(m, n)\|$, and as in the heat equation this can be used to show that the solution u is infinitely many times differentiable in $\mathbb{R}^+ \times \overline{\Omega}$ for any initial condition $f \in L^2(\Omega)$. ■

Solving the wave equation via eigenfunctions

As our final illustration of the eigenfunction method we turn to our standard example of a hyperbolic second-order partial differential equation, the wave equation.

Example 2.74 (Wave equation in two dimensions). Consider the wave equation

$$u_{tt} = c^2 \Delta u \quad \text{on} \quad \mathbb{R}_0^+ \times \Omega \quad \text{for} \quad \Omega = (0, 1)^2 \subset \mathbb{R}^2,$$

and subject to homogeneous Dirichlet boundary conditions on the left and right edges of Ω , and homogeneous Neumann boundary conditions along the top and bottom edges,

i.e., again subject to the conditions in (2.122). In this formulation, the equation can be considered as a model of a vibrating square membrane, which is fixed along the left and right edges, and which is allowed to move freely along the top and bottom edges.

By now the reader will be familiar with the eigenfunction method. Due to the boundary conditions (2.122) and the right-hand side operator $L = c^2 \Delta$, the eigenfunctions and eigenvalues of L are given by

$$\varphi_{m,n}(x, y) = \sin(m\pi x) \cos(n\pi y) \quad \text{and} \quad \lambda_{m,n} = c^2(m^2 + n^2)\pi^2$$

for $(m, n) \in \mathbb{N} \times \mathbb{N}_0$, and these functions form a complete orthogonal set in $L^2(\Omega)$. If we then write the solution u as a generalized Fourier series as in (2.120), with the obvious modifications due to the doubly-indexed eigenfunctions, then the unknown coefficient functions $\alpha_{m,n}(t)$ have to satisfy the second-order linear ordinary differential equation

$$\alpha''_{m,n}(t) + \lambda_{m,n} \alpha_{m,n} = 0.$$

This equation has the general solution

$$\alpha_{m,n}(t) = F_{m,n} \cos(\sqrt{\lambda_{m,n}} t) + G_{m,n} \sin(\sqrt{\lambda_{m,n}} t),$$

for arbitrary real numbers $F_{m,n}, G_{m,n} \in \mathbb{R}$, and therefore the general solution of the wave equation on the unit square subject to the boundary conditions (2.122) is given by the formal series

$$u(t, x) = \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} \left(F_{m,n} \cos(c\pi t \sqrt{m^2 + n^2}) + G_{m,n} \sin(c\pi t \sqrt{m^2 + n^2}) \right) \sin(m\pi x) \cos(n\pi y). \quad (2.128)$$

In this series, the unknown constants $F_{m,n}$ and $G_{m,n}$ can be determined through initial conditions for $u(0, \cdot, \cdot)$ and $u_t(0, \cdot, \cdot)$.

Rather than rehashing this step, we consider a specific initial value problem. According to our discussion at the beginning of this example, the wave equation subject to (2.122) models the vibrations of a square membrane, which is fixed at the left and right edges, and whose upper and lower edge are allowed to move freely. Suppose the membrane is initially at rest, but the region $[1/2, 3/4] \times [1/5, 1]$ is given a sudden uniform upward impulse. This can be modeled via the initial conditions

$$u(0, x, y) = 0 \quad \text{and} \quad u_t(0, x, y) = f(x, y) \quad \text{for all } (x, y) \in \Omega = (0, 1)^2,$$

where the function f is the piecewise constant function defined in (2.124). In this case, the coefficients in (2.128) are given by

$$F_{m,n} = 0 \quad \text{and} \quad G_{m,n} = \frac{A_{m,n}}{c\pi\sqrt{m^2 + n^2}} \quad \text{for all } (m, n) \in \mathbb{N} \times \mathbb{N}_0,$$

where the coefficients $A_{m,n}$ were defined in (2.125). The solution to this initial value problem is illustrated in Figure 2.28, for the parameter value $c = 1$. The four images in the figure show four solution snapshots $u(t, \cdot, \cdot)$ at times $t = 0.1, 0.3, 0.6$, and 1.0 . As especially the top left image shows, the solution appears to exhibit sharp corners, i.e., it is not everywhere differentiable. This demonstrates an earlier observation about solutions to the wave equation. Namely, since our initial condition for u_t is discontinuous, the solution u can only be expected to be continuous. Therefore this wave equation solution has to be considered as a weak solution. ■

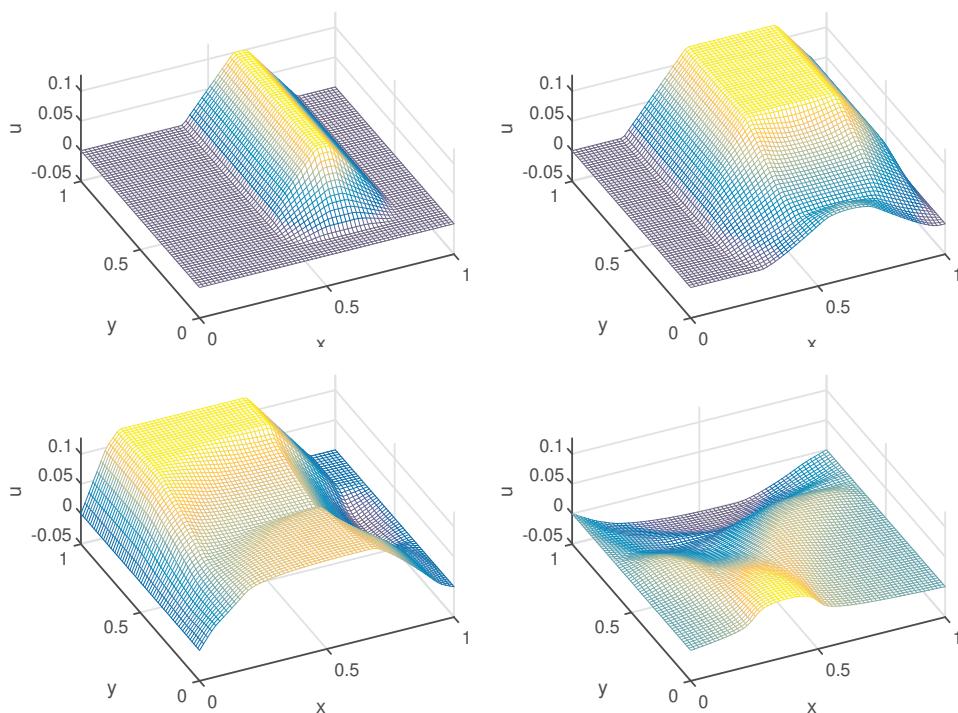


Figure 2.28. Surface plots of four solution snapshots for the wave equation considered in Example 2.74. The images show the functions $u(t, \cdot, \cdot)$ for the solution u which satisfies the two initial conditions $u(0, x, y) = 0$ and $u_t(0, x, y) = f(x, y)$ for all $(x, y) \in \Omega = (0, 1)^2$, where the function f was given in (2.124). From top left to bottom right the panels are for times $t = 0.1, 0.3, 0.6$, and 1.0 .

2.5 • Nonhomogeneous Equations and Boundary Conditions

So far, we have applied the method of separation of variables, as well as the related method of eigenfunctions, to derive explicit series solutions for a variety of partial differential equations. Notice, however, that all of the considered problems exhibited some degree of homogeneity. On the one hand, all equations treated until now were homogeneous partial differential equations. In other words, we assumed that the underlying partial differential equations were linear and that the zero function was a solution. On the other hand, all examples with time dependence, i.e., all evolution equations, were considered with associated homogeneous boundary conditions — and even for the few elliptic time-independent examples most parts of the domain boundary were subjected to homogeneous constraints. At first look, these homogeneity assumptions seem critical to the approach, since the basic idea behind the methods amount to

- finding as many solutions of the partial differential equation as possible by assuming that the solutions have product form, and then
- solving the actual specific problem using generalized Fourier series based on these solutions.

For the second step above, it is crucial that the equation is homogeneous, since otherwise a linear combination of solutions will in general no longer be a solution to the equation. In addition, homogeneity is also instrumental in the first step above. In particular, the

homogeneity of the boundary conditions on large parts of the boundary allow for the reduction of the problem to a one-dimensional Sturm-Liouville problem with a discrete set of eigenvalues. Furthermore, even in this reduced setting, on the parts of the boundary with homogeneous conditions, we use the fact that a linear combination of solutions remains a solution.

In the present section, we fully remove the restriction of homogeneity for both for the equation and for the boundary conditions. Our approach is reminiscent of the study of nonhomogeneous linear ordinary differential equations, where the set of all solutions is obtained by considering the solution set of the associated homogeneous equation, and then adding one particular solution to the nonhomogeneous problem. A similar approach can be used in the setting of partial differential equations as well.

2.5.1 • Solving Nonhomogeneous Linear Evolution Equations

In this section, we consider nonhomogeneous linear evolution equations which satisfy homogeneous boundary conditions. Solving equations of this type will be accomplished using the eigenfunction method of the last section. By the end of this section the reader will be able to construct series solutions for evolution equations of the form

$$u_t = Lu + F(t, x) \quad \text{or} \quad u_{tt} = Lu + F(t, x),$$

for arbitrary inhomogeneities $F : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}$. As before, we will assume that L denotes an elliptic differential operator such as the Laplacian. Notice that unless F is identically zero, neither of the above two equations is homogeneous, since the trivial solution no longer solves the partial differential equation.

Solving a simple nonhomogeneous heat equation

Before presenting the general method, we motivate the approach by returning to the one-dimensional heat equation, now with an added inhomogeneity. More precisely, we consider the problem

$$u_t = c^2 u_{xx} + F(t, x), \quad \text{for } (t, x) \in \mathbb{R}^+ \times (0, a),$$

subject to homogeneous Dirichlet boundary conditions $u(t, 0) = u(t, a) = 0$ for $t > 0$. We allow for a general initial condition of the form $u(0, x) = f(x)$ where the function $f \in L^2(0, a)$ is given.

In the previous section, we used the eigenfunction method to solve this heat equation problem for $F \equiv 0$. For this, we used knowledge about the complete orthogonal set of eigenfunctions for the operator $Lu = c^2 u_{xx}$ to search for solutions of the form

$$u(t, x) = \sum_{n=1}^{\infty} \alpha_n(t) \sin \frac{n\pi x}{a}, \tag{2.129}$$

where the scalar functions α_n are unknown. In the homogeneous case $F \equiv 0$ we could determine these coefficient functions through the solution of associated scalar first-order linear ordinary differential equations. The crucial ingredient for this complete decoupling of the partial differential equation into an infinite sequence of scalar equations was the fact that for this series, the terms u_t and $c^2 u_{xx}$ are still generalized Fourier series with respect to the same eigenfunctions, and we can therefore consider the resulting equations for each coefficient separately.

What if we now add a nonzero inhomogeneity F ? If as a first step it is given as a linear combination of some of the eigenfunctions of $Lu = c^2 u_{xx}$, then these new terms only modify the ordinary differential equations for the respective coefficients. This is demonstrated in the following simple example.

Example 2.75 (A simple nonhomogeneous heat equation). In this example, we consider the one-dimensional heat equation

$$u_t = c^2 u_{xx} + \sin \frac{3\pi x}{a} \quad \text{on the domain } \mathbb{R}^+ \times (0, a),$$

subject to homogeneous Dirichlet boundary conditions, and with the initial condition

$$u(0, x) = 2 \sin \frac{\pi x}{a} + 4 \sin \frac{3\pi x}{a} \quad \text{for all } 0 < x < a.$$

We look for a solution of this problem in the form (2.129). Substituting this solution representation into the slightly reformulated partial differential equation $u_t - c^2 u_{xx} = \sin(3\pi x/a)$ gives

$$\sum_{n=1}^{\infty} \left(\alpha'_n(t) + \frac{c^2 n^2 \pi^2}{a^2} \alpha_n(t) \right) \sin \frac{n\pi x}{a} = \sin \frac{3\pi x}{a}.$$

Using the abbreviation $\lambda_n = c^2 n^2 \pi^2 / a^2$, we can now match like sine terms on both sides, or in the formulation of generalized Fourier series, we can take the L^2 -inner product of both sides with respect to each of the basis functions $\varphi_\ell(x) = \sin(\ell \pi x / a)$, and then divide by $\|\varphi_\ell\|_{L^2(0,a)}^2$. For $\ell \neq 3$ the inner product of the right-hand with φ_ℓ vanishes, and one obtains the equation $\alpha'_\ell(t) + \lambda_\ell \alpha_\ell(t) = 0$. In the remaining case $\ell = 3$, one can easily verify that the above procedure yields $\alpha'_3(t) + \lambda_3 \alpha_3(t) = 1$, i.e., an *nonhomogeneous* first-order linear ordinary differential equation.

In an analogous fashion one can determine the values for $\alpha_\ell(0)$ from the initial condition. Based on its specific form, we have to have $\alpha_1(0) = 2$ and $\alpha_3(0) = 4$. Altogether, the unknown coefficient functions can be determined from the equations

$$\begin{aligned} \alpha'_1(t) + \lambda_1 \alpha_1(t) &= 0 \quad \text{with} \quad \alpha_1(0) = 2, \\ \alpha'_3(t) + \lambda_3 \alpha_3(t) &= 1 \quad \text{with} \quad \alpha_3(0) = 4, \quad \text{and} \\ \alpha'_\ell(t) + \lambda_\ell \alpha_\ell(t) &= 0 \quad \text{with} \quad \alpha_\ell(0) = 0, \quad \text{for } \ell \notin \{1, 3\}. \end{aligned}$$

The reader can easily verify that the initial value problems in the third line all have only the trivial solution $\alpha_\ell(t) = 0$, for arbitrary $\ell \notin \{1, 3\}$, and that the homogeneous equation in the first line produces the unique solution $\alpha_1(t) = 2e^{-\lambda_1 t}$. The remaining initial value problem in the second equation can be tackled using techniques from linear ordinary differential equations, such as using the method of integrating factors. This implies that its unique solution is given by

$$\alpha_3(t) = \frac{1}{\lambda_3} + \left(4 - \frac{1}{\lambda_3} \right) e^{-\lambda_3 t}.$$

Upon substitution into the series representation (2.129) for u , this produces the solution

$$u(t, x) = 2e^{-c^2 \pi^2 t / a^2} \sin \frac{\pi x}{a} + \left(\frac{a^2}{9c^2 \pi^2} + \left(4 - \frac{a^2}{9c^2 \pi^2} \right) e^{-9c^2 \pi^2 t / a^2} \right) \sin \frac{3\pi x}{a}$$

for our simple heat equation initial/boundary value problem. ■

Nonhomogeneous parabolic evolution equations and the eigenfunction method

We now generalize the above example. For this, we now consider an elliptic operator L over some domain $\Omega \subset \mathbb{R}^d$, subject to certain homogeneous boundary conditions. Moreover, as in the last section, we assume that the associated eigenvalue problem

$$Lu + \lambda u = 0 \quad \text{in} \quad \Omega \subset \mathbb{R}^d, \quad (2.130)$$

subject to the same homogeneous boundary conditions, has a complete set of eigenfunctions φ_n for $n \in \mathbb{N}$ with corresponding eigenvalues $\lambda_n \in \mathbb{R}$.

Now let $F : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}$ denote an arbitrary inhomogeneity, which usually will not be identically zero, and consider as a first case the linear parabolic partial differential equation

$$u_t = Lu + F(t, x) \quad \text{for} \quad (t, x) \in \mathbb{R}^+ \times \Omega, \quad (2.131)$$

again subject to the same homogeneous boundary conditions as the ones for L . Finally, consider an initial condition of the form $u(0, x) = f(x)$ for all $x \in \Omega$.

In order to solve the initial/boundary value problem for the above evolution equation, we begin by representing both the initial condition and the inhomogeneity in terms of the eigenfunctions of the operator L . More precisely, let

$$f(x) = \sum_{n=1}^{\infty} f_n \varphi_n(x) \quad \text{and} \quad F(t, x) = \sum_{n=1}^{\infty} F_n(t) \varphi_n(x), \quad (2.132)$$

for scalar constants $f_n \in \mathbb{R}$ and functions $F_n : \mathbb{R}^+ \rightarrow \mathbb{R}$. Recall that the coefficients in these series can be determined explicitly using the formula (2.7) given in Theorem 2.14. For these series to be guaranteed convergent¹⁷, one needs to assume that f and $F(t, \cdot)$ are contained in the space $L^2(\Omega)$, which is usually satisfied, and additionally that they satisfy the same homogeneous boundary conditions as (2.130). This boundary condition may not be satisfied, but we can still write down a formal series and later check for convergence.

In order to find the solution u to our parabolic problem, we again assume that for unknown coefficient functions $\alpha_n : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ we have the series representation

$$u(t, x) = \sum_{n=1}^{\infty} \alpha_n(t) \varphi_n(x). \quad (2.133)$$

If we now rewrite (2.131) in the form $u_t - Lu = F(t, x)$, use the fact that the functions φ_n solve the eigenvalue problem (2.130), and that the inhomogeneity F can be expanded as in (2.132), then substituting the series representation of u into the parabolic partial differential equation gives

$$\sum_{n=1}^{\infty} (\alpha'_n(t) + \lambda_n \alpha_n(t)) \varphi_n(x) = \sum_{n=1}^{\infty} F_n(t) \varphi_n(x),$$

and this implies as before that the coefficient functions have to satisfy the linear ordinary differential equations

$$\alpha'_n(t) + \lambda_n \alpha_n(t) = F_n(t) \quad \text{for all} \quad t \in \mathbb{R}_0^+ \quad \text{and} \quad n \in \mathbb{N}.$$

Moreover, if we compare the series representations for $u(0, x)$ and $f(x)$, one can immediately see that the coefficient functions also have to satisfy the initial conditions $\alpha_n(0) = f_n$.

¹⁷almost everywhere

for all $n \in \mathbb{N}$. Altogether, we have shown that the coefficient functions $\alpha_n(t)$ can be determined uniquely from the scalar initial value problems

$$\alpha'_n(t) + \lambda_n \alpha_n(t) = F_n(t) \quad \text{with} \quad \alpha_n(0) = f_n, \quad (2.134)$$

for all $t \in \mathbb{R}_0^+$ and $n \in \mathbb{N}$. In other words, assuming that the functions $F(t, x)$ and $f(x)$ are sufficiently well-behaved, i.e., as long as we have the series representations (2.132), we have reformulated our nonhomogeneous parabolic partial differential equation problem as an infinite system of first-order linear nonhomogeneous ordinary differential equations with initial conditions. Clearly this is an improvement over our original equation, since using ordinary differential equations techniques such as integrating factors can usually be used to find the solutions to these ordinary differential equations — and this in turn provides us with a series representation for u . We now demonstrate this procedure for a specific example.

Example 2.76 (A slightly more involved nonhomogeneous heat equation).

Consider the nonhomogeneous heat equation

$$u_t = c^2 u_{xx} + xt \quad \text{on the domain} \quad \mathbb{R}^+ \times (0, \pi),$$

subject to the boundary and initial conditions given by

$$u(t, 0) = u(t, \pi) = 0 \quad \text{and} \quad u(0, x) = 1,$$

respectively. This problem is exactly of the form studied above if we define the eigenfunctions as $\varphi_n(x) = \sin(nx)$ with corresponding eigenvalues $\lambda_n = c^2 n^2$ for $n \in \mathbb{N}$. The input data to the problem is given by $F(t, x) = tx$ and $f(x) = 1$, and these functions can be expanded as generalized Fourier series. More precisely, the identities in (2.132) hold with

$$f_n = \frac{2}{\pi} \int_0^\pi f(x) \sin(nx) dx = \frac{2}{\pi} \int_0^\pi \sin(nx) dx = \frac{2(1 - (-1)^n)}{n\pi},$$

as well as

$$F_n(t) = \frac{2}{\pi} \int_0^\pi F(t, x) \sin(nx) dx = \frac{2}{\pi} \int_0^\pi xt \sin(nx) dx = \frac{2t(-1)^{n+1}}{n}.$$

Plugging these expressions into the initial value problems (2.134), one can easily see that the coefficient functions $\alpha_n(t)$ have to satisfy

$$\alpha'_n(t) + c^2 n^2 \alpha_n(t) = \frac{2t(-1)^{n+1}}{n} \quad \text{with} \quad \alpha_n(0) = \frac{2(1 - (-1)^n)}{n\pi}.$$

All of these equations can be solved using the method of integrating factors. For this, multiply both sides of the differential equation by $e^{c^2 n^2 t}$, then we have

$$\frac{d}{dt} (e^{c^2 n^2 t} \alpha_n(t)) = e^{c^2 n^2 t} \alpha'_n(t) + c^2 n^2 e^{c^2 n^2 t} \alpha_n(t) = \frac{2(-1)^{n+1}}{n} t e^{c^2 n^2 t},$$

and integrating both sides from 0 to t further implies

$$\begin{aligned} e^{c^2 n^2 t} \alpha_n(t) - \alpha_n(0) &= \int_0^t \frac{d}{d\tau} (e^{c^2 n^2 \tau} \alpha_n(\tau)) d\tau = \frac{2(-1)^{n+1}}{n} \int_0^t \tau e^{c^2 n^2 \tau} d\tau \\ &= \frac{2(-1)^{n+1}}{n} \frac{1 + e^{c^2 n^2 t} (c^2 n^2 t - 1)}{c^4 n^4}. \end{aligned}$$

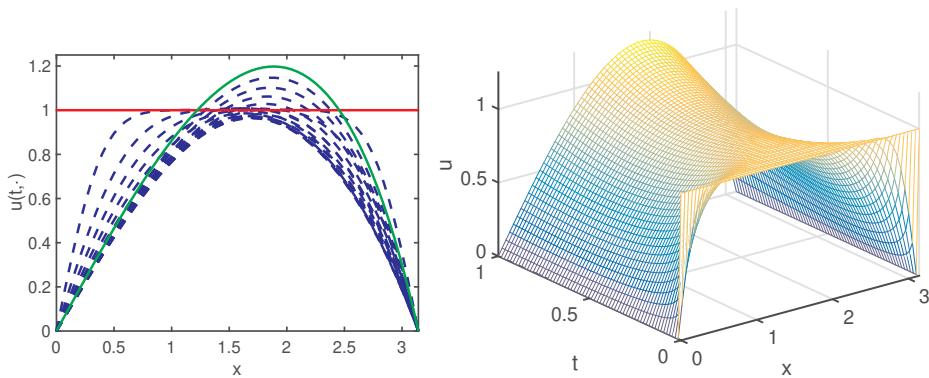


Figure 2.29. Solution of the nonhomogeneous heat equation on the domain $\Omega = (0, \pi)$ and with diffusion constant $c = 1$, as discussed in Example 2.76. In the left panel, solution snapshots $u(t, \cdot)$ are shown at times $t = 0$ (red curve), $t = 1$ (green curve), as well as fifteen times t between these two values (dashed blue curves). The image on the right shows the restriction $u : [0, 1] \times [0, \pi] \rightarrow \mathbb{R}$ as a surface plot.

With the above formula for $\alpha_n(0)$ this finally implies

$$\alpha_n(t) = \frac{2(1 - (-1)^n)}{n\pi} e^{-c^2 n^2 t} + 2(-1)^{n+1} \frac{e^{-c^2 n^2 t} + c^2 n^2 t - 1}{c^4 n^5},$$

which together with (2.133) and $\varphi_n(x) = \sin(nx)$ gives an explicit representation for the solution u . Plots of the solution are shown in Figure 2.29. ■

The nonhomogeneous wave equation and resonance

Even though we only discussed the nonhomogeneous heat equation in detail, the described method can clearly also be used in other contexts. Suppose for example that in the situation discussed before, we are interested in solving the nonhomogeneous wave equation

$$u_{tt} = Lu + F(t, x) \quad \text{for } (t, x) \in \mathbb{R}^+ \times \Omega,$$

where L is an elliptic differential operator as before, and where both L and the wave equation are considered subject to the same homogeneous boundary conditions. This time, the equation needs to be supplemented by initial conditions of the form $u(0, x) = f(x)$ and $u_t(0, x) = g(x)$ for all $x \in \Omega$.

In order to solve the initial/boundary value problem for this nonhomogeneous wave equation, we again let $\varphi_n(x)$ for $n \in \mathbb{N}$ denote the complete set of eigenfunctions of L , with associated eigenvalues λ_n . Furthermore, we consider series expansions for f and F as in (2.132), but now also assume that

$$g(x) = \sum_{n=1}^{\infty} g_n \varphi_n(x) \quad \text{for real numbers } g_n \in \mathbb{R}.$$

While we leave the details of the derivation to the reader, one can easily see that if u is of the form

$$u(t, x) = \sum_{n=1}^{\infty} \alpha_n(t) \varphi_n(x),$$

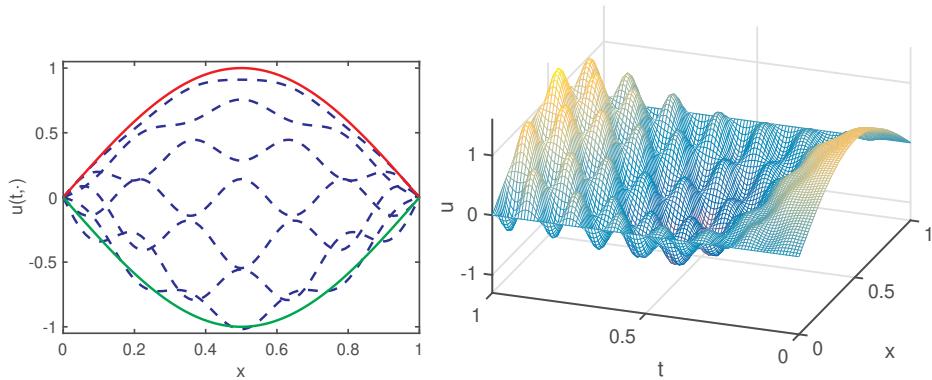


Figure 2.30. Solution of the nonhomogeneous wave equation on the domain $\Omega = (0, 1)$ as discussed in Example 2.77. In the left panel, solution snapshots $u(t, \cdot)$ are shown at times $t = 0$ (red curve), $t = 1/2$ (green curve), as well as seven times t between these two values (dashed blue curves). The image on the right shows the restriction $u : [0, 1]^2 \rightarrow \mathbb{R}$ as a surface plot.

then the function u can only solve the above initial/boundary value problem for the wave equation, if the coefficient functions $\alpha_n(t)$ satisfy the initial value problems

$$\alpha_n''(t) + \lambda_n \alpha_n(t) = F_n(t) \quad \text{with} \quad \alpha_n(0) = f_n \quad \text{and} \quad \alpha_n'(0) = g_n.$$

As before, each of these equations is a nonhomogeneous linear ordinary differential equation, this time, however, of order two. These equations can now be solved with standard techniques from ordinary differential equations.

Rather than providing a more involved example, we illustrate the above discussion with a simple wave equation problem. Nevertheless, this example will demonstrate that wave equations can lead to solutions which grow unboundedly as $t \rightarrow \infty$.

Example 2.77 (A nonhomogeneous wave equation with resonance). Consider the nonhomogeneous wave equation

$$u_{tt} = 4u_{xx} + 50 \cos(14\pi t) \sin(7\pi x) \quad \text{on the domain} \quad \mathbb{R}^+ \times (0, 1),$$

subject to the boundary and initial conditions given by

$$u(t, 0) = u(t, 1) = 0, \quad \text{as well as} \quad u(0, x) = \sin(\pi x) \quad \text{and} \quad u_t(0, x) = 0,$$

respectively. This problem is of the above form with $\varphi_n(x) = \sin(n\pi x)$ and $\lambda_n = 4\pi^2 n^2$ for $n \in \mathbb{N}$. The input data is chosen in such a way that $F_7(t) = 50 \cos(14\pi t)$ and $f_1 = 1$, while all other coefficients F_n , f_n , and g_n are zero. This implies that all but two coefficient functions $\alpha_n(t)$ satisfy a homogeneous second-order linear ordinary differential equation with initial position and initial velocity equal to zero. In other words, all of these functions have to be identically zero. The only exceptions occur for $n = 1$ and $n = 7$, which lead to the two initial value problems

$$\alpha_1''(t) + \lambda_1 \alpha_1(t) = 0 \quad \text{with} \quad \alpha_1(0) = 1, \quad \alpha_1'(0) = 0,$$

as well as

$$\alpha_7''(t) + \lambda_7 \alpha_7(t) = 50 \cos(14\pi t) \quad \text{with} \quad \alpha_7(0) = 0, \quad \alpha_7'(0) = 0.$$

We leave it as an easy exercise to the reader to verify that the corresponding solutions are given by

$$\alpha_1(t) = \cos(2\pi t) \quad \text{and} \quad \alpha_7(t) = \frac{25t}{14\pi} \sin(14\pi t),$$

respectively. This finally leads to the explicit solution formula

$$u(t, x) = \cos(2\pi t) \sin(\pi x) + \frac{25t}{14\pi} \sin(14\pi t) \sin(7\pi x).$$

Plots of this solution can be found in Figure 2.30. We would like to point out that this solution of the nonhomogeneous wave equation actually becomes unbounded as $t \rightarrow \infty$, due to the multiplicative factor t in the second term. This is a consequence of the phenomenon of *resonance*, since the frequency of the forcing term on the right-hand side of the partial differential equation is an integer multiple of the frequency which is inherent in the homogeneous problem. ■

The nonhomogeneous Laplace equation and time-independence

All of our examples so far have considered time-dependent evolution equations with inhomogeneities, but this was done purely for presentation purposes. In fact, the eigenfunction method can also be used to solve nonhomogeneous elliptic problems. In this case, the solution technique becomes even simpler.

To explain this, assume again that L is an elliptic operator on a domain $\Omega \subset \mathbb{R}^d$, subject to homogeneous boundary conditions, and let $\varphi_n(x)$ for $n \in \mathbb{N}$ denote the complete set of eigenfunctions of L , with associated eigenvalues λ_n . In this last subsection on nonhomogeneous equations, we are interested in solving the nonhomogeneous elliptic equation

$$Lu + f(x) = 0 \quad \text{for } x \in \Omega,$$

subject to the same homogeneous boundary conditions as L . In order to find the solution $u : \Omega \rightarrow \mathbb{R}$, we again use generalized Fourier series to express both the known inhomogeneity f and the unknown solution u in the form

$$f(x) = \sum_{n=1}^{\infty} f_n \varphi_n(x) \quad \text{and} \quad u(x) = \sum_{n=1}^{\infty} \alpha_n \varphi_n(x),$$

this time for real numbers $f_n, \alpha_n \in \mathbb{R}$. Upon substituting these expansions into the partial differential equation one obtains

$$\sum_{n=1}^{\infty} (-\lambda_n \alpha_n + f_n) \varphi_n(x) = 0, \quad \text{and therefore} \quad \lambda_n \alpha_n = f_n \quad \text{for } n \in \mathbb{N}.$$

If we now assume that $\lambda_n \neq 0$ for all $n \in \mathbb{N}$, then the unique solution u is given by

$$u(x) = \sum_{n=1}^{\infty} f_n \lambda_n^{-1} \varphi_n(x),$$

i.e., we obtain the Fourier coefficients directly, without the need to solve any differential equation at all. Note also that if $\lambda_k = 0$ for some index $k \in \mathbb{N}$, one can only find a solution if $f_k = 0$ holds as well. In fact, one can show that in this case, the nonhomogeneous elliptic problem has infinitely many solutions.

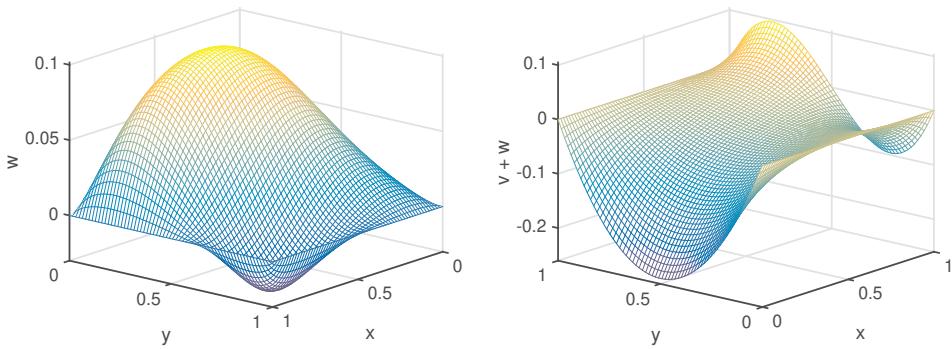


Figure 2.31. The left panel shows a surface plot of the solution w to the nonhomogeneous Laplace equation subject to homogeneous Dirichlet boundary conditions considered in Example 2.78. The right panel is a surface plot of the shifted function $v + w$ where $v(x, y) = 2xy^2 + y^2 - y - 2xy^3$, which will be further discussed in Example 2.83 in the next section.

We close our discussion of nonhomogeneous partial differential equations subject to homogeneous boundary conditions with a simple elliptic problem. This example will be of interest in the next section as well.

Example 2.78 (A nonhomogeneous Laplace problem). Consider the nonhomogeneous Laplace equation

$$\Delta w + 2 + 4x - 12xy = 0 \quad \text{on the square domain} \quad \Omega = (0, 1)^2,$$

subject to homogeneous Dirichlet boundary conditions. Notice that due to forthcoming discussions of this particular example, we used the letter w for the unknown solution, rather than our usual u .

We have already seen that for the elliptic operator $L = \Delta$ on the unit square a complete orthogonal set of eigenfunctions is given by

$$\varphi_{k,\ell}(x, y) = \sin(k\pi x)\sin(\ell\pi y) \quad \text{with eigenvalues} \quad \lambda_{k,\ell} = (k^2 + \ell^2)\pi^2,$$

for $k, \ell \in \mathbb{N}$. Using our results on generalized Fourier series, one can then show that

$$2 + 4x - 12xy = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{8}{k\ell\pi^2} \left(1 - 3(-1)^k - (-1)^\ell - 3(-1)^{k+\ell}\right) \varphi_{k,\ell}(x, y),$$

and since all eigenvalues of $L = \Delta$ are nonzero, the solution w has the series expansion

$$w(x, y) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{8(1 - 3(-1)^k - (-1)^\ell - 3(-1)^{k+\ell})}{k\ell(k^2 + \ell^2)\pi^4} \sin(k\pi x)\sin(\ell\pi y).$$

This solution is shown in the left panel of Figure 2.31. The right panel of this figure shows the shifted function $v + w$, where $v(x, y) = 2xy^2 + y^2 - y - 2xy^3$. While this might certainly seem a bit arbitrary at the moment, we will return to this graph in the next section. ■

2.5.2 • Nonhomogeneous Boundary Conditions

The last section provided a fairly thorough discussion of how nonhomogeneous linear partial differential equations can be solved using the method of separation of variables and the related method of eigenfunctions. Yet, these approaches still required the use of homogeneous boundary conditions. This last restriction will be removed in the following.

Separating the boundary from the domain via superposition

In order to fix our ideas, we begin by considering linear parabolic partial differential equations of the form

$$u_t = Lu + F(t, x) \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \Omega \subset \mathbb{R}^+ \times \mathbb{R}^d, \quad (2.135)$$

where L is again an elliptic linear operator on Ω , and the function $F : \mathbb{R}_0^+ \times \Omega \rightarrow \mathbb{R}$ denotes an arbitrary inhomogeneity. Furthermore, suppose that we study (2.135) subject to boundary conditions of the form

$$Bu = g(t, x) \quad \text{on the boundary } \partial\Omega \quad \text{and for all } t > 0, \quad (2.136)$$

where $g : \mathbb{R}^+ \times \partial\Omega \rightarrow \mathbb{R}$ denotes an arbitrary function, and the linear operator B encodes the specific type of boundary condition. For example, in the case of Dirichlet boundary conditions one uses the definition $Bu(t, x) = u(t, x)$, while in the Neumann case we define the operator $Bu(t, x) = (\partial u / \partial n)(t, x)$. If $g \equiv 0$, the considered boundary conditions are homogeneous, so we will assume from now on that g is not identically zero. Finally, suppose that the problem is complemented by an initial condition of the form

$$u(0, x) = f(x) \quad \text{on the domain } \Omega, \quad (2.137)$$

where $f : \Omega \rightarrow \mathbb{R}$ is at the very least square integrable.

The basic idea for studying problems of the above type can easily be formulated. Instead of trying to find the solution u directly, we try to separate the boundary conditions from the differential equation. More precisely, we try to find two separate functions v and w such that

$$u(t, x) = v(t, x) + w(t, x)$$

solves the parabolic problem (2.135), (2.136), and (2.137), but where the function v is responsible for incorporating the boundary conditions, while the function w makes sure the differential equation is satisfied.

To set this up in more detail, we first consider the boundary conditions. Assume that the functions v and w satisfy the two conditions

$$Bv = g(t, x) \quad \text{and} \quad Bw = 0 \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \partial\Omega. \quad (2.138)$$

In other words, the function v satisfies the original boundary conditions $Bv = g(t, x)$, while the function w satisfies homogeneous boundary conditions. Needless to say, under these two conditions, the function $u = v + w$ will satisfy the correct boundary conditions, since the operator B is linear.

We now turn our attention to the differential equation. Seeing that v is already incorporating the nontrivial boundary conditions, our goal is to avoid any further constraints on this function. Thus, we need to make sure that w is chosen in such a way that $u = v + w$

solves the original differential equation. Substituting the sum into (2.135) shows that one needs to have

$$v_t + w_t = Lv + Lw + F(t, x),$$

which can be rewritten as

$$w_t = Lw + (F(t, x) + Lv(t, x) - v_t(t, x)) \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \Omega. \quad (2.139)$$

Notice that since the function v will be known explicitly, in fact, we will be able to choose it freely as long as it satisfies $Bv = g(t, x)$, this equation is a nonhomogeneous parabolic problem for the function w , but subject to homogeneous boundary conditions. The new parabolic problem for w has the same structure as the original one given in (2.135), only the inhomogeneity is in general different. Finally, in order to satisfy the correct initial condition one has to make sure that w satisfies

$$w(0, x) = f(x) - v(0, x) \quad \text{for all } x \in \Omega. \quad (2.140)$$

This discussion can be summarized as follows:

- Suppose that v is a sufficiently smooth function which satisfies the boundary conditions $Bv = g(t, x)$, and let w denote the solution to the nonhomogeneous parabolic equation (2.139) and subject to homogeneous boundary conditions $Bw = 0$, then the function $u = v + w$ solves the parabolic problem (2.135), (2.136), and (2.137), as long as w satisfies the shifted initial condition (2.140).

In other words, the principle of superposition allows us to reduce the case of *nonhomogeneous boundary conditions* to the case of *nonhomogeneous equations with homogeneous boundary conditions* which was discussed at length in the previous section.

Before bringing this abstract method to life with a number of examples, we briefly comment further on the role played by the function v . In the above discussion, the primary task of the function v is to satisfy the boundary conditions. This can usually be achieved through simple mathematical techniques such as for example taking a convex combination, or fitting a polynomial of a certain degree. Nevertheless, this is the step that usually involves some thought.

In addition, the constructed function v has to be smooth enough so that the additional inhomogeneity $Lv(t, x) - v_t(t, x)$ is still amenable to our techniques based on generalized Fourier series. In certain situations, one can get away with fairly simple functions v . If, for example, the boundary conditions (2.136) do not depend on time, one might try to find a function $v = v(x)$, which in turn would make the term v_t disappear in the new inhomogeneity. Even more daring, one might try to find v in such a way that it satisfies $Lv = 0$ as well, and then the inhomogeneity in (2.139) is the same as the original one in (2.135). We will encounter both of these situations in the following examples.

One-dimensional diffusion with constant boundary conditions

In order to demonstrate the method, we consider the diffusion of a substance in a narrow pipe or conduit of length one filled with stationary fluid. Note that in order to consider the case of a pipe with fluid in motion, we would need to include an additional transport term. As we discovered in the introduction, this can be modeled by the diffusion or heat equation in the form

$$u_t = c^2 u_{xx} \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \Omega \quad \text{with } \Omega = (0, 1), \quad (2.141)$$

where the narrow pipe is oriented along the x -axis between $x = 0$ and the point $x = 1$. For the examples in this subsection, we assume that the equation models the concentration of a pollutant in a pipe, which initially has a pollutant concentration described by a given function $f : (0, 1) \rightarrow \mathbb{R}^+$. In other words, we assume that the initial condition

$$u(0, x) = f(x) \quad \text{for all } x \in (0, 1) \quad (2.142)$$

holds. The equation and the initial condition have to be complemented by boundary conditions at the endpoints $x = 0$ and $x = 1$ of the interval. For this subsection, we suppose that the boundary conditions do not vary with time, and therefore are of the general form

$$\alpha_0 u(t, 0) - \beta_0 u_x(t, 0) = u_0 \quad \text{and} \quad \alpha_1 u(t, 1) + \beta_1 u_x(t, 1) = u_1$$

for all $t > 0$. Note that the negative sign in front of the derivative $u_x(t, 0)$ is due to the fact that mixed boundary conditions are based on the outward normal derivative.

It has already been mentioned in the introduction that different boundary conditions correspond to different modeling situations for the pollutant evolution in the pipe. In the following, we present three different situations, and explain their implications. We would like to preface this by saying that the heat equation is a very simple first model for diffusion processes, and that our conclusions have to be taken with the proverbial grain of salt.¹⁸

Example 2.79 (Pollutant in a pipe connecting two large reservoirs). For our first example we consider a pipe which is connecting two large reservoirs. Each of the reservoirs contains the pollutant at a constant concentration, and both reservoirs are considered large enough so that one can neglect the pollutant concentration change due to the attached pipe. Thus, we can assume that at the endpoints of the pipe the pollutant concentration equals the one in the respective reservoir, i.e., the function $u(t, x)$ has to satisfy constant Dirichlet boundary conditions of the form

$$u(t, 0) = u_0 \quad \text{and} \quad u(t, 1) = u_1 \quad \text{for } t > 0, \quad (2.143)$$

where u_0 and u_1 denote the pollutant concentrations in the two attached reservoirs.

In order to solve the initial/boundary value problem (2.141), (2.142), and (2.143), we first need to find a function v which incorporates the boundary conditions. Since the latter are constant, it is natural to try for a function v which only depends on the spatial variable x . Moreover, since the underlying domain is the interval $\Omega = (0, 1)$, it seems also natural to try a linear function which interpolates between the boundary values. In other words, we can choose the function

$$v(x) = u_0 + (u_1 - u_0)x \quad \text{for } x \in [0, 1].$$

In fact, this choice provides even more than we asked for. Clearly we have $v_t = 0$, but one can immediately see that also $v_{xx} = 0$. This implies that the transformed equation (2.139) is actually the same equation as (2.141). Thus, the sum $u(t, x) = v(x) + w(t, x)$ satisfies our parabolic problem if and only if the function w satisfies the heat equation (2.141) subject to homogeneous Dirichlet boundary conditions, as well as the initial condition

$$w(0, x) = f(x) - u_0 - (u_1 - u_0)x \quad \text{for all } x \in (0, 1).$$

¹⁸Come to think of it, we could formulate the next few problems for salinity in water.

But we certainly know how to solve the remaining problem. Consider the complete orthogonal set $\varphi_k(x) = \sin(k\pi x)$ for $k \in \mathbb{N}$, with associated eigenvalues for $Lu = c^2 u_{xx}$ given by $\lambda_k = c^2 k^2 \pi^2$. If we let $f_k = 2 \int_0^1 (f(x) - u_0 - u_1 x + u_0 x) \varphi_k(x) dx$ then

$$f(x) - u_0 - (u_1 - u_0)x = \sum_{k=1}^{\infty} f_k \varphi_k(x),$$

and the function

$$w(t, x) = \sum_{k=1}^{\infty} \alpha_k(t) \varphi_k(x)$$

is a solution to the above initial/boundary value problem for w as long as the coefficient functions $\alpha_k(t)$ solve the initial value problems

$$\alpha'_k(t) + c^2 k^2 \pi^2 \alpha_k(t) = 0, \quad \text{with} \quad \alpha_k(0) = f_k.$$

This finally implies that the solution u to our pipe problem has the explicit representation

$$u(t, x) = u_0 + (u_1 - u_0)x + \sum_{k=1}^{\infty} e^{-c^2 k^2 \pi^2 t} f_k \sin(k\pi x).$$

One can show that as $t \rightarrow \infty$, the series in this formula converges uniformly to zero on the interval $[0, 1]$. This shows that after some initial transient behavior, the pollutant concentration along the pipe interpolates linearly between the pollutant concentrations of the two reservoirs, regardless of the initial concentration f . ■

Example 2.80 (Pollutant in a pipe attached to a large reservoir). The second example considers a slightly modified situation. We now suppose that at its left end, the pipe is connected to a large reservoir with constant pollutant concentration u_0 . In contrast to the previous example, at the other end of the pipe pollutant is removed or added at a constant rate. This situation corresponds to the boundary conditions

$$u(t, 0) = u_0 \quad \text{and} \quad u_x(t, 1) = u_1 \quad \text{for} \quad t > 0, \tag{2.144}$$

where u_1 indicates the rate of change of the addition of pollutant for $u_1 > 0$, and $|u_1|$ denotes the rate of change of pollutant removal if $u_1 < 0$. In the case $u_1 = 0$ the right end of the pipe is insulated against pollutant transport.

As in the first example, we solve the initial/boundary value problem (2.141), (2.142), and (2.144), by determining a function v which incorporates the boundary conditions. Owing to the time-independence of the boundary conditions, our goal is again a function $v = v(x)$, and in fact we can try to find a harmonic function v with the correct boundary conditions. Integrating the equation $v_{xx} = 0$ twice and matching the behavior at interval endpoints $x = 0$ and $x = 1$ readily implies

$$v(x) = u_0 + u_1 x \quad \text{for} \quad x \in [0, 1].$$

As in the first example, the fact that $v_t = v_{xx} = 0$ shows that the transformed partial differential equation does not change, and therefore the sum $u(t, x) = v(x) + w(t, x)$ satisfies our parabolic problem if and only if the function w satisfies the original heat equation (2.141) subject to the homogeneous boundary conditions $w(t, 0) = w_x(t, 1) = 0$, as well as with the initial condition

$$w(0, x) = f(x) - u_0 - u_1 x \quad \text{for all} \quad x \in (0, 1).$$

To solve the remaining problem using the method of eigenfunctions we refer the reader back to Example 2.63. There it was shown that for $k \in \mathbb{N}_0$ the functions

$$\varphi_k(x) = \sin \frac{(2k+1)\pi x}{2}, \quad \text{with eigenvalues} \quad \lambda_k = \frac{c^2(2k+1)^2\pi^2}{4},$$

form a complete orthogonal set in $L^2(0,1)$ for the operator $Lu = c^2 u_{xx}$ subject to the homogeneous boundary conditions $u(0) = u'(1) = 0$. Furthermore, we have

$$f(x) - u_0 - u_1 x - u_0 x = \sum_{k=1}^{\infty} f_k \varphi_k(x),$$

as long as $f_k = 2 \int_0^1 (f(x) - u_0 - u_1 x) \varphi_k(x) dx$. Then the series

$$w(t, x) = \sum_{k=1}^{\infty} \alpha_k(t) \varphi_k(x)$$

is a solution to the above initial/boundary value problem for w as long as the coefficient functions $\alpha_k(t)$ solve the initial value problems

$$\alpha'_k(t) + \frac{c^2(2k+1)^2\pi^2}{4} \alpha_k(t) = 0, \quad \text{with} \quad \alpha_k(0) = f_k.$$

This finally shows that the solution u to our second pipe problem has the explicit representation

$$u(t, x) = u_0 + u_1 x + \sum_{k=0}^{\infty} e^{-c^2(2k+1)^2\pi^2 t/4} f_k \sin \frac{(2k+1)\pi x}{2}.$$

As in the first example, one can show that as $t \rightarrow \infty$, the series in this formula converges uniformly to zero on the interval $[0, 1]$. Thus, after some initial transient behavior the pollutant concentration along the pipe follows a line segment with slope u_1 and which equals u_0 at $x = 0$. Again, this long-term behavior can be observed for any choice of the initial concentration f . ■

Example 2.81 (Pipe with pollutant inflow or outflow at both ends). Our final example with constant boundary conditions considers a pipe with pollutant inflow or outflow at both ends. This situation is modeled by the boundary conditions

$$u_x(t, 0) = -u_0 \quad \text{and} \quad u_x(t, 1) = u_1 \quad \text{for} \quad t > 0, \quad (2.145)$$

where positive values of u_0 or u_1 indicate inflow at the respective end of the pipe, while negative values correspond to outflow.

In order to solve the initial/boundary value problem (2.141), (2.142), and (2.145), we again seek a function $v = v(x)$ which assumes the correct boundary conditions. This time, however, it is no longer possible to find a harmonic function with this property. In fact, the reader can easily see by twice integrating the identity $v_{xx} = 0$ that this would only be possible if $u_0 + u_1 = 0$. We therefore attempt to find a simple function which assumes the boundary conditions $v'(0) = -u_0$ and $v'(1) = u_1$, and one natural first try is a quadratic polynomial $v(x)$. One can readily verify that

$$v(x) = \frac{u_1 + u_0}{2} x^2 - u_0 x \quad \text{for} \quad x \in [0, 1]$$

is indeed such a choice. While we still have $v_t = 0$, the second derivative of v is given by $v_{xx} = u_1 + u_0$. This implies that the sum $u(t, x) = v(x) + w(t, x)$ satisfies the original parabolic problem if and only if the function w satisfies the nonhomogeneous heat equation

$$w_t = c^2 w_{xx} + c^2(u_1 + u_0) \quad \text{for all } (t, x) \in \mathbb{R}^+ \times (0, 1),$$

subject to homogeneous Neumann boundary conditions and the initial condition

$$w(0, x) = f(x) - \frac{u_1 + u_0}{2}x^2 + u_0x \quad \text{for all } x \in (0, 1).$$

Again we use the eigenfunction method to solve this remaining problem. Due to the homogeneous Neumann boundary conditions, the complete orthogonal set of eigenfunctions is given by $\varphi_k(x) = \cos(k\pi x)$ for $k \in \mathbb{N}_0$, and the associated eigenvalues of the operator $Lu = c^2 u_{xx}$ are given by $\lambda_k = c^2 k^2 \pi^2$. If we let $f_k = 2 \int_0^1 (f(x) - v(x)) \varphi_k(x) dx$, then

$$f(x) - \frac{u_1 + u_0}{2}x^2 + u_0x = \sum_{k=1}^{\infty} f_k \varphi_k(x),$$

and in particular, we have

$$f_0 = \frac{2u_0 - u_1}{6} + \int_0^1 f(x) dx.$$

If we write the sought-after solution w in the form

$$w(t, x) = \sum_{k=0}^{\infty} \alpha_k(t) \varphi_k(x),$$

then the function w is a solution to the above initial/boundary value problem as long as the coefficient functions $\alpha_k(t)$ solve the initial value problems

$$\alpha'_0(t) = c^2(u_1 + u_0), \quad \text{with } \alpha_0(0) = f_0,$$

and

$$\alpha'_k(t) + c^2 k^2 \pi^2 \alpha_k(t) = 0, \quad \text{with } \alpha_k(0) = f_k, \quad \text{for } k \in \mathbb{N}.$$

After solving these initial value problems, this finally shows that the solution u to our second pipe problem has the explicit representation

$$\begin{aligned} u(t, x) &= \frac{u_1 + u_0}{2}x^2 - u_0x + \frac{2u_0 - u_1}{6} + c^2(u_1 + u_0)t + \int_0^1 f(x) dx \\ &\quad + \sum_{k=1}^{\infty} e^{-c^2 k^2 \pi^2 t} f_k \cos(k\pi x). \end{aligned}$$

As in the previous two examples, the series in this formula converges uniformly to zero on the interval $[0, 1]$ as $t \rightarrow \infty$. Thus, for large times t the expression in the first line describes the long-term behavior of the solution. This expression consists of the time-independent profile function

$$p(x) = \frac{u_1 + u_0}{2}x^2 - u_0x + \frac{2u_0 - u_1}{6}$$

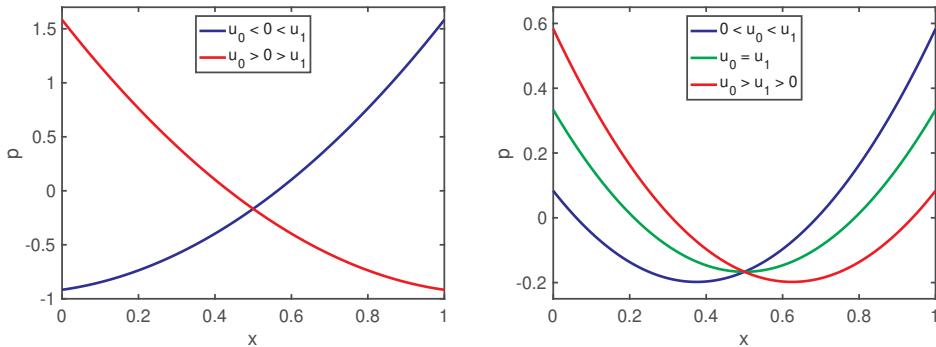


Figure 2.32. Sample images of the profile function $p(x)$ in Example 2.81, which describes the relative pollutant distribution along the pipe for large times. In the left image, the curves are for the two cases $u_0 < 0 < u_1$ (blue curve) and $u_0 > 0 > u_1$ (red curve), while the right image is for $0 < u_0 < u_1$ (blue curve), $u_0 = u_1$ (green curve), as well as $u_0 > u_1 > 0$ (red curve). In all cases, we use $u_0 + u_1 = 4$.

and the spatially constant linear growth term

$$g(t) = c^2(u_1 + u_0)t + \int_0^1 f(x)dx.$$

The function $g(t)$ is an illustration of the fact that the net amount of inflow or outflow determines whether the overall pollutant concentration in the pipe increases or decreases. While for $u_1 + u_0 > 0$ one observes an increase, for $u_1 + u_0 < 0$ the pollutant concentration decreases linearly. In the balanced case $u_1 + u_0 = 0$, the net pollutant change is zero.

More interesting is the profile function $p(x)$. For large times, this time-independent function describes the relative pollutant distribution along the pipe. Some sample graphs for the net inflow case $u_1 + u_0 > 0$ and for different choices of u_0 and u_1 are shown in Figure 2.32. The left image illustrates the qualitative form of the profile function $p(x)$ for the cases $u_0 < 0 < u_1$ and $u_0 > 0 > u_1$ in blue and red, respectively. In both of these cases, the profile is monotone in x . If, on the other hand, both boundary values u_0 and u_1 are positive, then the profile exhibits a unique minimum in the interval $(0, 1)$. This is shown in the right panel, which depicts the cases $0 < u_0 < u_1$, $u_0 = u_1$, and $u_0 > u_1 > 0$ in blue, green, and red, respectively. It is worth noting that also in this last example, the actual form of the initial condition does not affect the long-term relative pollutant distribution within the pipe. ■

Heat evolution under non-constant boundary conditions

The approach used in the last three examples can easily be extended to cover the case of non-constant, i.e., time-varying, boundary conditions, and this will be illustrated in the current brief subsection, in which we consider a narrow metal rod with time-varying temperatures on the two ends of the rod, and we use the heat equation to recover temperature information along the entire rod. To model this situation, we consider the one-dimensional heat equation

$$u_t = c^2 u_{xx} \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \Omega \quad \text{with } \Omega = (0, 1), \quad (2.146)$$

subject to the time-varying Dirichlet boundary conditions

$$u(t, 0) = u_0(t) \quad \text{and} \quad u(t, 1) = u_1(t) \quad \text{for } t > 0, \quad (2.147)$$

with an initial condition of the form

$$u(0, x) = f(x) \quad \text{for all } x \in (0, 1). \quad (2.148)$$

To apply the method outlined in the beginning of the section, one needs to find a time-varying function $v(t, x)$ which satisfies the boundary conditions. Since there are no further constraints on v , it is natural to choose $v(t, \cdot)$ as the straight-line connection between $u_0(t)$ and $u_1(t)$, i.e., we can set

$$v(t, x) = u_0(t) + (u_1(t) - u_0(t))x \quad \text{for all } t > 0 \quad \text{and} \quad x \in (0, 1).$$

Notice that $v_t(t, x) = u'_0(t) + (u'_1(t) - u'_0(t))x$ and $v_{xx}(t, x) = 0$, and therefore this choice of v will lead to a nonhomogeneous transformed equation.

We now try to find a solution of the form $u(t, x) = v(t, x) + w(t, x)$. This superposition solves the above parabolic problem, as long as the function w solves the nonhomogeneous parabolic partial differential equation

$$w_t = c^2 w_{xx} - u'_0(t) - (u'_1(t) - u'_0(t))x \quad \text{for all } (t, x) \in \mathbb{R}^+ \times \Omega,$$

subject to homogeneous Dirichlet boundary conditions and the initial condition

$$w(0, x) = f(x) - u_0(0) - (u_1(0) - u_0(0))x \quad \text{for all } x \in (0, 1).$$

The boundary conditions associated with w are homogeneous Dirichlet boundary conditions, and therefore the method of eigenfunctions can be applied with the complete orthogonal set $\varphi_k(x) = \sin(k\pi x)$ with eigenvalues $\lambda_k = c^2 k^2 \pi^2$, for all $k \in \mathbb{N}$. If we let

$$f(x) = \sum_{k=1}^{\infty} f_k \varphi_k(x)$$

and note that

$$1 = \sum_{k=1}^{\infty} \frac{1 - (-1)^k}{k\pi} \varphi_k(x) \quad \text{and} \quad x = \sum_{k=1}^{\infty} \frac{2(-1)^{k+1}}{k\pi} \varphi_k(x),$$

then the function

$$w(t, x) = \sum_{k=1}^{\infty} \alpha_k(t) \varphi_k(x)$$

solves the above nonhomogeneous parabolic problem for w if and only if the coefficient functions $\alpha_k(t)$ solve the ordinary differential equation problems

$$\alpha'_k(t) + c^2 k^2 \pi^2 \alpha_k(t) = -u'_0(t) \frac{1 - (-1)^k}{k\pi} - (u'_1(t) - u'_0(t)) \frac{2(-1)^{k+1}}{k\pi} \quad (2.149)$$

with the initial condition

$$\alpha_k(0) = f_k - u_0(0) \frac{1 - (-1)^k}{k\pi} - (u_1(0) - u_0(0)) \frac{2(-1)^{k+1}}{k\pi}. \quad (2.150)$$

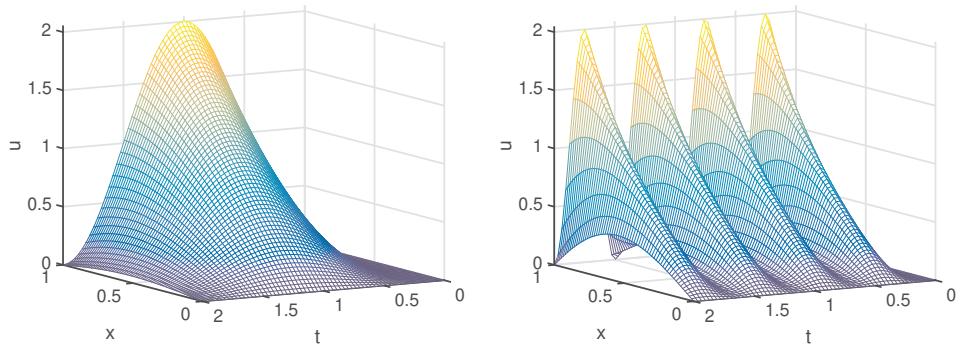


Figure 2.33. Heat diffusion with periodic temperature changes. The two panels illustrate solutions u to the parabolic problem discussed in Example 2.82, which models the evolution of heat in a thin rod of length one. While the left endpoint $x = 0$ of the rod is kept at constant temperature zero, the right endpoint $x = 1$ experiences the time-varying temperature $u_1(t) = 1 - \cos(\nu t)$. The left panel shows the solution for $\nu = \pi$, and the right panel is for $\nu = 4\pi$.

In principle, solving this infinite family of initial value problems leads to the explicit derivation of the coefficient functions $\alpha_k(t)$, and therefore to formulas for both w and u . Note, however, that this procedure is considerably more involved than the constant boundary condition case discussed in the previous subsection, where at most one of the ordinary differential equations was nonhomogeneous.

Rather than presenting explicit formulas for the solutions of (2.149) and (2.150), we close this subsection by considering one simple example.

Example 2.82 (Heat diffusion with periodic temperature changes). We consider the heat equation problem (2.146), (2.147), and (2.148) subject to the initial condition $f(x) = 0$ for all $x \in (0, 1)$, with diffusion constant $c = 1$, and for the specific Dirichlet boundary condition functions

$$u_0(t) = 0 \quad \text{and} \quad u_1(t) = 1 - \cos(\nu t) \quad \text{for all } t > 0,$$

where $\nu > 0$ denotes a real parameter. This initial/boundary value problem models the evolution of heat in a rod which is initially at the uniform temperature zero, whose left end is kept at this zero temperature for all positive times, and whose right end is exposed to periodic temperature fluctuations. Periodic temperature fluctuations occur naturally in many situations, such as daily temperature variations and seasonal temperature variation, depending on which time scale being considered.

In this specific situation, the infinite number of ordinary differential equation initial value problems given in (2.149) and (2.150) simplify to the equations

$$\alpha'_k(t) + k^2 \pi^2 \alpha_k(t) = \frac{2\nu(-1)^k}{k\pi} \sin(\nu t) \quad \text{with} \quad \alpha_k(0) = 0$$

for all $k \in \mathbb{N}$. We leave it to the reader to verify that these initial value problems have the solutions

$$\alpha_k(t) = \frac{2\nu k \pi (-1)^k}{\nu^2 + k^4 \pi^4} \sin(\nu t) + \frac{2\nu^2 (-1)^k}{k\pi (\nu^2 + k^4 \pi^4)} (e^{-k^2 \pi^2 t} - \cos(\nu t)).$$

This shows that the solution u to the heat equation problem (2.146), (2.147), and (2.148) and with the above specific choices of initial and boundary values is given by

$$\begin{aligned} u(t, x) = & (1 - \cos(\nu t))x + \sum_{k=1}^{\infty} \left(\frac{2\nu k \pi (-1)^k}{\nu^2 + k^4 \pi^4} \sin(\nu t) \right. \\ & \left. + \frac{2\nu^2 (-1)^k}{k \pi (\nu^2 + k^4 \pi^4)} (e^{-k^2 \pi^2 t} - \cos(\nu t)) \right) \sin(k \pi x). \end{aligned}$$

This solution is illustrated in Figure 2.33. In the left panel we show u for $\nu = \pi$, while the right image contains the solution for the parameter value $\nu = 4\pi$. Notice that in the former case, the periodic temperature changes at the $x = 1$ endpoint occur slowly enough that the zero temperature from the $x = 0$ endpoint can more or less penetrate the length of the rod. In contrast, for $\nu = 4\pi$ the temperature changes at the right endpoint happen so fast that the temperature in the middle of the rod stays positive for all sufficiently large times. ■

Harmonic functions with varying boundary values

Our above discussion of removing the homogeneity requirement from the boundary conditions does not rely on the equation being parabolic. For hyperbolic and elliptic problems one can also separate the boundary conditions from the differential equation via superposition, and then transform the original problem to a new nonhomogeneous problem with homogeneous boundary conditions. We now demonstrate this approach for a specific elliptic problem.

To this end, we return to the study of harmonic functions, i.e., to solutions of the Laplace equation. We showed in Corollary 1.47 of the introduction that the elliptic boundary value problem

$$\Delta u = f \quad \text{in } \Omega, \quad u = g \quad \text{on } \partial\Omega,$$

can only have a unique solution, if it has a solution at all. This result is true for arbitrary continuous functions $f : \Omega \rightarrow \mathbb{R}$ and $g : \partial\Omega \rightarrow \mathbb{R}$, as long as $\Omega \subset \mathbb{R}^d$ is a bounded open domain. For the special case $f \equiv 0$, this result states that for any boundary function g , there exists at most one harmonic function u which equals g on the boundary Ω .

We now show how such functions can be computed explicitly in terms of series solutions. For this, let us assume that we can find a function $v : \bar{\Omega} \rightarrow \mathbb{R}$ which is at least twice continuously differentiable and which satisfies

$$v(x) = g(x) \quad \text{for all } x \in \partial\Omega.$$

We would like to emphasize that *the only condition* on the smooth function v is the fact that it *satisfies the boundary conditions*, it does not have to satisfy any partial differential equation.

We now seek a solution u to the original problem in the form $u = v + w$. Then the reader can easily verify that u solves the original problem if and only if the function w solves the elliptic boundary value problem

$$\Delta w = f - \Delta v \quad \text{in } \Omega, \quad w = 0 \quad \text{on } \partial\Omega.$$

This is a nonhomogeneous problem subject to homogeneous boundary conditions, and it can be solved as in the previous section. We illustrate this procedure with a simple example.

Example 2.83 (A harmonic function with nonzero boundary conditions). Consider the homogeneous Laplace equation

$$\Delta u = 0 \quad \text{on the square domain } \Omega = (0, 1)^2,$$

subject to the Dirichlet boundary conditions

$$u(x, 0) = u(x, 1) = 0, \quad u(0, y) = y(y - 1), \quad u(1, y) = y(y - 1)(1 - 2y),$$

for all $x, y \in (0, 1)$. In order to determine a series representation for the solution, we need to define a smooth function v which has the correct boundary values. Since both $y(y - 1)$ and $y(y - 1)(1 - 2y)$ vanish for $y = 0$ and $y = 1$, we can use the convex combination

$$v(x, y) = (1 - x)y(y - 1) + x y(y - 1)(1 - 2y) = 2xy^2 + y^2 - y - 2xy^3.$$

One can easily verify that $\Delta v(x, y) = 2 + 4x - 12xy$. Thus, the function $u = v + w$ satisfies the above boundary value problem if and only if w is a solution of

$$\Delta w + 2 + 4x - 12xy = 0 \quad \text{on the square domain } \Omega = (0, 1)^2,$$

and subject to homogeneous Dirichlet boundary conditions. In fact, this nonhomogeneous problem was already solved in Example 2.78. This shows that the harmonic function u with the above boundary values is given by

$$u(x, y) = 2xy^2 + y^2 - y - 2xy^3 + \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{8(1 - 3(-1)^k - (-1)^\ell - 3(-1)^{k+\ell})}{k\ell(k^2 + \ell^2)\pi^4} \sin(k\pi x) \sin(\ell\pi y).$$

This solution was shown in the right panel of Figure 2.31. From this image, one can see that the function u attains both its global maximum and its global minimum on the boundary $\partial\Omega$, meaning that u satisfies the maximum principle. ■

2.5.3 • A Deeper Look: The Laplace Transform

All of the solution methods in this chapter have concentrated on finding solutions of nonhomogeneous partial differential equations on bounded domains. However, it is frequently more appropriate to consider such equations on unbounded domains, and this clearly requires a new method for finding solutions. To this end, this section introduces the Laplace transform without any assumption of prior knowledge of this topic. While the transform is useful in many different situations, we give only a very straightforward and brief introduction, with an emphasis on how it can be used to solve partial differential equations on the half line $\Omega = (0, \infty)$. We also point out that the Laplace transform is a good way for working with problems with discontinuities. Note, however, that solving differential equations using Laplace transforms has a severe limitation — it is best-suited for linear equations. The material in this section is based on Churchill's *Operational Mathematics* [14], which unfortunately is no longer in print.

Definition and basic examples

The Laplace transform takes as input a function $f : [0, \infty) \rightarrow \mathbb{R}$, which is considered as a function of the time variable t , and it returns as output a function $L[f]$ of a new variable, which we denote by s . More precisely, we have the following definition.

Definition 2.84 (Laplace transform). Let $f : [0, \infty) \rightarrow \mathbb{R}$ be a given function. Then the Laplace transform $L[f]$ of f is defined as

$$L[f](s) = \int_0^\infty f(t)e^{-st} dt,$$

for all $s \in \mathbb{R}$ for which the improper integral exists. We would like to point out that for most functions, there are many values of s for which the improper integral does not exist. This will be discussed in more detail later on.

Rather than providing theoretical results on which functions f lead to a convergent integral $L[f](s)$, we begin our discussion by computing the Laplace transform for a number of simple functions and by providing some basic properties.

Example 2.85 (Laplace transform of a constant and of the identity map). We start with the simple case of the constant function $f(t) = 1$ for all $t \in \mathbb{R}$. Then for all real numbers $s > 0$, the Laplace transform is defined by the improper integral

$$L[1](s) = \int_0^\infty e^{-st} dt = \lim_{T \rightarrow \infty} \left(-\frac{e^{-st}}{s} \Big|_{t=0}^{t=T} \right) = \frac{1}{s} \left(1 - \lim_{T \rightarrow \infty} e^{-sT} \right) = \frac{1}{s}.$$

We leave it to the reader to verify that for all $s \leq 0$, the improper integral $L[1](s)$ is in fact divergent, i.e., the Laplace transform is undefined.

We now turn our attention to the identity function $f(t) = t$. For arguments $s > 0$, standard integration by parts gives the identity

$$L[t](s) = \int_0^\infty te^{-st} dt = \lim_{T \rightarrow \infty} \left(-e^{-st} \left(\frac{1}{s^2} + \frac{t}{s} \right) \Big|_{t=0}^{t=T} \right) = \frac{1}{s^2}.$$

Also in this case, the Laplace transform is divergent for all arguments $s \leq 0$. ■

The above example shows that even for simple functions f , the Laplace transform does not necessarily exist for all arguments $s \in \mathbb{R}$. Nevertheless, in both cases it was defined for all $s > 0$. We will see in our further examples below, that for most functions of interest one can expect existence for all $s > \alpha$, for some constant α .

In our second example, we demonstrate that the Laplace transform shares one of the basic features of both differentiation and integration — it is a linear operation.

Example 2.86 (The Laplace transform is linear). Assume that the Laplace transforms of two given functions f and g exist for all arguments $s > \alpha$. Furthermore, let a and b denote two arbitrary real numbers. Then for all $s > \alpha$ we have

$$\begin{aligned} L[af + bg](s) &= \int_0^\infty (af(t) + bg(t))e^{-st} dt \\ &= a \int_0^\infty f(t)e^{-st} dt + b \int_0^\infty g(t)e^{-st} dt = aL[f](s) + bL[g](s). \end{aligned}$$

In other words, the Laplace transform of the linear combination $af + bg$ is defined for all arguments $s > \alpha$, and we have $L[af + bg] = aL[f] + bL[g]$, i.e., the Laplace transform is a linear operator. Together with the previous example, this implies

$$L[a + bt](s) = \frac{a}{s} + \frac{b}{s^2} \quad \text{for all } s > 0,$$

where a and b are arbitrary real constants. ■

Our next example is concerned with the Laplace transform of functions which involve exponentials, and after that we demonstrate how elementary deformations of f affect its transform.

Example 2.87 (Laplace transforms involving exponentials). In this example, we consider the exponential function $f(t) = e^{at}$ for some constant $a \in \mathbb{R}$. One can easily see that if $s > a$, then we have

$$L[e^{at}](s) = \int_0^\infty e^{at} e^{-st} dt = \int_0^\infty e^{-(s-a)t} dt = \int_0^\infty e^{-ut} dt,$$

where $u = s - a > 0$. This last integral is equal to $L[1](s - a)$, and this in turn implies

$$L[e^{at}](s) = \frac{1}{s - a} \quad \text{for all } s > a.$$

Moreover, one can readily see that for all $s \leq a$ the Laplace transform integral is divergent.

From the calculation above, we can recover even more. In particular, the Laplace transform of a given function $f(t)$ is related in an easy way to the Laplace transform of this same function f multiplied by the exponential e^{at} . That is, assume that $F(s) = L[f](s)$ denotes the Laplace transform for f , and suppose further that it is defined for all $s > \alpha$. Then one obtains

$$L[e^{at} f(t)](s) = \int_0^\infty f(t) e^{-(s-a)t} dt = F(s - a), \quad \text{for all } s > a + \alpha.$$

In other words, multiplication by an exponential function results in shifting the Laplace transform. ■

Example 2.88 (Laplace transform of shifts, dilations, and contractions). Inspired by the previous example, we examine how shifting a given function changes its Laplace transform. Specifically, assume that the Laplace transform of $f(t)$ is defined for $s > \alpha$, and consider an arbitrary positive constant $a > 0$. If we then define

$$g(t) = \begin{cases} 0 & \text{for all } 0 \leq t < a, \\ f(t - a) & \text{for all } t \geq a, \end{cases}$$

then the definition of the Laplace transform for g implies for all $s > \alpha$ the identity

$$L[g](s) = \int_0^a 0 dt + \int_a^\infty f(t - a) e^{-st} dt = \int_0^\infty f(u) e^{-s(u+a)} du = e^{-sa} L[f](s),$$

since $g(t) = 0$ for all $0 \leq t < a$. In other words, while multiplication of $f(t)$ by an exponential function results in shifting the Laplace transform, shifting the function f to the right leads to the multiplication of the transform by an exponential. Note that the function f in this example is discontinuous as long as $f(0) \neq 0$. This illustrates the ease of use of the Laplace transform for a function with a discontinuity.

In addition to shifts of functions, the elementary deformations of dilation and contraction lead to straightforward changes in the Laplace transform. To see this, assume

that $L[f]$ exists for all arguments $s > \alpha$. Then for $a > 0$ the Laplace transform of $f(at)$ can be computed using the change of variables $u = at$, and this immediately implies

$$L[f(at)](s) = \int_0^\infty f(at)e^{-st} dt = \int_0^\infty \frac{f(u)e^{-su/a}}{a} du = \frac{L[f(t)](s/a)}{a}.$$

Since we assumed that $L[f](s)$ exists for all $s > \alpha$, this formula holds for all $s > a\alpha$. ■

The above examples provide some initial intuition for a number of specific Laplace transforms. In general, there are many functions for which the Laplace transform can be explicitly calculated. In fact, numerous tables of Laplace transforms exist in the literature, and in many cases they can be computed using symbolic algebra packages.

On the importance of the Laplace transform for differential equations

We now address the question of existence of Laplace transforms. That is, under which reasonably general assumption does the improper integral in the definition of $L[f](s)$ exist? As the following result shows, a simple mild growth condition is all that it takes.

Lemma 2.89 (Transformable functions). *Suppose that $f : [0, \infty) \rightarrow \mathbb{R}$ is a piecewise continuous function such that on every bounded interval, the function f has only a finite number of points of discontinuity. If in addition there exist constants $M > 0$ and $\alpha \in \mathbb{R}$ such that*

$$|f(t)| \leq M e^{\alpha t} \quad \text{for all } t \geq 0,$$

then the Laplace transform $L[f]$ exists for all $s > \alpha$.

The simple proof of this statement follows from the comparison theorem for integrals, since we have already seen that the Laplace transform of the exponential $M e^{\alpha t}$ exists.

Our discussion so far has shown that the Laplace transform can often be easily determined, exists for all $s > \alpha$ under fairly mild assumptions, and that the effect of elementary function transformations has simple correspondences on the transform level. What we have not addressed so far is why it is so fundamentally important for the solution of differential equations.

The answer to this question comes in two parts. The first is theoretical in nature, and it shows that the function f is uniquely determined by its Laplace transform $L[f]$. In other words, making statements about f is equivalent to making transformed statements about $L[f]$. More precisely, we have the following result, which we state without proof.

Theorem 2.90 (Uniqueness of the inverse Laplace transform). *If f and g both satisfy the criteria for transformable functions from Lemma 2.89, and if for all arguments $s > \alpha$ we have $L[f](s) = L[g](s)$, then one necessarily has $f(t) = g(t)$ for all $t \geq 0$, except possibly at points of discontinuity. In other words, the mapping $f \mapsto L[f]$ is one-to-one.*

In addition to being one-to-one, the reader might wonder whether the transport process is also onto. That is, given an arbitrary function $F(s)$, is it possible to produce a function $f(t)$ such that $L[f] = F$? Unfortunately, the answer to this question is in general no. Both the general criteria required for $f(t)$ to exist and the methods for finding it in the general case involve complex analysis, and are therefore beyond the scope of this book. However, if we happen to recognize the Laplace transform, we can deduce what the underlying function $f(t)$ is.

The second ingredient for the importance of the Laplace transform in the context of the solution of differential equations stems from the fact that for a given differentiable function f , the Laplace transform of its derivatives can easily be found, as we show in the next example.

Example 2.91 (Laplace transforms of derivatives). Suppose that $f(t)$ is a continuously differentiable function which satisfies the conditions of Lemma 2.89. In this case, the Laplace transform $L[f]$ exists for all $s > \alpha$. Moreover, the definition of the Laplace transform applied to the first derivative f' can be written as

$$\begin{aligned} L[f'](s) &= \int_0^\infty f'(t)e^{-st} dt = \lim_{T \rightarrow \infty} (f(t)e^{-st}|_{t=0}^{t=T}) + \int_0^\infty f(t)se^{-st} dt \\ &= \lim_{T \rightarrow \infty} (f(T)e^{-sT}) - f(0) + sL[f](s) = sL[f](s) - f(0), \end{aligned}$$

as long as we assume again $s > \alpha$. Notice that the second equality uses integration by parts, while the last equality uses the fact that f satisfies the assumptions of Lemma 2.89 and we consider only $s > \alpha$.

The above computation shows that the Laplace transform of the derivative f' can be determined from $L[f]$ by an algebraic operation which involves the value $f(0)$ and the independent variable s . One can easily see that if the function f is more than one time continuously differentiable, then this procedure can be iterated. In fact, we leave it to the reader to verify that if f is n -times continuously differentiable, then the Laplace transform of $f^{(n)}$ is given by

$$L[f^{(n)}](s) = s^n L[f](s) - s^{n-1} f(0) - s^{n-2} f'(0) - \dots - f^{(n-1)}(0),$$

for all $s > \alpha$. Notice that this expression involves the values of $f^{(k)}(0)$ for all orders $k = 0, \dots, n-1$. ■

Solving the heat equation via Laplace transforms

After these preparations, we are in a position to explain how Laplace transforms can be used to solve boundary value problems for differential equations. Namely, by applying the Laplace transform to a *differential equation* which determines a solution u , one obtains an *algebraic equation* for the Laplace transform $L[u]$. The latter equation can then be solved to provide an explicit form of this Laplace transform, and if we can recognize the underlying function u , we have obtained the solution.

Before demonstrating this procedure in the context of a specific example, we still need one final definition. Namely, we need to introduce the error function erf , which arises in the context of Gaussian random variables in probability theory.

Definition 2.92 (The Gauss error function erf). The Gauss error function erf is defined by the integral

$$\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2} d\tau = \frac{1}{\sqrt{\pi}} \int_{-t}^t e^{-\tau^2} d\tau,$$

where the second identity is a consequence of the even symmetry of the integrand $e^{-\tau^2}$. Note that in our discussion regarding Definition 1.78, we showed that the last expression represents the probability that a normally (or Gaussian) distributed random variable with mean zero

and standard deviation $1/\sqrt{2}$ lies within the interval $(-t, t)$. Due to the above definition, the error function satisfies $\text{erf}(0) = 0$. Moreover, since the probability over the entire space is always one, we have $\lim_{t \rightarrow \infty} \text{erf}(t) = 1$.

The error function is important in a number of applications, despite the fact that it has no closed-form representation. That is, the integral in its definition cannot be expressed using standard functions. Even without an explicit formula for the integral, for a suitable transformed version of the error function, it is possible to determine its Laplace transform. This is the subject of the following example.

Example 2.93 (Laplace transform of a transformed error function). As a continuous and bounded function, the error function has a Laplace transform, defined for all $s > 0$. For our application below, we will need to be able to recognize the Laplace transform of the composition of the error function with the map $1/(2\sqrt{t})$. While the derivation of the Laplace transform of this composition is not too difficult, as it only involves integration by parts with a suitable substitution, we skip the details and merely present the result.

$$L[\text{erf}(1/(2\sqrt{t}))](s) = \frac{1 - e^{-\sqrt{s}}}{s} \quad \text{for all } s > 0,$$

where the function $\text{erf}(1/(2\sqrt{t}))$ is defined as 1 at $t = 0$. Also note that if $a > 0$ is an arbitrary constant, then one obtains

$$L[\text{erf}(1/(2\sqrt{at}))](s) = \frac{1 - e^{-\sqrt{s/a}}}{s} \quad \text{for all } s > 0. \quad (2.151)$$

This is derived by combining the above explicit Laplace transform formula with the dilation formula established earlier. ■

We are now finally in a position to put the Laplace transform to work in a specific example. In the following example, we use the Laplace transform to solve a heat equation on the infinite domain $\Omega = (0, \infty)$ by transforming the partial differential equation for the unknown function u into an ordinary differential equation for its Laplace transform.

Example 2.94 (Diffusion on the half line). This example is concerned with the evolution of dye in a long thin rod. More precisely, consider a long thin rod which is initially filled with dye-free fluid such that at one end of the rod, dye is being added with the concentration $f(t)$ at time t . We are interested in determining the concentration of dye in the rod as a function of both space and time.

To model this situation, we assume that the rod is infinitely long and covers the domain $\Omega = (0, \infty)$, and that dye is being added at the end of the rod which corresponds to $x = 0$. Then the diffusion of dye is governed by the heat or diffusion equation

$$u_t = k^2 u_{xx} \quad \text{on the domain } \mathbb{R}_0^+ \times \Omega, \quad (2.152)$$

subject to the dye-free initial condition

$$u(0, x) = 0 \quad \text{for all } x \in \Omega = (0, \infty), \quad (2.153)$$

and the Dirichlet boundary condition

$$u(t, 0) = f(t) \quad \text{for all } t > 0. \quad (2.154)$$

The last condition represents the addition of dye at the left end of the rod. Rather than considering the case of general f , we focus on the step function

$$f(t) = \begin{cases} A & \text{for all } t < T_0, \\ 0 & \text{for all } t \geq T_0. \end{cases}$$

In order to pin down a unique solution to the dye evolution problem, we also need to describe its behavior at the infinite end of Ω . For this, assume that

$$\lim_{x \rightarrow \infty} u(t, x) = 0 \quad \text{for all } t > 0.$$

Notice that while this assumption may seem unphysical at first glance — since no rod is infinitely long — it is equivalent to assuming that no dye can make it to the far end of the rod within the time frame of the experiment.

To solve the above partial differential equation problem, we begin by taking the Laplace transform of both sides of this equation with respect to t . To justify this step, we make the physically reasonable assumption that for each position $x \in \Omega$, the function value size $|u(t, x)|$ is bounded with respect to t , as this implies that the function $u(\cdot, x)$ has a Laplace transform which is defined for all $s > 0$. Similarly, we need to determine the Laplace transform for the boundary data $f(t)$. We denote these transforms by

$$U(s, x) = L[u(\cdot, x)](s) = \int_0^\infty u(t, x) e^{-st} dt \quad \text{for all } (s, x) \in \mathbb{R}^+ \times \Omega$$

and

$$F(s) = L[f](s) = \int_0^{T_0} e^{-st} A dt = \frac{A(1 - e^{-sT_0})}{s} \quad \text{for all } s > 0.$$

Notice that since the Laplace transform U acts only on the independent variable t , one can differentiate the integrand in the definition of U with respect to x in order to compute the corresponding derivatives of U with respect to x . That is,

$$L[u_{xx}(\cdot, x)](s) = U_{xx}(s, x) \quad \text{for all } (s, x) \in \mathbb{R}^+ \times \Omega.$$

Next we need to understand how the Laplace transform U relates to the Laplace transform of $u_t(\cdot, x)$. To see this, fix an arbitrary $x \in \Omega$. Then we can use the Laplace transform derivative formula from Example 2.91 to $u(\cdot, x)$ to obtain

$$L[u_t(\cdot, x)](s) = sU(s, x) - u(0, x) = sU(s, x) \quad \text{for all } s > 0,$$

where for the last identity we used the initial condition $u(0, x) = 0$. Finally, recall that at the $x = 0$ part of the boundary of Ω we assumed the Dirichlet boundary condition $u(t, 0) = f(t)$ for all $t > 0$, and this is equivalent to the identity

$$U(s, 0) = L[f](s) \quad \text{for all } s > 0.$$

In other words, every part of the initial/boundary value problem for the heat equation on $\Omega = (0, \infty)$ can be lifted to the level of Laplace transforms.

What does this reformulation of the original problem (2.152), (2.153), and (2.154), accomplish? By combining all the pieces of our above discussion one can immediately see that the Laplace transform $U(s, x)$ satisfies the ordinary differential equation

$$k^2 U_{xx}(s, x) = sU(s, x) \quad \text{for all } x \in \Omega = (0, \infty), \quad \text{for every } s > 0, \quad (2.155)$$

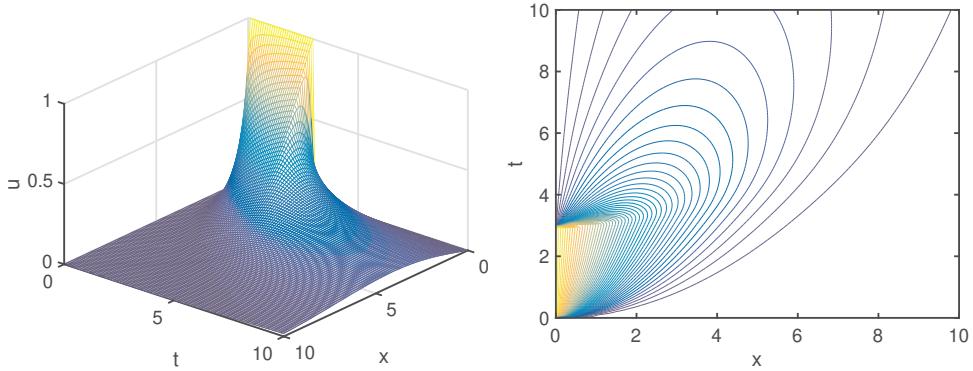


Figure 2.34. The diffusion of dye in an infinite rod. The images show the solution of the heat equation on the infinite domain $\Omega = (0, \infty)$ which is discussed in Example 2.94. For both images, the dye is added with concentration $A = 1$ for times between zero and $T_0 = 3$, and the diffusion constant is chosen as $k = 1$. The left panel contains a surface plot, while the right one shows the contour lines of the solution.

which is now a differential equation only in x , and which depends on the parameter s . This equation is supplemented by the conditions

$$U(s, 0) = F(s) \quad \text{and} \quad \lim_{x \rightarrow \infty} U(s, x) = 0 \quad \text{for all } s > 0, \quad (2.156)$$

i.e., an initial condition at $x = 0$ and a limiting boundary condition as $x \rightarrow \infty$.

After these preparations and transformations, solving the problem (2.155) subject to (2.156) is straightforward. First of all, notice that for every fixed $s > 0$ the differential equation (2.155) is a second-order linear ordinary differential equation with respect to x , which has the general solution

$$U(s, x) = \alpha(s)e^{\sqrt{s}x/k} + \beta(s)e^{-\sqrt{s}x/k} \quad \text{for all } x \geq 0,$$

where both the constants in the linear combination and the exponents of the fundamental solutions depend on the parameter $s > 0$. Due to the limiting condition in (2.156) we necessarily have $\alpha(s) = 0$, and the initial condition in (2.156) further implies $\beta(s) = F(s)$ for all parameter values $s > 0$. Together with the explicit form of F derived earlier, this yields for all $(s, x) \in \mathbb{R}^+ \times \Omega$ the explicit formula

$$U(s, x) = F(s)e^{-\sqrt{s}x/k} = A \frac{e^{-\sqrt{sx^2/k^2}}}{s} - Ae^{-sT_0} \frac{e^{-\sqrt{sx^2/k^2}}}{s}. \quad (2.157)$$

For every $x \in \Omega$, the Laplace transform $U(\cdot, x)$ determines a unique function $u(\cdot, x)$, and all we have to do is recognize which one. For this, recall that the identity $L[1] = 1/s$, in combination with (2.151) and the linearity of the transform, implies

$$L\left[1 - \operatorname{erf}\left(1/(2\sqrt{at})\right)\right] = \frac{e^{-\sqrt{s/a}}}{s}.$$

Moreover, for any transformable function $g : \mathbb{R} \rightarrow \mathbb{R}$ with $g(t) = 0$ for all $t < 0$ we have already seen that $L[g(t - b)](s) = e^{-sb} L[g(t)](s)$. In other words, if we introduce the abbreviation $G(s) = L[g(t)](s)$, then one obtains

$$L^{-1}\left[e^{-sb} G(s)\right](t) = \begin{cases} 0 & \text{for all } t < b, \\ g(t - b) & \text{for all } t \geq b. \end{cases}$$

where L^{-1} denotes the inverse Laplace transform. This formula is the last piece of our puzzle. If we let $a = k^2/x^2$ and $b = T_0$, then one obtains $\sqrt{1/a} = 1/\sqrt{(k^2/x^2)} = x/k$. We now consider two cases. To begin with, if $t < T_0$, then the second term in the last expression of (2.157) has the inverse Laplace transform of zero, and this yields

$$u(t, x) = A - A \operatorname{erf} \frac{x}{2k\sqrt{t}} \quad \text{for all } 0 \leq t < T_0 \quad \text{and } x \in \Omega.$$

On the other hand, for all $t \geq T_0$ the second term in the last expression of (2.157) becomes relevant. One can readily see that in this case we have

$$\begin{aligned} u(t, x) &= A \left(\left(1 - \operatorname{erf} \frac{x}{2k\sqrt{t}} \right) - \left(1 - \operatorname{erf} \frac{x}{2k\sqrt{t-T_0}} \right) \right) \\ &= A \operatorname{erf} \frac{x}{2k\sqrt{t-T_0}} - A \operatorname{erf} \frac{x}{2k\sqrt{t}} \quad \text{for all } t \geq T_0, \end{aligned}$$

where the first term is equal to A for $t = T_0$. After some straightforward simplifications, we finally see that the solution to the initial/boundary value problem for the diffusion equation (2.152), subject to (2.153) and (2.154), is given by the formula

$$u(t, x) = \begin{cases} A \int_{x/(2k\sqrt{t})}^{\infty} e^{-\tau^2} d\tau & \text{for all } 0 < t \leq T_0, \\ A \int_{x/(2k\sqrt{t})}^{x/(2k\sqrt{t-T_0})} e^{-\tau^2} d\tau & \text{for all } t > T_0. \end{cases}$$

Notice that there is a discontinuous change in the dye concentration at $x = 0$ and $t = T_0$, but the solution is smooth and continuous everywhere else. An illustration of the solution can be found in Figure 2.34. ■

2.6 • Separation of Variables in Different Coordinate Systems

So far, we have applied the technique of separation of variables only to rectangular domains Ω . Clearly this is to be expected, since we represent the desired solution as the product of functions of each of the coordinate variables, which in turn is then defined on products of intervals. However, in many applications one would like to study partial differential equations on domains with curved boundaries. Such domains can be treated using separation of variables, as long as they can be transformed into rectangular domains via a suitable coordinate transformation. In the following, we demonstrate how this can be achieved for the most common transformations.

2.6.1 • Polar Coordinates in the Plane

In planar systems with some type of rotational symmetry, a rectilinear separation of variables assumption does not apply. Rather, one has to express the underlying partial differential equation in polar coordinates and make the assumption that the polar and radial variables can be separated. In this way, separation of variables can still be used. In the current section, we illustrate this approach by first expressing the Laplacian operator in polar coordinates, and then using the transformed form to find radially symmetric solutions, as well as more general polar separated variables solutions to Laplace's equation.

The Laplacian in polar coordinates

Throughout this section, we will make use of polar coordinates to study partial differential equations on planar domains Ω which are *polar rectangles*. Such domains include disks and annuli, as well as circular sectors. Recall from calculus that the associated change of variables formulas from polar into rectilinear coordinates are given by the equations

$$x = r \cos \theta \quad \text{and} \quad y = r \sin \theta, \quad (2.158)$$

where r and θ are the new real independent variables. Using this new coordinate system, we can consider any function $u(x, y)$ equivalently as a function $u(r, \theta)$, where we usually abuse the notation and use the same letter u in both formulations.

Since the Laplacian operator forms the basis for most elliptic partial differential equations, it is natural to wonder how this differential operator changes under the above coordinate transformation. We will now show that the formula for the Laplacian in polar coordinates is the second-order singular operator given by

$$\Delta u = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} = u_{rr} + \frac{u_r}{r} + \frac{u_{\theta\theta}}{r^2}. \quad (2.159)$$

Notice that this operator is not defined at the origin, i.e., for the radius $r = 0$. We will need to take this into account during our applications below.

How can (2.159) be derived from the coordinate transformation (2.158)? For this, we begin by differentiating (2.158) directly, and then collect all partial derivatives in the following matrix

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{pmatrix}.$$

This matrix is nothing but the transpose of the Jacobian matrix for the change of variables function $(r, \theta) \mapsto (x, y)$. But what about the partial derivatives of r and θ with respect to x and y ? While for single-variable functions we have the convenient formula that the partial derivatives dy/dx and dx/dy are reciprocals, this is most definitely not the case once we consider two-variable functions. Therefore we need to work a bit in order to find partial derivatives for the change of variable formula from rectilinear coordinates into polar coordinates. To derive these expressions, we consider a generic function $u(x, y)$ and write it in polar coordinates as $u(r \cos \theta, r \sin \theta)$. Then the chain rule implies

$$\begin{pmatrix} u_r \\ u_\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix},$$

and after inverting the matrix one obtains

$$\begin{pmatrix} u_x \\ u_y \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \cos \theta & -\sin \theta \\ r \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} u_r \\ u_\theta \end{pmatrix}.$$

This last identity gives us the partial derivatives for the change of variables from rectilinear to polar coordinates. Namely,

$$\begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial \theta}{\partial x} \\ \frac{\partial r}{\partial y} & \frac{\partial \theta}{\partial y} \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \cos \theta & -\sin \theta \\ r \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -\frac{\sin \theta}{r} \\ \sin \theta & \frac{\cos \theta}{r} \end{pmatrix}.$$

This set of derivatives for the change of variables formulas — combined with multiple uses of the chain rule — finally allows us to express the Laplacian $\Delta u = u_{xx} + u_{yy}$ in terms of polar coordinates. More precisely, one obtains

$$\begin{aligned}
 \Delta u &= \left(\cos \theta u_r - \frac{\sin \theta}{r} u_\theta \right)_x + \left(\sin \theta u_r + \frac{\cos \theta}{r} u_\theta \right)_y \\
 &= \left(\cos \theta u_r - \frac{\sin \theta}{r} u_\theta \right)_r \cos \theta - \left(\cos \theta u_r - \frac{\sin \theta}{r} u_\theta \right)_\theta \frac{\sin \theta}{r} \\
 &\quad + \left(\sin \theta u_r + \frac{\cos \theta}{r} u_\theta \right)_r \sin \theta + \left(\sin \theta u_r + \frac{\cos \theta}{r} u_\theta \right)_\theta \frac{\cos \theta}{r} \\
 &= \cos^2 \theta u_{rr} - \cos \theta \sin \theta \left(\frac{u_\theta}{r} \right)_r - \frac{1}{r} (-\sin^2 \theta u_r + \cos \theta \sin \theta u_{r\theta}) \\
 &\quad + \frac{\sin \theta}{r^2} (\sin \theta u_\theta)_\theta + \sin^2 \theta u_{rr} + \sin \theta \cos \theta \left(\frac{u_\theta}{r} \right)_r \\
 &\quad + \frac{1}{r} (\cos^2 \theta u_r + \cos \theta \sin \theta u_{r\theta}) + \frac{\cos \theta}{r^2} (\cos \theta u_\theta)_\theta \\
 &= u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta},
 \end{aligned}$$

which is exactly the formula provided in (2.159). In other words, through repeated applications of the chain and product rules we have transformed the Laplacian into polar coordinates.

The Laplace equation on circular domains

What is the motivation for finding this expression? In the following sequence of examples, we demonstrate that a polar separation of variables can be used to find solutions for Laplace's equation with certain symmetries, or even solutions to nonhomogeneous boundary value problems. As in the case of rectilinear separation of variables, we will assume that the solution of interest can be written as a product of a function of r and a function of θ . This assumption then allows us to reduce the Laplace equation to two ordinary differential equations. In this new coordinate system, we are able to construct new solutions that could not be found using separation of variables in Cartesian coordinates.

As our first example, we consider the important special case of planar solutions for Laplace's equation which are radially symmetric. Namely, we find solutions which are invariant under all rotations of the plane around the origin. One can easily see that a function $u(r, \theta)$ is radially symmetric if and only if u does not depend on θ . This leads to the following example.

Example 2.95 (Radially symmetric solutions of Laplace's equation). In this example, our goal is to find all radially symmetric functions in the plane which satisfy Laplace's equation $\Delta u = 0$. As was mentioned above, this amounts to finding a solution u which in polar coordinates only depends on the radial variable r , i.e., we assume that the solution is of the form $u = u(r)$. If we substitute this form into the Laplacian operator in polar

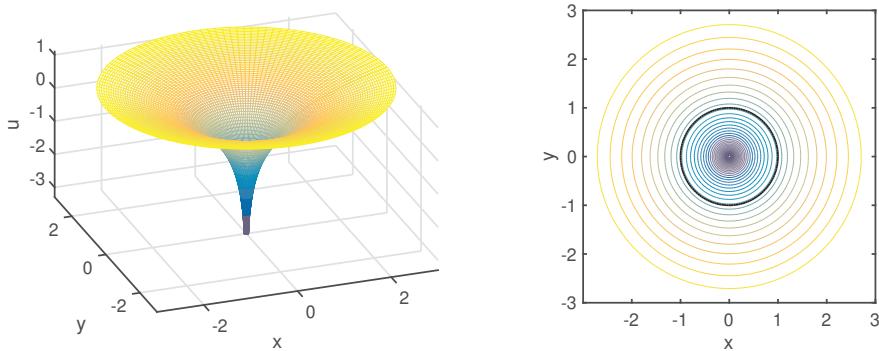


Figure 2.35. A radially symmetric solution to Laplace's equation. The solution $u(r, \theta) = \ln r$ is singular at the origin $r = 0$, since it limits to $-\infty$ as $r \rightarrow 0$. The left panel contains a surface plot of the solution, while the image on the right shows its level curves. The zero set of the solution is shown in black.

coordinates given in (2.159), then one obtains

$$0 = \Delta u = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right),$$

which implies $(ru_r)_r = 0$, and therefore $ru_r = A$ for some constant $A \in \mathbb{R}$. Solving this identity for u_r and integrating one more time with respect to r shows that

$$u(r) = A \ln r + B \quad \text{for some constants } A, B \in \mathbb{R}.$$

Notice that for $A \neq 0$ this solution limits to $-\infty$ and ∞ , respectively, as r approaches 0 and ∞ . One can interpret the radially symmetric solution to Laplace's equation as the electric potential for the repulsion between parallel charged wires of like charge. The specific form of the solution u then implies that the work required to bring these wires arbitrarily close to each other is infinite. For an illustration of the radially symmetric solution for $A = 1$ and $B = 0$ we refer the reader to Figure 2.35. ■

We now turn our attention to the general case. Rather than just studying radially symmetric solutions, we consider solutions of Laplace's equation which can be written as the product of a function of r and of a function of θ .

Example 2.96 (Polar separation of variables). As in the previous example, we consider the Laplace equation $\Delta u = 0$ on the plane, but this time our goal is to find as many solutions as possible which are of the form $u(r, \theta) = R(r)\Theta(\theta)$. Substituting this product into the polar coordinate form of Laplace's equation, we get

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r R' \Theta \right) + \frac{1}{r^2} R \Theta'' = 0,$$

and therefore

$$R'' \Theta + \frac{1}{r} R' \Theta + \frac{1}{r^2} R \Theta'' = 0.$$

Collecting the functions R and Θ on opposite sides of the equal sign leads to the identity

$$\frac{r^2 R'' + r R'}{R} = \frac{-\Theta''}{\Theta} = x \quad \text{for some constant } x \in \mathbb{R},$$

since each of the first two expressions depends on a different independent variable. As earlier, this in turn gives rise to two second-order linear ordinary differential equations for R and Θ . This time, however, we need to be more careful about the additional constraints. First of all, since we are looking for functions which are consistent with the polar coordinate representation, we have to impose both

$$\Theta(0) = \Theta(2\pi) \quad \text{and} \quad \Theta'(0) = \Theta'(2\pi),$$

and together with the equation $\Theta'' + x\Theta = 0$, we therefore obtain a Sturm-Liouville problem for Θ . Notice that this particular problem has already been solved in Example 2.43. Thus, we see that the only nontrivial solutions of the problem can be found for $x = m^2$, where $m \in \mathbb{N}_0$, and in this case all solutions are given by

$$\Theta(\theta) = A_m \cos(m\theta) + B_m \sin(m\theta) \quad \text{for arbitrary } A_m, B_m \in \mathbb{R}.$$

We now turn our attention to the ordinary differential equation for R , which for $x = m^2$ can be rewritten as

$$r^2 R'' + r R' - m^2 R = 0. \quad (2.160)$$

Some readers might immediately recognize this equation as the Cauchy-Euler equation, but rather than assume this prior knowledge, we show how to solve the equation directly. One can easily see that for $m = 0$ all solutions of this equation are given by

$$R(r) = C_0 + D_0 \ln r, \quad \text{where } C_0, D_0 \in \mathbb{R}.$$

In fact, this case is just the radially symmetric case discussed in the previous example.

In contrast, for the case $m \in \mathbb{N}$ it is not possible to integrate the equation directly. Nevertheless, since the equation for R is a second-order linear ordinary differential equation, we know a priori that the set of all possible solutions is a linear space spanned by two linearly independent solutions — and all we have to do is find such a fundamental set. For this, we make a formal power series assumption on R , i.e., we assume that R can be written in the form $R(r) = \sum_{j=-\infty}^{\infty} a_j r^j$. For now, we do not worry about the convergence of this series, but merely use it as a tool for finding two linearly independent solutions which will then be independently verified afterwards. By formal differentiation, the power series assumption gives rise to the expressions

$$R'(r) = \sum_{j=-\infty}^{\infty} j a_j r^{j-1} \quad \text{and} \quad R''(r) = \sum_{j=-\infty}^{\infty} j(j-1) a_j r^{j-2},$$

and substituting the three power series into the equation (2.160) leads to

$$\sum_{j=-\infty}^{\infty} (j(j-1) + j - m^2) a_j r^j = 0.$$

Due to the linear independence of the functions r^j for distinct $j \in \mathbb{Z}$, each term of this power series must be zero in order for the series to equal zero. Nontrivial coefficients a_j can therefore only occur when $j(j-1) + j - m^2 = 0$ is satisfied, which is equivalent to the identity $j^2 = m^2$. In other words, our formal calculation implies that for any $m \in \mathbb{N}$ there exists a solution of (2.160) of the form

$$R_m(r) = C_m r^m + D_m r^{-m} \quad \text{for arbitrary } C_m, D_m \in \mathbb{R},$$

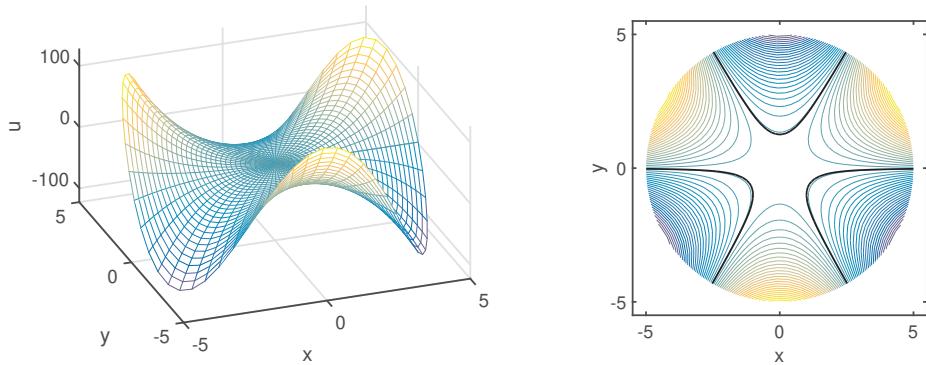


Figure 2.36. Laplace equation on a disk subject to Dirichlet boundary conditions. The images show the solution of $\Delta u = 0$ on the disk of radius 5, centered at the origin, which satisfies the boundary condition $u(5, \theta) = 2 + 125 \sin(3\theta)$ for $\theta \in [0, 2\pi]$, and which is given by $u(r, \theta) = 2 + r^3 \sin(3\theta)$. The left panel contains a surface plot of the solution, while the image on the right shows its level curves. The zero set of the solution is shown in black.

since one can easily verify that both r^m and r^{-m} are in fact linearly independent solutions of the linear second-order equation, and thus a fundamental set.

We can now combine our results for both R and Θ with the superposition principle. This shows that the general solution of $\Delta u = 0$ in polar coordinates is of the form

$$u(r, \theta) = C_0 + D_0 \ln r + \sum_{m=1}^{\infty} (C_m r^m + D_m r^{-m}) (A_m \cos(m\theta) + B_m \sin(m\theta)), \quad (2.161)$$

as long as the series converges. Once boundary conditions are given on a suitable planar domain, the coefficients A_m , B_m , C_m , and D_m can be determined. ■

We close this section with two concrete boundary value problems, which illustrate how the last formula can be used to find specific solutions. First, we consider the Laplace equation on a disk.

Example 2.97 (Laplace's equation on a disk). Consider the Laplace equation $\Delta u = 0$ on the disk of radius 5 given by $\Omega = \{(r, \theta) : r < 5\}$. In addition, we suppose that the solution satisfies the Dirichlet boundary condition

$$u(5, \theta) = 2 + 125 \sin(3\theta) \quad \text{for all } \theta \in [0, 2\pi].$$

This problem can be solved using the general solution given in (2.161). Since we are seeking a differentiable function u on Ω , it necessarily has to take finite function values there. This in turn implies that $D_m = 0$ for all integers $m \geq 0$. From the boundary conditions, we further see that

$$u(5, \theta) = 2 + 125 \sin(3\theta) = C_0 + \sum_{m=1}^{\infty} 5^m (C_m A_m \cos(m\theta) + C_m B_m \sin(m\theta)).$$

Thus by inspection one obtains $C_0 = 2$ and $C_3 B_3 = 1$, while all other coefficients appearing in this formula are zero. In other words, the above Dirichlet problem for the Laplace equation has the solution

$$u(r, \theta) = 2 + r^3 \sin(3\theta),$$

which is depicted in Figure 2.36. ■

Our second example considers a slightly more complicated domain. More precisely, we study the Laplace equation on an annulus whose boundary consists of two concentric circles. Again we only consider the case of Dirichlet boundary conditions.

Example 2.98 (Laplace's Equation on an annulus). Assume now that the function u satisfies Laplace's equation on the annulus $\Omega = \{(r, \theta) : 1 < r < 3\}$, subject to the Dirichlet boundary conditions

$$u(1, \theta) = \sin(5\theta) \quad \text{and} \quad u(3, \theta) = 6 \cos(5\theta) - 10 \sin(4\theta) \quad \text{for } \theta \in [0, 2\pi].$$

One can easily see that the solution expansion (2.161) for $u(r, \theta)$ can be rewritten as

$$u(r, \theta) = E_0 + F_0 \ln r + \sum_{m=1}^{\infty} \left(\left(E_m^c r^m + \frac{F_m^c}{r^m} \right) \cos(m\theta) + \left(E_m^s r^m + \frac{F_m^s}{r^m} \right) \sin(m\theta) \right). \quad (2.162)$$

We would like to point out that for fixed r , this expression for u is a generalized Fourier series with respect to θ in terms of the family of sines and cosines described in Examples 2.6 and 2.11. The boundary condition at the inner circle $r = 1$ then gives

$$\sin(5\theta) = E_0 + \sum_{m=1}^{\infty} ((E_m^c + F_m^c) \cos(m\theta) + (E_m^s + F_m^s) \sin(m\theta))$$

for all $\theta \in [0, 2\pi]$, and using the second boundary condition at the outer circle $r = 3$ one further obtains

$$\begin{aligned} 6 \cos(5\theta) - 10 \sin(4\theta) &= \\ E_0 + F_0 \ln 3 + \sum_{m=1}^{\infty} ((E_m^c 3^m + F_m^c 3^{-m}) \cos(m\theta) + (E_m^s 3^m + F_m^s 3^{-m}) \sin(m\theta)) &= \end{aligned}$$

By inspection, we can now match the cosine and sine terms on the left of each equation with the infinite series on the right sides. For each integer $m \in \mathbb{N}$, this provides us with four equations for the four unknowns E_m^c , F_m^c , E_m^s , and F_m^s , while for $m = 0$ we obtain exactly two equations for two unknowns. In the latter case, one can easily see that both

$$E_0 = 0 \quad \text{and} \quad E_0 + F_0 \ln 3 = 0$$

has to be satisfied, which leads to the unique solution $E_0 = F_0 = 0$. In the case $m = 4$, since the left-hand side of the second equation contains the term $-10 \sin(4\theta)$, one obtains

$$\begin{aligned} E_4^c + F_4^c &= 0, & E_4^s + F_4^s &= 0, \\ 3^4 E_4^c + 3^{-4} F_4^c &= 0, & 3^4 E_4^s + 3^{-4} F_4^s &= -10, \end{aligned}$$

which in turn yields the solution $E_4^c = F_4^c = 0$, $E_4^s \approx -0.1235$, and $F_4^s \approx 0.1235$. A fast way to solve this linear system is using Matlab. For example, to solve for the unknowns E_4^s and F_4^s above, one can use the commands

```
A = [1, 1; 3^4 3^(-4)];
b = [0; -10];
A\b
```

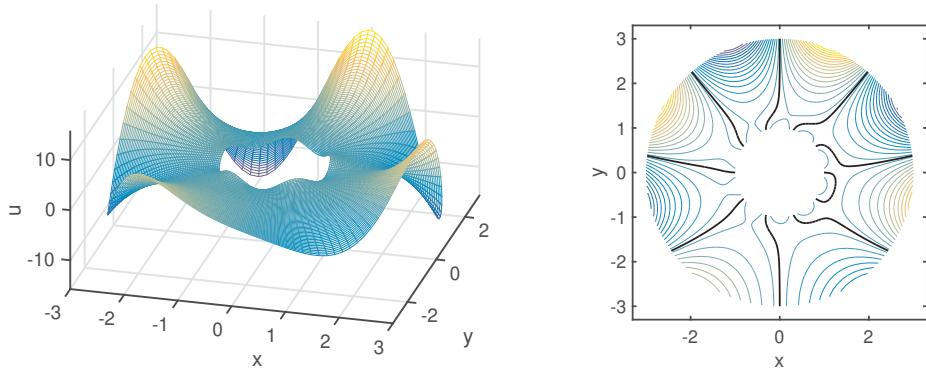


Figure 2.37. Laplace equation on an annulus subject to Dirichlet boundary conditions. The images show the solution of $\Delta u = 0$ on the annulus $1 < r < 3$, which satisfies the nonhomogeneous boundary conditions $u(1, \theta) = \sin(5\theta)$ and $u(3, \theta) = 6\cos(5\theta) - 10\sin(4\theta)$ for $\theta \in [0, 2\pi]$. The left panel contains a surface plot of the solution, while the image on the right shows its level curves. The zero set of the solution is shown in black.

Similarly, for $m = 5$ the left-hand side of the first equation contains the term $\sin(5\theta)$, while the left-hand side of the second equation contains the term $6\cos(5\theta)$, which results in the system

$$\begin{aligned} E_5^c + F_5^c &= 0, & E_5^s + F_5^s &= 1, \\ 3^5 E_5^c + 3^{-5} F_5^c &= 6, & 3^5 E_5^s + 3^{-5} F_5^s &= 0, \end{aligned}$$

and gives the unique solution $E_5^c \approx 0.0247$, $F_5^c \approx -0.0247$, $E_5^s \approx -1.69 \cdot 10^{-5}$, and $F_5^s \approx 1$. We leave it to the reader to verify that for all remaining natural numbers $m \in \mathbb{N}_0 \setminus \{0, 4, 5\}$, the analogous linear system has right-hand sides identical to zero, which furnishes the unique solutions $E_m^c = F_m^c = E_m^s = F_m^s = 0$.

The above discussion shows that by comparing the series expansion (2.162) with the boundary conditions, all coefficients in (2.162) can be uniquely determined. Together, this implies that the desired solution is given by the explicit formula

$$u(r, \theta) = \left(E_4^s r^4 + \frac{F_4^s}{r^4} \right) \sin(4\theta) + \left(E_5^c r^5 + \frac{F_5^c}{r^5} \right) \cos(5\theta) + \left(E_5^s r^5 + \frac{F_5^s}{r^5} \right) \sin(5\theta),$$

where the coefficients E_4^s , F_4^s , E_5^c , F_5^c , E_5^s , and F_5^s were defined above. This solution is illustrated in Figure 2.37. While the left panel in the figure shows a surface plot of u , the right image shows some of its level curves. ■

Note that a similar example is considered later in the book in Section 4.5.3, where we consider solutions to the wave equation in two spatial dimensions (and one time dimension) using polar coordinates.

2.6.2 • Cylindrical Coordinates and Bessel Functions

We now leave two space dimensions and consider cylindrical coordinates in \mathbb{R}^3 . This new coordinate system is very much related to the polar coordinate case, as it uses polar coordinates in the x - y -directions, and just adds on the z -coordinate. This leads to the transformation

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = z.$$

This specific form immediately implies that the Laplacian in cylindrical coordinates differs from the Laplacian in polar coordinates only by an additional second partial derivative with respect to z . In other words, in the three-dimensional cylindrical coordinate system we have

$$\Delta u = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2},$$

which is again a singular differential operator at $r = 0$.

Separation of variables under radial Dirichlet conditions

We now compute solutions to Laplace's equation using separation of variables in cylindrical coordinates. In other words, we seek a function u which satisfies $\Delta u = 0$ and is of the product form $u(r, \theta, z) = R(r)\Theta(\theta)Z(z)$. Rather than considering the most general case, we only try to find solutions u which satisfy a homogeneous Dirichlet boundary condition on the cylinder $r = 1$, i.e., we consider the boundary condition

$$u(1, \theta, z) = 0 \quad \text{for all } \theta \in [0, 2\pi) \quad \text{and } z \in \mathbb{R}.$$

In view of our separation of variables assumption, this boundary condition is equivalent to the requirement $R(1) = 0$. In addition, as in the case of polar coordinates we need to assume that the angular variable θ satisfies periodic boundary conditions, i.e., one has to impose both $\Theta(0) = \Theta(2\pi)$ and $\Theta'(0) = \Theta'(2\pi)$.

We can now turn our attention to the differential equation. After substituting the assumed product form of u into the cylindrical version of the Laplacian, one ends up with the equation

$$0 = \frac{1}{r} \frac{\partial}{\partial r} \left(r R' \Theta Z \right) + \frac{1}{r^2} R \Theta'' Z + R \Theta Z'',$$

and a slight reformulation of this identity implies

$$\frac{1}{rR} \frac{\partial}{\partial r} (rR') + \frac{\Theta''}{r^2 \Theta} = -\frac{Z''}{Z} = -\lambda. \quad (2.163)$$

Since the expression to the left of the first equality sign depends only on the independent variables r and θ , while the expression between the two equality signs depends merely on the independent variable z , both expressions necessarily have to be constant. As indicated, we denote this constant by $-\lambda \in \mathbb{R}$, and this leads to two new independent equations. The first of these can be rewritten as

$$\frac{1}{r} \frac{\partial}{\partial r} (rR') \Theta + \frac{R \Theta''}{r^2} = -\lambda R \Theta.$$

We would like to point out that this equation is actually an eigenvalue problem as discussed in Section 2.4.3. To see this, let $w(r, \theta) = R(r)\Theta(\theta)$. Then the left-hand side of the above equation is nothing but the two-dimensional polar coordinate version of the Laplacian of $w(r, \theta)$. In other words, the equation is equivalent to the eigenvalue problem $\Delta_{\text{polar}} w + \lambda w = 0$, where Δ_{polar} denotes the Laplacian in polar coordinates.

Rather than pursuing this line of inquiry further, we return to (2.163) and rewrite it in the fully separated form $r(R' + rR'')/R + \Theta''/\Theta = -\lambda r^2$, which implies

$$\frac{rR' + r^2R'' + \lambda r^2R}{R} = -\frac{\Theta''}{\Theta} = \chi.$$

Clearly, each side depends on a different independent variable, so each has to be equal to a constant $x \in \mathbb{R}$.

After these preparations, one can finally determine the solutions of the Laplace problem. First of all, Example 2.43 can be used to solve the equation $\Theta'' + x\Theta = 0$ subject to periodic boundary conditions. Thus, we know that $x = \ell^2$ for some $\ell \in \mathbb{N}_0$ and

$$\Theta(\theta) = A \cos(\ell\theta) + B \sin(\ell\theta), \quad \text{for } A, B \in \mathbb{R}.$$

Now fix an integer $\ell \in \mathbb{N}_0$, and consider the equation for $R(r)$ given by

$$r^2 R'' + r R' + (\lambda r^2 - \ell^2) R = 0,$$

which is supplemented by the Dirichlet condition $R(1) = 0$. Furthermore, we are looking for a solution that is defined at $r = 0$, so we have to assume that $R(0)$ is finite. As the astute reader might recall, the resulting boundary value problem is the singular Sturm-Liouville problem that has already been discussed in Example 2.46. In particular, we showed in the example that there are infinitely many linearly independent solutions, each of which is a scaled Bessel function $R(r) = J_\ell(v_{\ell,j} r)$, where $0 < v_{\ell,1} < v_{\ell,2} < \dots \rightarrow \infty$. The values $v_{\ell,j}$ are precisely the zeroes of the Bessel functions J_ℓ of the first kind, and the smallest possible value of $v_{\ell,j}$ is given by $v_{0,1} \approx 2.4048$. For more details, we refer the reader again to the detailed discussion in Example 2.46. The corresponding eigenvalues λ are $\lambda_{\ell,j} = v_{\ell,j}^2$. This completes the solution of the radial component $R(r)$.

As the last step, we only have to determine the solution of the differential equation for Z . For this, we leave $\ell \in \mathbb{N}_0$ fixed, and in addition fix an arbitrary Bessel function zero $v_{\ell,j} > 0$ from the above solution for $R(r)$. Then the equation for Z can be rewritten in the form $Z'' = \lambda_{\ell,j} Z = v_{\ell,j}^2 Z$, and its general solution is given by

$$Z(z) = C e^{v_{\ell,j} z} + D e^{-v_{\ell,j} z}.$$

Altogether, for fixed integers $j \in \mathbb{N}$ and $\ell \in \mathbb{N}_0$, we therefore obtain a solution $u_{\ell,j}(r, \theta, z)$ of the Laplace equation subject to homogeneous Dirichlet boundary conditions on the cylinder $r = 1$ given by

$$u_{\ell,j}(r, \theta, z) = (A_{\ell,j} \cos(\ell\theta) + B_{\ell,j} \sin(\ell\theta)) (C_{\ell,j} e^{v_{\ell,j} z} + D_{\ell,j} e^{-v_{\ell,j} z}) J_\ell(v_{\ell,j} r).$$

Using superposition, the general solution is the infinite linear combination

$$u(r, \theta, z) = \sum_{\ell=0}^{\infty} \sum_{j=1}^{\infty} (A_{\ell,j} \cos(\ell\theta) + B_{\ell,j} \sin(\ell\theta)) (C_{\ell,j} e^{v_{\ell,j} z} + D_{\ell,j} e^{-v_{\ell,j} z}) J_\ell(v_{\ell,j} r)$$

of these terms. The coefficients in this series representation have to be determined through the specification of additional boundary conditions.

2.6.3 ■ Spherical Coordinates and Spherical Harmonics

As our final example for separation of variables in different coordinate systems, we consider the case of spherical coordinates in three-dimensional Euclidean space \mathbb{R}^3 . These coordinates are given by the nonlinear transformation

$$x = r \cos\theta \sin\varphi, \quad y = r \sin\theta \sin\varphi, \quad z = r \cos\varphi.$$

In this formulation, the variable θ is the azimuthal angle in the xy -plane from the x -axis, with values between 0 and 2π , and φ denotes the polar angle from the positive z -axis, and taking values between 0 and π . By a similar calculation as the one performed for both polar and cylindrical coordinates, Laplace's equation in spherical coordinates can be rewritten in the form

$$\Delta u = \frac{1}{r^2} \left(\frac{\partial}{\partial r} \left(r^2 \frac{\partial u}{\partial r} \right) + \frac{1}{\sin^2 \varphi} \frac{\partial^2 u}{\partial \theta^2} + \frac{1}{\sin \varphi} \frac{\partial}{\partial \varphi} \left(\sin \varphi \frac{\partial u}{\partial \varphi} \right) \right).$$

Using this representation, the goal of the present section is to construct solutions of Laplace's equation $\Delta u = 0$ which can be written in the product form

$$u(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi),$$

i.e., we try to employ a spherical separation of variables assumption. In the following, we first consider the radially symmetric case in which both Θ and Φ are assumed to be constant. After that, we turn our attention to the general case.

Radially symmetric solutions to Laplace's equation and the wave equation

As our first step, we seek solutions of Laplace's equation with radial symmetry. In this case, the solution u is independent of the two angular variables θ and φ , i.e., it only depends on the distance r from the origin. Thus, we assume the form $u(r, \theta, \varphi) = u(r)$, and Laplace's equation becomes

$$0 = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{du}{dr} \right).$$

After multiplying by r^2 and integration once with respect to r one obtains

$$r^2 \frac{du}{dr} = c,$$

which in turn implies $du/dr = c/r^2$. Another integration then shows that all solutions are of the form

$$u(r) = \frac{A}{r} + B, \quad \text{for arbitrary } A, B \in \mathbb{R}.$$

As in the two-dimensional case, this solution can be viewed as a three-dimensional electric potential, in this case for two particles of like charge. As the distance r limits to zero, the term $1/r$ approaches infinity faster than $|\ln r|$, and this implies that the work required to move charged particles close together grows even faster than the work required to bring parallel charged wires together.

As a second example of radially symmetric equations, we seek solutions to the wave equation with radial symmetry. That is, we assume the form $u(t, r, \theta, \varphi) = u(t, r)$, and the wave equation becomes

$$u_{tt} = c^2 \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial u}{\partial r} \right).$$

For specificity, we assume the boundary conditions

$$u(t, 0) \text{ is finite, and } u(t, L) = 0.$$

A simple differentiation shows that

$$\frac{\partial^2}{\partial r^2}(ru) = \frac{1}{r} \frac{\partial}{\partial r} \left(r^2 \frac{\partial u}{\partial r} \right).$$

Therefore, multiplying both sides of the wave equation by r gives

$$(ru)_{tt} = c^2 \frac{\partial^2}{\partial r^2}(ru),$$

If we let $v = ru$, this is the one-dimensional wave equation for v . Putting this another way, $u(t, r) = v(t, r)/r$, where $v(t, x)$ solves the one-dimensional wave equation. We now establish the solution for v and subsequently u based on our boundary conditions. Recall that for $p > 0$ and arbitrary $A, B \in \mathbb{R}$, the spatial piece of the solution to the one-dimensional wave equation was either of the form $Ae^{px} + Be^{-px}$ or of the form $A\cos px + B\sin px$. Observe that by our finiteness assumption, we can eliminate three of the functions in this form. Namely, for any $p \neq 0$, the functions $e^{pr}/r, e^{-pr}/r$, and $\cos(pr)/r$ all grow without bound as $r > 0$ limits to zero. In contrast, from a standard calculus result $\lim_{r \rightarrow 0} \sin(pr)/r = p$. Combining this with the second boundary condition, we get

$$u(t, r) = \sum_{n=1}^{\infty} \frac{1}{r} \sin \frac{n\pi r}{L} \left(F_n \cos \frac{cn\pi t}{L} + G_n \sin \frac{cn\pi t}{L} \right),$$

where F_n and G_n are determined via prescribed initial conditions. We conclude that the oscillatory nature of solutions persists for radially symmetric solutions of the wave equation, but the amplitude of the oscillations decreases.

General spherical harmonics

We now look for functions of all three variables which are solutions to Laplace's equation under the spherical separation of variables assumption

$$u(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi).$$

Such solutions are called *spherical harmonics*, and some of them are illustrated below in Figures 2.38 and 2.39. Clearly, their illustration is a bit more problematic, since these are functions of three variables. For this reason, we only illustrate the behavior of some solutions $u(1, \theta, \varphi)$, which are restricted to the unit sphere, represented through a surface plot on this sphere. In the images, the function values are indicated by color variations, with positive values represented by yellow, and negative ones by blue. The specific notation used in the figure captions will be explained in more detail below.

We begin our derivation of spherical harmonics with the comment that as in the cases of polar and cylindrical coordinates, the azimuthal angle θ addresses the same points for its extreme values $\theta = 0$ and $\theta = 2\pi$. Thus, one has to assume that the function Θ satisfies the periodic boundary conditions $\Theta(0) = \Theta(2\pi)$ and $\Theta'(0) = \Theta'(2\pi)$. In contrast, the polar angle φ corresponds to the different poles, so we only impose a finiteness assumption on the function Φ and its derivative at both $\varphi = 0$ and $\varphi = \pi$. If we now substitute the product form of u into Laplace's equation, multiply the result by $r^2 \sin^2 \varphi$, and finally divide through by u , then one obtains the identity

$$\sin^2 \varphi \frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{\Theta} \frac{d^2 \Theta}{d\theta^2} + \frac{\sin \varphi}{\Phi} \frac{d}{d\varphi} \left(\sin \varphi \frac{d\Phi}{d\varphi} \right) = 0.$$

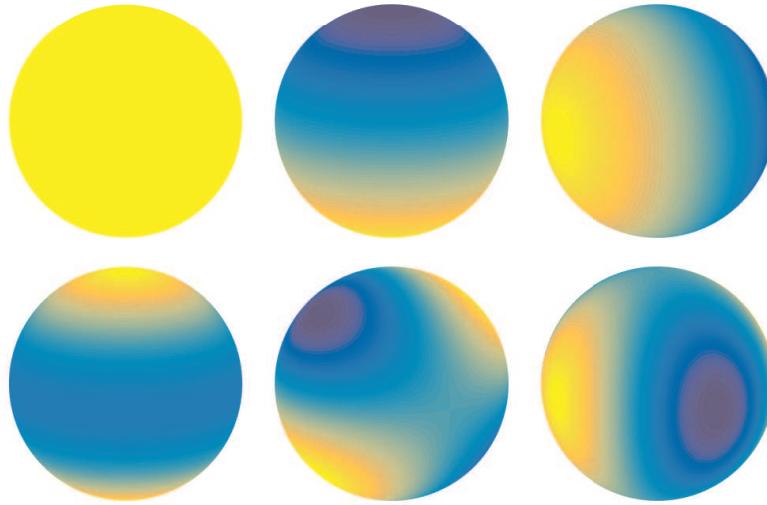


Figure 2.38. Spherical harmonics evaluated on the surface of the unit ball. The images show the functions $h_n^k(r, \theta, \varphi) = L_n^k(\cos \varphi) r^n \cos(k\theta)$ evaluated at $r = 1$. From top left to bottom right, the panels show the functions h_0^0 , h_1^0 , h_1^1 , h_2^0 , h_2^1 , and h_2^2 . In each case, the function values are indicated by color variations, with positive values represented by yellow, and negative ones by blue.

Since the second term of the left-hand expression depends only on the variable θ , while the other two terms are independent from θ , we conclude that

$$\frac{\Theta''}{\Theta} = -\lambda \quad \text{and} \quad \frac{1}{R} \frac{d}{dr} (r^2 R') + \frac{(\sin \varphi \Phi')'}{\Phi \sin \varphi} = \frac{\lambda}{\sin^2 \varphi},$$

for some constant $\lambda \in \mathbb{R}$. The first equality, combined with the necessary 2π -periodicity of the angular function Θ , implies that we have to have $\lambda = k^2$ for some integer $k \in \mathbb{N}_0$, as well as

$$\Theta(\theta) = A \cos(k\theta) + B \sin(k\theta) \quad \text{for } A, B \in \mathbb{R}.$$

See again Example 2.43 for details. Using the λ -values found above, the second equation can now be rewritten in the form

$$\frac{r^2 R'' + 2r R'}{R} = -\frac{(\sin \varphi \Phi')'}{\Phi \sin \varphi} + \frac{k^2}{\sin^2 \varphi}.$$

Since the left-hand side has no dependence on φ and the right-hand side is independent of the variable r , one therefore obtains both

$$\frac{r^2 R'' + 2r R'}{R} = \beta \quad \text{and} \quad -\frac{(\sin \varphi \Phi')'}{\Phi \sin \varphi} + \frac{k^2}{\sin^2 \varphi} = \beta.$$

The equation for $\Phi(\varphi)$ is equivalent to the singular Sturm-Liouville problem

$$\frac{d}{d\varphi} \left(\sin \varphi \frac{d\Phi}{d\varphi} \right) - \frac{k^2}{\sin \varphi} \Phi + \beta \sin \varphi \Phi = 0,$$

where we assume that the solution Φ and its derivative take finite values at $\varphi = 0$ and $\varphi = \pi$. While this problem could be solved directly, we approach it through a change of variables. If we consider the new variable $t = \cos \varphi$, i.e., we define the new function

$$Q(t) = Q(\cos \varphi) = \Phi(\varphi) \quad \text{for} \quad -1 \leq t \leq 1,$$

then the t -range clearly corresponds to the inequality constraint $0 \leq \varphi \leq \pi$. Moreover, the new function Q satisfies

$$\frac{d\Phi}{d\varphi} = \frac{dQ}{dt} \frac{dt}{d\varphi} = -\sin \varphi \frac{dQ}{dt},$$

which in turn leads to

$$-\sin \varphi \frac{d}{dt} \left(\sin \varphi \left(-\sin \varphi \frac{dQ}{dt} \right) \right) - \frac{k^2}{\sin \varphi} Q + \beta \sin \varphi Q = 0.$$

Due to the identity $\sin^2 \varphi = 1 - t^2$, and after division by $\sin \varphi$, this last equation can be rewritten in the form

$$\frac{d}{dt} \left((1-t^2) \frac{dQ}{dt} \right) - \frac{k^2}{1-t^2} Q + \beta Q = 0,$$

which turns out to be an equation which was discussed earlier in this book. More precisely, this equation is the associated Legendre Sturm-Liouville problem from Example 2.45, and it was shown there that nontrivial bounded solutions exist precisely when β is of the form $\beta = n(n+1)$, where $n \geq k$ denotes an arbitrary integer. In this case, the corresponding solutions for Φ are given by $\Phi(\varphi) = Q(\cos \varphi) = L_n^k(\cos \varphi)$, where $L_n^k(t)$ denotes the associated Legendre polynomials.

After the successful treatment of the two angular variables, it only remains to solve the equation for the radial function R . For this, fix arbitrary integers $n \geq k \geq 0$. Since in the last paragraph we saw that $\beta = n(n+1)$, one can now rewrite the differential equation for R in the form

$$r^2 R'' + 2rR' - n(n+1)R = 0.$$

The reader will immediately recognize that this equation is quite similar to the problem (2.160) which was studied during our discussion of polar coordinates. We therefore assume that the solution $R(r)$ can be represented as a, for now formal, series

$$R(r) = \sum_{j=-\infty}^{\infty} c_j r^j.$$

After substituting this representation into the differential equation and collecting like terms one readily obtains

$$\sum_{j=-\infty}^{\infty} (j(j-1) + 2j - n(n+1)) c_j r^j = 0.$$

As before, nonzero coefficients c_j can only occur if the expression preceding it vanishes. This, however, is equivalent to the identity

$$j(j-1) + 2j - n(n+1) = j(j+1) - n(n+1) = 0,$$

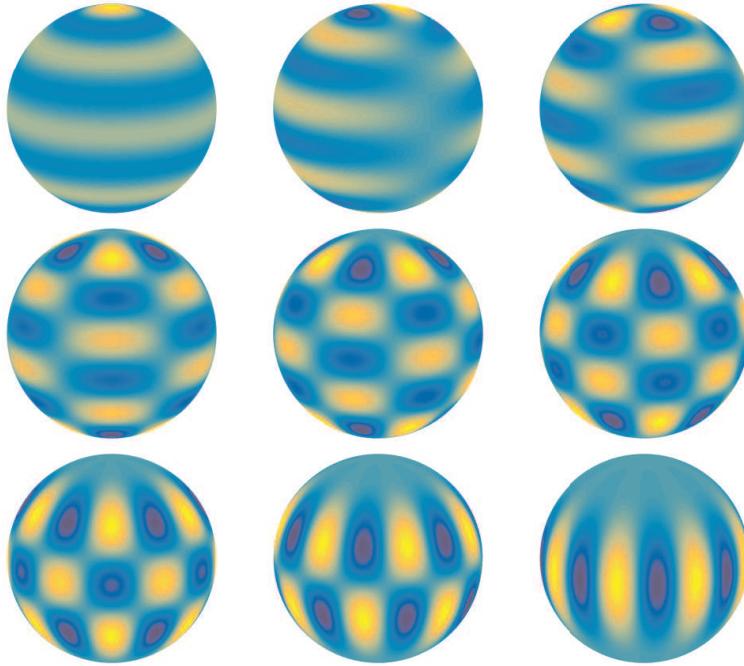


Figure 2.39. Spherical harmonics evaluated on the surface of the unit ball. The images show the functions $h_n^k(r, \theta, \varphi) = L_n^k(\cos \varphi) r^n \cos(k\theta)$ evaluated at $r = 1$. From top left to bottom right, the panels show the functions h_n^k for $k = 0, \dots, 8$. In each case, the function values are indicated by color variations, with positive values represented by yellow, and negative ones by blue.

which can only be true if either $j = n$ or $j = -(n + 1)$. This easily implies

$$R(r) = Cr^n + Dr^{-(n+1)} \quad \text{for } C, D \in \mathbb{R},$$

since both functions do in fact satisfy the differential equation for R , are linearly independent, and therefore form a fundamental set of solutions.

All that remains to be done at this time is to collect the above pieces. Specifically, by combining our results with the principle of superposition, one obtains that the solutions to Laplace's equation using separation of variables in spherical coordinates are of the form

$$u(r, \theta, \varphi) = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} L_n^k(\cos \varphi) \left(C_{k,n} r^n + \frac{D_{k,n}}{r^{n+1}} \right) (A_{k,n} \cos(k\theta) + B_{k,n} \sin(k\theta)), \quad (2.164)$$

for arbitrary real constants $A_{k,n}, B_{k,n}, C_{k,n}, D_{k,n} \in \mathbb{R}$, as long as the series can be guaranteed to converge and result in a sufficiently smooth function. As usual, the specific coefficients in the series have to be determined from suitable further boundary conditions.

As we mentioned earlier, some of the spherical harmonics, i.e., the terms after the double sum in (2.164), are illustrated in Figures 2.38 and 2.39. In these two figures, we only indicate the behavior of the respective series terms for $r = 1$ through function value

intensity plots on the unit sphere. In the images, positive function values are represented by yellow, and negative ones by blue.

Example 2.99 (Laplace's equation on a ball). We close this section with a quick application of (2.164). For this, we try to determine the solution to Laplace's equation on the open ball $\Omega = \{(r, \theta, \varphi) : r < 2\}$, and subject to the Dirichlet boundary condition

$$u(2, \theta, \varphi) = 6 \sin \varphi \cos \theta \quad \text{for all } \theta \in [0, 2\pi) \quad \text{and} \quad \varphi \in [0, \pi].$$

For this, we make use of the general solution given in (2.164). Since the domain Ω contains the origin and we are interested in a smooth solution, one has to have $D_{k,n} = 0$ for all integers $n \geq k \geq 0$. Recall also that it follows from the identity given in (2.45) that

$$L_1^1(s) = -\sqrt{1-s^2}.$$

This first implies $L_1^1(\cos \varphi) = -\sin \varphi$, and together with the imposed boundary condition we further obtain

$$\begin{aligned} u(2, \theta, \varphi) &= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} L_n^k(\cos \varphi) 2^n (C_{k,n} A_{k,n} \cos(k\theta) + C_{k,n} B_{k,n} \sin(k\theta)) \\ &= -6L_1^1(\cos \varphi) \cos \theta. \end{aligned}$$

By direct inspection, this immediately leads to the choice $C_{1,1} = 1$ and $A_{1,1} = -3$, while the remaining coefficients $A_{k,n}$, $B_{k,n}$, and $C_{k,n}$ are chosen as zero. Thus, the solution to the above Dirichlet boundary value problem is given by

$$u(r, \theta, \varphi) = -3r \sin \varphi \cos \theta.$$

We have to admit that this example is rather contrived. In the general case one has to find the generalized Fourier series for a trigonometric function in terms of associated Legendre polynomials. ■

2.7 ▪ Review Questions and Problems

2.7.1 ▪ Review Questions

Review Question 2.7.1. Find constants α and β such that the functions 1, x , and $1 + \alpha x + \beta x^2$ are orthogonal on the interval $[-1, 1]$.

Review Question 2.7.2. Show that the functions $\cos \frac{n\pi x}{L}$ for $n \in \mathbb{N}_0$ are orthogonal on $[0, L]$.

Review Question 2.7.3. Verify the following statements: (a) $e^{-x^2} \in L^2(1, \infty)$, (b) $1/\sqrt{x} \notin L^2(1, \infty)$, (c) $1/\sqrt[4]{x} \in L^2(0, 1)$, and (d) $1/\sqrt{x} \notin L^2(0, 1)$.

Review Question 2.7.4. Show that any function which is continuous on $[0, 1]$ is contained in $L^2([0, 1])$. Use this to show that $L^2([0, 1])$ contains nondifferentiable functions.

Review Question 2.7.5. Let $j, k \in \mathbb{N}_0$, and let $p_j(x) = x^j$, and $p_k(x) = x^k$. Find the inner product $\langle p_j, p_k \rangle_{L^2(-1, 1)}$. Conclude that the set of monomials is not an orthogonal set on this interval.

Review Question 2.7.6. Use Theorem 2.19 to show that the approximating series in Example 2.16 converges uniformly to the function x .

Review Question 2.7.7. For each of the following functions, find the coefficients γ_k of the generalized Fourier series $\sum_{k=1}^{\infty} \gamma_k \sin(k\pi x)$ for f on $(0, 1)$. (a) $f(x) = \sin(5\pi x) + 2\sin(9\pi x)$, (b) $f(x) = 10\sin(3\pi x)\cos(4\pi x) - 10\cos(3\pi x)\sin(4\pi x)$, (c) $f(x) = \sin(\pi x)\cos(\pi x)$. (Hint: In (b) and (c) use trigonometric identities to write these expressions as a sum of members of the sine family.)

Review Question 2.7.8. For each of the following functions, find coefficients γ_k of the generalized Fourier series $\sum_{k=0}^{\infty} \gamma_k \cos(k\pi x)$ for f on $(0, 1)$. (a) $f(x) = \cos(2\pi x)$, (b) $f(x) = \cos^2(13\pi x)$, (c) $f(x) = 1 - \sin^2(4\pi x)$, (d) $f(x) = 4\cos(2\pi x)\cos(5\pi x) + 4\sin(2\pi x)\sin(5\pi x)$. (Hint: In (b), (c), and (d) use trigonometric identities to write these expressions as a sum of members of the cosine family.)

Review Question 2.7.9. Show that the functions 1 , $1-x$, and $1-2x+x^2/2$ are orthogonal with respect to the weight function $w(x) = e^{-x}$ on the interval $[0, \infty)$.

Review Question 2.7.10. Let $T_n(x) = \cos(n \arccos x)$ for $n \in \mathbb{N}_0$. Show that the set of functions defined in this way is orthogonal with respect to the weight function $w(x) = 1/\sqrt{1-x^2}$ on the interval $[-1, 1]$.

Review Question 2.7.11. Show that a matrix $A \in \mathbb{R}^{d \times d}$ is symmetric if and only if the identity $(Aw) \cdot z = w \cdot (Az)$ holds for all $w, z \in \mathbb{R}^d$, where \cdot denotes the standard inner product in \mathbb{R}^d .

Review Question 2.7.12. For the domain $I = [0, L]$, work out the details to show that the functions $\varphi_k(x) = \cos(k\pi x/L)$ for $k \in \mathbb{N}_0$ are eigenfunctions of the Sturm-Liouville problem

$$u'' + \lambda u = 0 \quad \text{with} \quad u'(0) = 0 \quad \text{and} \quad u'(L) = 0.$$

What are the corresponding eigenvalues? Justify that this family of eigenfunctions is complete. Hint: Use Corollary 2.38 to show completeness.

Review Question 2.7.13. Show that the first four Hermite polynomials $1, x, x^2 - 1$, and $x^3 - 3x$ given in Example 2.24 satisfy the differential equation $u'' - xu' + \lambda u = 0$, each with a different value of λ . What is λ in each case?

Review Question 2.7.14. Using separation of variables, find a solution for the vibrating string given in (2.86-2.88), where the string is of length 5, $c = 2$, with fixed ends, with initial condition $u_t(0, x) = \sin(11\pi x/L)$, and such that the string initially has zero displacement: $u(0, x) = 0$. Hint: Use the solution formula in (2.94-2.96).

Review Question 2.7.15. Find a solution to the wave equation given in (2.86-2.88) with $c = 1$ on the unit interval with $u(t, 0) = u(t, 1) = 0$, $u(0, x) = \sin(3\pi x)$, and $u_t(0, x) = 4\sin(2\pi x)$. Hint: Use the solution formula in (2.94-2.96).

Review Question 2.7.16. Consider the heat equation $u_t = ku_{xx}$ on the domain $\Omega = [0, 1]$, subject to homogeneous Dirichlet boundary conditions. Furthermore, let $u(0, x) = f(x)$ for $x \in [0, 1]$. Use the solution formula given in (2.112, 2.113) to answer the following.

- (a) If $f(x) = 4\sin(2\pi x)$, find a solution for the equation.

- (b) If $f(x) = 4\sin(2\pi x) + 12\sin(3\pi x) + 3\sin(5\pi x)$, find a solution for the equation.
(c) If $f(x) = 4x(1-x)$, find a solution for the equation.

Review Question 2.7.17. Use the method of eigenfunctions to find a solution for the one-dimensional wave equation $u_{tt} = c^2 u_{xx}$ on the interval $\Omega = [0, L]$ where $c = 1$, with homogeneous Neumann boundary conditions $u_x(t, 0) = u_x(t, L) = 0$ for all $t \geq 0$. Apply your general form to solve the equation with initial conditions $u(0, x) = \sin^2(\pi x/L)$, $u_t(0, x) = 0$ for $x \in [0, L]$. Hint: Rewrite the initial condition in terms of an appropriate generalized Fourier series.

Review Question 2.7.18. A function $u(t, x)$ is said to be time periodic with least period p if $u(t + p, x) = u(t, x)$, and p is the smallest number to make this true. Find a solution $u(t, x)$ to the wave equation with $c = 1$ and homogeneous Dirichlet boundary conditions on $[0, 1]$ such that u has period $2/3$. Hint: Consider the solutions given in (2.92).

2.7.2 ■ Problems

Problem 2.7.1. Show that the only continuous function $f \in L^2(I)$ such that $\|f\|_{L^2(I)} = 0$ is the *zero function* on I , i.e. for every $x \in I$, $f(x) = 0$.

Problem 2.7.2. Consider the function

$$f(x) = x - 1. \quad (2.165)$$

Use the best approximation formula to find coefficients $\{a_k\}$ and $\{b_k\}$ for a Fourier series expansion for f on the interval $(0, 1)$. That is,

$$f(x) \stackrel{L^2}{=} a_0 + \sum_{k=1}^{\infty} (a_k \cos(2k\pi x) + b_k \sin(2k\pi x)).$$

Problem 2.7.3. For the function $f(x) = 3x^2 - 2x^3$ on the interval $[0, 1]$, do the following.

- (a) Find the coefficients γ_k for the generalized Fourier series $\sum_{k=0}^{\infty} \gamma_k \cos(k\pi x)$ for the function $f(x)$. Let $f_N = \sum_{k=0}^N \gamma_k \cos(k\pi x)$. This can either be done by hand or numerically. Graph your partial sum approximation f_N for $N = 10, 20, 50$. In addition, graph $|f_N - f|$ as a function of x . Does this appear to converge uniformly? If so, prove that this series converges uniformly. If not, indicate the mismatch that causes slow convergence.
- (b) Do the same exercise using the sine family.

Problem 2.7.4. Consider the boundary value problem on the interval $I = [0, 1]$ given by

$$u'' + \lambda u = 0 \quad \text{with} \quad u(0) = u'(0) \quad \text{and} \quad u(1) = u'(1)$$

where λ is a real parameter. For which values of λ can you find nontrivial solutions? Which solutions do you obtain?

Problem 2.7.5. Find an equation for the eigenvalues λ of the Sturm-Liouville problem

$$u'' + \lambda u = 0 \quad \text{on the interval} \quad I = [0, 1],$$

subject to the boundary conditions

$$u(0) = 0 \quad \text{and} \quad u(1) + u'(1) = 0,$$

and show that this equation has infinitely many solutions. Then express the eigenfunctions via the eigenvalues.

Problem 2.7.6. For the domain $I = [0, L]$, find all eigenvalues and eigenfunctions of the Sturm-Liouville problem

$$u'' + \lambda u = 0 \quad \text{with} \quad u(0) = 0 \quad \text{and} \quad u'(L) = 0,$$

and justify that the family of eigenfunctions that you constructed is complete. Hint: Use Corollary 2.38 to show completeness.

Problem 2.7.7 (Periodic boundary conditions). For the domain $I = [0, L]$, find all eigenvalues and eigenfunctions of the Sturm-Liouville problem

$$u'' + \lambda u = 0 \quad \text{with} \quad u(0) = u(L) \quad \text{and} \quad u'(0) = u'(L).$$

Can you use Corollary 2.38 to justify that the family of eigenfunctions that you constructed is complete? (Hint: Is this a regular Sturm-Liouville problem?)

Problem 2.7.8 (Legendre-Fourier series). In this problem, you find a generalized Fourier series expansion of a function in $L^2(-1, 1)$ using the Legendre polynomials. These functions are defined in Example 2.44 and are shown to be a complete orthogonal set on $L^2(-1, 1)$.

- (a) For the function $f(x) = (1 - x^2)e^x$, find an integral expression for the coefficients γ_k in the generalized Fourier series $\sum_{k=0}^{\infty} \gamma_k L_k(x)$, where $L_k(x)$ is the k^{th} Legendre polynomial. Evaluate this expression for $k = 0, 1$. Hint: Use formula (2.7).
- (b) Let f_N be the partial sum containing the first N terms of the Fourier series for $f(x)$. Graph your partial sum approximation f_N and the error $|f_N(x) - f(x)|$ for $N = 3, 7, 10$. Hint: Rather than evaluating by hand, find the coefficients numerically by adapting the code in `fouriercoeff.m` using the Matlab built in Legendre polynomials. (The command `legendre(N, x)` gives a length N vector, with first component $L_N(x)$. The full vector consists of the generalized Legendre polynomials $L_N^m(x)$, $m = 0, 1, \dots, N-1$.)

Problem 2.7.9 (Bessel-Fourier series). The goal of this problem is to write a generalized Fourier series expansion using the set of functions $\{J_1(\nu_{1,j}x)\}$, $j = 1, 2, \dots$ given in Example 2.46. The function J_1 is a Bessel function of the first kind, and $\nu_{1,j}$ are the zeroes of J_1 . These functions form a complete orthogonal set in $L_x^2(0, 1)$

- (a) Find an integral expression for the coefficients γ_k in the generalized Fourier series $\sum_{j=1}^{\infty} \gamma_j \varphi_j(x)$ where $\varphi_j(x) = J_1(\nu_{1,j}x)$ and $f(x) = (x - x^2)e^x$. Hint: Use formula (2.7) on $L_x^2(0, 1)$.
- (b) Evaluate γ_j for $j = 1, \dots, 5$ to give the values of $\nu_{1,j}$. This part of the problem is designed to be done numerically using your expressions from part (a). To do so, you can adapt the code `fouriercoeff.m` using the Matlab built in Bessel functions `besselj(N, x)`, along with the values given in Table 2.1. Graph your partial sum approximation f_N and the error $|f_N(x) - f(x)|$ for $N = 3, 5$. How well does your partial sum approximate $f(x)$?

Problem 2.7.10 (Hermite polynomials). The Hermite polynomials are an important set of complete orthogonal polynomials on $L_w^2(\mathbb{R})$ with the weight function $w(x) = e^{-x^2/2}$. They solve the Hermite Sturm-Liouville problem, given by

$$\frac{d}{dx} \left(e^{-x^2/2} \frac{du}{dx} \right) + \lambda e^{-x^2/2} u = 0 \quad \text{for } x \in I = \mathbb{R}.$$

Since there is no boundary for the whole real line, the boundary conditions are replaced by the assumption that the solutions grow no faster than x^n for some $n \in \mathbb{N}$ as $|x| \rightarrow \infty$.

- (a) Find $p(x)$, $q(x)$, and $r(x)$ for this Sturm-Liouville problem. Note that this problem is a singular Sturm-Liouville problem, since the interval $I = \mathbb{R}$ is unbounded.
- (b) Show that the above differential equation is equivalent to the differential equation

$$u'' - xu' + \lambda u = 0.$$

- (c) Assume that $u(x)$ is given by a power series $\sum_{k=0}^{\infty} a_k x^k$. Using the differential equation in part (b), derive a recursive relationship for the coefficients a_{k+2} in terms of a_k .
- (d) The assumption that solutions grow no faster than polynomially implies that this series must only contain a finite number of terms. Conclude that eigenvalues of this problem are $\lambda = n \in \mathbb{N}$, and that the series is either even or odd depending on whether λ is even or odd. That is, in order for the series to terminate, it is necessary to set either $a_0 = 0$ or $a_1 = 0$. In order to fully specify the solution, choose either $a_1 = 1$ or $a_0 = 1$ (depending on which one you have already set equal to zero). The eigenfunction corresponding to the eigenvalue n is a polynomial of degree n .
- (e) For the eigenvalues $\lambda = 0, 1, 2, 3$, use your recursive relation for a_k to find the first four Hermite polynomials. (Compare your answer to the functions listed in Example 2.24.)

Problem 2.7.11 (Chebyshev polynomials). The Chebyshev differential equation is given by

$$\frac{d}{dx} \left(\sqrt{1-x^2} \frac{du}{dx} \right) + \frac{\lambda}{\sqrt{1-x^2}} u = 0 \quad \text{for } x \in I = (-1, 1),$$

with boundary condition $u(1) = 1$, and the assumption that $u(-1)$ is bounded.

- (a) Find $p(x)$, $q(x)$, and $r(x)$ for this Sturm-Liouville problem. Is this a regular or a singular Sturm-Liouville problem?
 - (b) Show that the Sturm-Liouville differential equation is equivalent to the differential equation
- $$(1-x^2)u'' - xu' + \lambda u = 0.$$
- (c) Use the method of series solution to solve the differential equation in part (a). That is, assume $u(x) = \sum_{k=0}^{\infty} a_k (x-1)^k$, where by the boundary condition we know that $a_0 = 1$. Use the differential equation to show the recursive relation

$$a_{k+1} = \frac{\lambda - k^2}{1 + 3k + 2k^2} a_k.$$

- (d) The series is required to converge at $x = -1$. Just like in the case of the Legendre equation, this is the indeterminate case of the ratio test. However, if $\lambda = n^2$ for $n \in \mathbb{N}_0$, show that the series terminates after exactly n terms, implying that the solution is a degree n polynomial.
- (e) Let $T_n(x)$ correspond to the degree n polynomial eigenfunction with $\lambda = n^2$. Show that $T_0(x) = 1$, $T_1(x) = x$, $T_2(x) = 2x^2 - 1$, and $T_3(x) = 4x^3 - 3x$.
- (f) Use Sturm-Liouville theory to show that if $n, m \in \mathbb{N}_0$ are distinct, then $T_n(x)$ and $T_m(x)$ are orthogonal with respect to the weighting $r(x)$. Use the Stone-Weierstraß theorem to conclude that the set $\{T_n(x)\}_{n=0}^\infty$ is a complete orthogonal set on $L_r^2(-1, 1)$.

Problem 2.7.12 (Laguerre polynomials). Laguerre polynomials solve

$$\frac{d}{dx} \left(x e^{-x} \frac{du}{dx} \right) + \lambda e^{-x} u = 0 \quad \text{for } x \in I = (0, \infty),$$

with boundary condition $u(0) = 1$, and the assumption that u grows slower than x^k for some $k \in \mathbb{N}$.

- (a) For the above Sturm-Liouville problem, find $p(x)$, $q(x)$, and $r(x)$. This problem is a singular Sturm-Liouville problem since the interval $I = (0, \infty)$ is unbounded. Use the Sturm-Liouville results to conclude that any two eigenfunctions for this problem with distinct eigenvalues are orthogonal with respect to the weight $r(x)$.
- (b) Show that the following differential equation is equivalent to the Sturm-Liouville differential equation above:

$$x u'' + (1-x) u' + \lambda u = 0.$$

- (c) Assume a series solution $u(x) = \sum_{k=0}^\infty a_k x^k$. Use the differential equation to obtain the recursion relation

$$a_{k+1} = \frac{\lambda - k}{(k+1)^2} a_k.$$

From the boundary condition, conclude that $a_0 = 1$. From the growth condition, we know that this series must terminate after a finite number of terms. Therefore conclude that the eigenvalues are $\lambda = n \in \mathbb{N}_0$. Furthermore, for $\lambda = n$, the corresponding eigenfunction is a degree n polynomial.

- (d) Let $L_n(x)$ denote the eigenfunction corresponding to eigenvalue $\lambda = n$. Use part (a) combined with the Stone-Weierstraß theorem to conclude that $\{L_n(x)\}_{n=0}^\infty$ forms a complete orthogonal set in $L_r^2(0, \infty)$.
- (e) Use the recursion relation in part (b) to determine the first four Laguerre polynomials.

Problem 2.7.13. Consider the equation

$$u_t = \frac{1}{4} u_{xx}, \text{ on } (0, 1), \text{ such that } u_x(t, 0) = 0, u(t, 1) = 0. \quad (2.166)$$

- (a) Use separation of variables to write down the general solution to this equation.

(b) Write down a solution if $u(0, x) = 7 \cos(5\pi x/2)$.

Problem 2.7.14. Using separation of variables, find a solution for the wave equation on the interval $\Omega = [0, 1]$ where $u(t, 0) = u_x(t, 1) = 0$ for all $t \geq 0$, and $u(0, x) = x - x^2/2$ and $u_t(0, x) = 0$ for $x \in [0, 1]$. Show that your series solution converges uniformly for all $t > 0$.

Problem 2.7.15. Using separation of variables, find a set of solutions to the Laplace equation

$$u_{xx} + u_{yy} = 0 \quad \text{on } \Omega = [0, a] \times [0, a]$$

subject to the boundary conditions

$$u(x, 0) = u(x, a) = u(a, y) = 0 \quad \text{for } x \in [0, a] \text{ and } y \in [0, a],$$

and such that

$$u(0, y) = f(y) \quad \text{for } y \in [0, a].$$

In particular, solve the case for which $a = 1$ and

$$f(y) = 4y(1-y).$$

Problem 2.7.16. Find a series solution for the partial differential equation

$$u_{xx} + u_{yy} + u_y - u = 0 \quad \text{for } x, y \in (0, 1),$$

subject to the boundary conditions $u(0, y) = u(1, y) = u(x, 0) = 0$, as well as $u(x, 1) = x^2 - x^3$.

Problem 2.7.17. Find a series solution for the partial differential equation

$$u_{tt} + u_t = u_{xx} \quad \text{for } t \geq 0 \quad \text{and } x \in (0, 1),$$

subject to homogeneous Dirichlet boundary conditions $u(t, 0) = u(t, 1) = 0$ for $t \geq 0$, and for initial conditions $u(0, x) = f(x)$ and $u_t(0, x) = g(x)$ for $x \in (0, 1)$. Write an explicit solution when $f(x) = \sin(2\pi x)$, and $g(x) = \sin(\pi x) + 5 \sin(3\pi x)$.

Problem 2.7.18. Using separation of variables, find all eigenvalues and eigenfunctions of the Helmholtz equation

$$\Delta u + \lambda u = 0 \quad \text{on the domain } \Omega = (0, a) \times (0, b) \subset \mathbb{R}^2,$$

subject to the boundary conditions

$$u(0, y) = u(a, y) = u(x, 0) = 0 \quad \text{and } u_y(x, b) = 0$$

for all $x \in (0, a)$ and $y \in (0, b)$.

Problem 2.7.19. Assume that $u(x) = u(x_1, x_2, \dots, x_d)$ is an eigenfunction for the Helmholtz equation homogeneous Dirichlet boundary conditions on the bounded domain $\Omega \in \mathbb{R}^d$ with eigenvalue λ . That is,

$$\Delta u + \lambda u = 0 \text{ on } \Omega, \text{ and } u(x) = 0 \text{ for } x \in \partial\Omega.$$

Define $\tilde{\Omega} \in \mathbb{R}^d$ as follows:

$$\tilde{\Omega} = \left\{ x \in \mathbb{R}^d : \frac{x}{a} \in \Omega \right\}.$$

(That is, $\tilde{\Omega}$ is a uniform scaling of Ω by a). Show that

$$w(x_1, \dots, x_n) = u\left(\frac{x_1}{a}, \frac{x_2}{a}, \dots, \frac{x_n}{a}\right)$$

is an eigenfunction for the Helmholtz equation with homogeneous Dirichlet boundary conditions on the domain $\tilde{\Omega}$. What is the corresponding eigenvalue?

Problem 2.7.20. (Wave equation using eigenfunctions) Use the method of eigenfunctions as in Example 2.74 to solve the wave equation in two different cases.

(a) Solve the equation

$$u_{tt} = c^2 \Delta u \quad \text{on the domain } \Omega = (0, a) \times (0, b) \subset \mathbb{R}^2,$$

subject to the boundary conditions

$$u(t, 0, y) = u(t, a, y) = u(t, x, 0) = u(t, x, b) = 0$$

for all $x \in (0, a)$ and $y \in (0, b)$, where $u(0, x, y) = f(x, y)$, and $u_t(0, x, y) = g(x, y)$, where $f(x, y) = 5 \sin(2\pi x/a) \sin(3\pi y/b)$ and $g(x) = 3 \sin(4\pi x/a) \sin(7\pi y/b)$.

(b) More abstractly, let $\Omega \in \mathbb{R}^d$ be a bounded domain, and assume that we have constructed a complete orthogonal set of eigenfunctions $\varphi_n(x)$ with corresponding eigenvalues λ_n for the Helmholtz equation with homogeneous Neumann (or Dirichlet) boundary conditions. We assume $\lambda_n > 0$ for all n . Solve the wave equation

$$u_{tt} = c^2 \Delta u \quad \text{on the domain } \Omega \subset \mathbb{R}^d,$$

subject to the same homogeneous Neumann (or Dirichlet) boundary conditions as are satisfied by $\varphi_n(x)$, and such that $u(0, x) = f(x)$, and $u_t(0, x) = g(x)$, where $f(x) = 4\varphi_1(x) - 7\varphi_3(x)$, and let $g(x) = 3\varphi_2(x) + 8\varphi_3(x)$.

Problem 2.7.21. Use the method of eigenfunctions to solve the problem

$$u_t = \Delta u + 5u \quad \text{on the domain } \Omega = (0, 1) \times (0, 1) \subset \mathbb{R}^2,$$

subject to the boundary conditions

$$u(t, 0, y) = u(t, 1, y) = u(t, x, 0) = 0 \quad \text{and} \quad u_y(t, x, 1) = 0$$

on the unit square $(0, 1)^2$ with initial condition $u(0, x, y) = f(x, y)$. Apply your general series solution to solve the equation with the initial condition $f(x, y) = \sin(3\pi x) \sin(7\pi y/2)$. Hint: Use the results from Problem 2.7.18.

Problem 2.7.22. Solve the nonhomogeneous heat equation $u_t = c^2 u_{xx} + \sin(5\pi x)$ for all $0 < x < 1$, $t > 0$ subject to homogeneous Dirichlet boundary conditions $u(t, 0) = u(t, 1) = 0$ and initial condition $u(0, x) = 4 \sin(3\pi x) + 9 \sin(7\pi x)$.

Problem 2.7.23. Solve the nonhomogeneous wave equation $u_{tt} = 9u_{xx} + xt$ for all $0 < x < 1$, $t > 0$ subject to homogeneous Dirichlet boundary conditions $u(t, 0) = u(t, 1) = 0$ and initial conditions $u(0, x) = 0$ and $u_t(0, x) = 0$. Graph your answer for $t = 0, 0.25, 0.5, 0.75$, and 1 .

Problem 2.7.24. Solve the nonhomogeneous Laplace equation $\Delta w + f(x, y) = 0$ on the unit square $(0, 1)^2$, where $f(x, y) = 7 \sin(3\pi x) \sin(12\pi y) + 4 \sin(5\pi x) \sin(6\pi y)$, subject to homogeneous Dirichlet boundary conditions. That is, $u = 0$ on all four edges of the square.

Problem 2.7.25. Solve the parabolic problem

$$u_t = u_{xx} - u \quad \text{on } \Omega = (0, 1),$$

subject to the nonhomogeneous boundary conditions

$$u(t, 0) = 1 \quad \text{and} \quad u(t, 1) = 0 \quad \text{for } t \geq 0,$$

and the initial condition

$$u(0, x) = f(x) \quad \text{for } x \in (0, 1),$$

for some given function $f : (0, 1) \rightarrow \mathbb{R}$.

Problem 2.7.26. Solve the parabolic problem

$$u_t = u_{xx} + u \quad \text{on } \Omega = (0, 1),$$

subject to the nonhomogeneous boundary conditions

$$u(t, 0) = 1 \quad \text{and} \quad u(t, 1) = 1 \quad \text{for } t \geq 0,$$

and the initial condition

$$u(0, x) = f(x) \quad \text{for } x \in (0, 1),$$

for some given function $f : (0, 1) \rightarrow \mathbb{R}$.

Problem 2.7.27. Consider the case of a pollutant diffusing in a pipe $u_t = c^2 u_{xx}$ for $0 < x < 1$, $t > 0$, such that $u_x(t, 0) = -u_0$, $u(t, 1) = u_1$, and $u(0, x) = f(x)$. Express solutions in terms of an explicit series representation. What is the long-term behavior of solutions (independent of the initial condition)?

Problem 2.7.28. Find a solution to the wave equation $u_{tt} = c^2 u_{xx}$ for $0 < x < 1$, $t > 0$ with nonhomogeneous boundary conditions $u(t, 0) = u_0$, $u(t, 1) = u_1$ and initial conditions $u(0, x) = f(x)$, $u_t(0, x) = g(x)$. Write your solution in terms of an explicit series representation.

Problem 2.7.29. Show that $f(t) = 1/\sqrt{t}$ is an eigenfunction for the Laplace transform. That is, find a value of λ such that for all $s > 0$, $L[f(t)] + \lambda f(s) = 0$.

Problem 2.7.30. Use the Laplace transform to solve an ordinary differential equations on the half line via the following steps.

(a) Find the Laplace transforms of $\cos(kt)$ and $\sin(kt)$. The transforms are only defined for $s > 0$.

(b) Find the Laplace transform of the initial value problem

$$u'' + \lambda u = 0, \text{ for } x > 0, u(0) = a, u'(0) = b.$$

- (c) For $\lambda > 0$, use part (a) to solve the differential equation in part (b). Notice that unlike the case of a boundary value problem on an interval, every $\lambda > 0$ corresponds to a solution.
- (d) For $\lambda < 0$ for the differential equation in part (b), use partial fractions to write Laplace transform as a sum of two terms. Use this to solve the differential equation. Again, this differs from the problem on the interval, since values of $\lambda < 0$ give rise to nontrivial solutions.

Problem 2.7.31. Use Laplace transforms to find the solution for the heat equation

$$u_t = k^2 u_{xx} \text{ for } x > 0, t > 0,$$

$$u(0, x) = 0 \text{ for } x > 0,$$

$$\lim_{x \rightarrow \infty} u(t, x) = 0 \text{ for } t > 0, \text{ and}$$

$$u(t, 0) = \begin{cases} A & \text{for } T_1 < t < T_2 \\ 0 & \text{otherwise.} \end{cases}$$

Problem 2.7.32. Use Laplace transforms to find a solution for the following wave equation on the half line:

$$u_{tt} = c^2 u_{xx} \text{ for } x > 0, t > 0,$$

$$u(0, x) = u_t(0, x) = 0, \text{ for } x > 0,$$

$$\lim_{x \rightarrow \infty} u(t, x) = 0, \text{ and}$$

$$u(t, 0) = \begin{cases} A & \text{for } t < T_0 \\ 0 & \text{for } t \geq T_0. \end{cases}$$

Problem 2.7.33. Find a solution for Laplace's equation in the unit disk where $u = \cos^2 \theta$ (in terms of polar coordinates) on the unit circle.

Problem 2.7.34. Solve Laplace's equation in the planar region $x^2 + y^2 < 4$, with the boundary condition $u(x, y) = x^4$ for $x^2 + y^2 = 4$. (Hints: Do all your calculations in polar coordinates. Leave your answer in polar coordinates! Also note that in order to be a solution on a domain D , a function must be defined everywhere in D . Since the domain specified in this problem includes the origin, your answer must be a function that is defined at the point $r = 0$.)

Problem 2.7.35. Find the solution to the elliptic boundary value problem

$$\Delta u = 0 \quad \text{on} \quad \Omega = \left\{ x \in \mathbb{R}^2 : 1 < \sqrt{x_1^2 + x_2^2} < 3 \right\},$$

i.e., the domain Ω is an annulus, subject to the following boundary conditions for $\theta \in [0, 2\pi)$,

$$\begin{aligned} u(1 \cos \theta, 1 \sin \theta) &= 1 + 2 \cos(4\theta) - 5 \sin(\theta), \quad \text{and} \\ u(3 \cos \theta, 3 \sin \theta) &= 2 + 3 \cos(4\theta) + 2 \sin(4\theta). \end{aligned}$$

Problem 2.7.36. Use separation of variables in polar coordinates to find the solution to the wave equation on the disk of radius one with homogeneous Dirichlet boundary conditions on the boundary. Your general answer will be in the form of a double infinite series. For zero initial velocity and an initial displacement of the form $f(r, \theta)$, describe the procedure you would need to use to find the coefficients of your series. What is the minimum period of the solutions?

Problem 2.7.37 (Helmholtz equation on the unit disk). Find the eigenvalues and eigenfunctions for the Helmholtz equation on the unit disk subject to homogeneous Dirichlet boundary conditions. Hint: Adapt the methods used for solving Laplace's equation in cylindrical coordinates in Section 2.6.2.

Problem 2.7.38 (Laplacian for radially symmetric functions). Let $u : \mathbb{R}_0^+ \rightarrow \mathbb{R}$ be a twice differentiable function and define the function $v : \mathbb{R}^d \rightarrow \mathbb{R}$ by setting $v(x) = u(\|x\|)$ for all $x \in \mathbb{R}^d$, where we assume $d \geq 2$. If we use the abbreviation $r = \|x\|$, verify that

$$\Delta v(x) = \frac{\partial^2 u}{\partial r^2}(r) + \frac{d-1}{r} \frac{\partial u}{\partial r}(r) = \frac{1}{r^{d-1}} \frac{\partial}{\partial r} \left(r^{d-1} \frac{\partial u}{\partial r}(r) \right).$$

This formula generalizes the form of the Laplacian for radially symmetric functions which we have found using polar and spherical coordinates to the case of arbitrary dimensions.