Sandro Salsa

Partial Differential Equations in Action

From Modelling to Theory

Sandro Salsa

Partial Differential Equations in Action

From Modelling to Theory

Springer

Sandro Salsa
Dipartimento di Matematica
Politecnico di Milano

Contents

Preface

This book is designed as an advanced undergraduate or a first-year graduate course for students from various disciplines like applied mathematics, physics, engineering. It has evolved while teaching courses on partial differential equations (PDE) during the last few years at the Politecnico of Milan.

The main purpose of these courses was twofold: on the one hand, to train the students to appreciate the interplay between theory and modelling in problems arising in the applied sciences, and on the other hand to give them a solid theoretical background for numerical methods, such as finite elements.

Accordingly, this textbook is divided into two parts.

The **first one**, chapters 2 to 5, has a rather elementary character with the goal of developing and studying basic problems from the macro-areas of *diffusion, propagation and transport, waves and vibrations*. I have tried to emphasize, whenever possible, ideas and connections with concrete aspects, in order to provide intuition and feeling for the subject.

For this part, a knowledge of advanced calculus and ordinary differential equations is required. Also, the repeated use of the method of separation of variables assumes some basic results from the theory of Fourier series, which are summarized in appendix A.

Chapter 2 starts with the heat equation and some of its variants in which transport and reaction terms are incorporated. In addition to the classical topics, I emphasized the connections with simple stochastic processes, such as random walks and Brownian motion. This requires the knowledge of some elementary probability. It is my belief that it is worthwhile presenting this topic as early as possible, even at the price of giving up to a little bit of rigor in the presentation. An application to financial mathematics shows the interaction between probabilistic and deterministic modelling. The last two sections are devoted to two simple non linear models from flow in porous medium and population dynamics.

Chapter 3 mainly treats the Laplace/Poisson equation. The main properties of harmonic functions are presented once more emphasizing the probabilistic motivations. The second part of this chapter deals with representation formulas in

terms of potentials. In particular, the basic properties of the single and double layer potentials are presented.

Chapter 4 is devoted to first order equations and in particular to first order scalar conservation laws. The methods of characteristics and the notion of integral solution are developed through a simple model from traffic dynamics. In the last part, the method of characteristics is extended to quasilinear and fully nonlinear equations in two variables.

In chapter 5 the fundamental aspects of waves propagation are examined, leading to the classical formulas of d'Alembert, Kirchhoff and Poisson. In the final section, the classical model for surface waves in deep water illustrates the phenomenon of dispersion, with the help of the method of stationary phase.

The main topic of the **second part**, from chapter 6 to 9, is the development of Hilbert spaces methods for the *variational formulation* and the analysis of *linear boundary* and *initial-boundary value problems.* Given the abstract nature of these chapters, I have made an effort to provide intuition and motivation about the various concepts and results, running the risk of appearing a bit wordy sometimes.

The understanding of these topics requires some basic knowledge of Lebesgue measure and integration, summarized in appendix B.

Chapter 6 contains the tools from functional analysis in Hilbert spaces, necessary for a correct variational formulation of the most common boundary value problems. The main theme is the solvability of abstract variational problems, leading to the Lax-Milgram theorem and Fredholm's alternative. Emphasis is given to the issues of compactness and weak convergence.

Chapter 7 is divided into two parts. The first one is a brief introduction to the theory of distributions of L. Schwartz. In the second one, the most used Sobolev spaces and their basic properties are discussed.

Chapter 8 is devoted to the variational formulation of elliptic boundary value problems and their solvability. The development starts with one-dimensional problems, continues with Poisson's equation and ends with general second order equations in divergence form. The last section contains an application to a simple control problem, with both distributed observation and control.

The issue in chapter 9 is the variational formulation of evolution problems, in particular of initial-boundary value problems for second order parabolic operators in divergence form and for the wave equation. Also, an application to a simple control problem with final observation and distributed control is discussed.

At the end of each chapter, a number of exercises is included. Some of them can be solved by a routine application of the theory or of the methods developed in the text. Other problems are intended as a completion of some arguments or proofs in the text. Also, there are problems in which the student is required to be more autonomous. The most demanding problems are supplied with answers or hints.

The order of presentation of the material is clearly a consequence of my ... prejudices. However, the exposition if flexible enough to allow substantial changes

without compromising the comprehension and to facilitate a selection of topics for a one or two semester course.

In the first part, the chapters are in practice mutually independent, with the exception of subsections 3.3.6 and 3.3.7, which presume the knowledge of section 2.6.

In the second part, which, in principle, may be presented independently of the first one, more attention has to be paid to the order of the arguments. In particular, the material in chapter 6 and in sections 7.1–7.4 and 7.7–7.10 is necessary for understanding chapter 8, while chapter 9 uses concepts and results from section 7.11.

To Anna, my wife

1

Introduction

Mathematical Modelling – Partial Differential Equations – Well Posed Problems – Basic Notations and Facts – Smooth and Lipschitz Domains – Integration by Parts Formulas

1.1 Mathematical Modelling

Mathematical modelling plays a big role in the description of a large part of phenomena in the applied sciences and in several aspects of technical and industrial activity.

By a "mathematical model" we mean a set of equations and/or other mathematical relations capable of capturing the essential features of a complex natural or artificial system, in order to describe, forecast and control its evolution. The applied sciences are not confined to the classical ones; in addition to *physics* and *chemistry*, the practice of mathematical modelling heavily affects disciplines like *finance, biology, ecology, medicine, sociology.*

In the industrial activity (e.g. for aerospace or naval projects, nuclear reactors, combustion problems, production and distribution of electricity, traffic control, etc.) the mathematical modelling, involving first the analysis and the numerical simulation and followed by experimental tests, has become a common procedure, necessary for innovation, and also motivated by economic factors. It is clear that all of this is made possible by the enormous computational power now available.

In general, the construction of a mathematical model is based on two main ingredients: *general laws* and *constitutive relations.* In this book we shall deal with general laws coming from continuum mechanics and appearing as conservation or balance laws (e.g. of mass, energy, linear momentum, etc.).

The constitutive relations are of an experimental nature and strongly depend on the features of the phenomena under examination. Examples are the Fourier law of heat conduction, the Fick law for the diffusion of a substance or the way the speed of a driver depends on the density of cars ahead.

The outcome of the combination of the two ingredients is usually a *partial differential equation or a system of them.*

1.2 Partial Differential Equations

A partial differential equation is a relation of the following type:

$$F\left(x_1, ..., x_n, u, u_{x_1}, ..., u_{x_n}, u_{x_1 x_1}, u_{x_1 x_2}..., u_{x_n x_n}, u_{x_1 x_1 x_1}, ...\right) = 0 \qquad (1.1)$$

where the unknown $u = u\left(x_1, ...x_n\right)$ is a function of n variables and $u_{x_j}, ..., u_{x_i x_j}, ...$ are its partial derivatives. The highest order of differentiation occurring in the equation is the *order of the equation*.

A first important distinction is between *linear* and *nonlinear* equations.

Equation (1.1) is *linear* if F is linear with respect to u and all its derivatives, otherwise it is *nonlinear*.

A second distinction concerns the types of nonlinearity. We distinguish:

– *Semilinear* equations where F is nonlinear only with respect to u but is linear with respect to all its derivatives;
– *Quasi-linear* equations where F is linear with respect to the highest order derivatives of u;
– *Fully nonlinear equations where F is nonlinear with respect to the highest order derivatives of u.

The theory of linear equations can be considered sufficiently well developed and consolidated, at least for what concerns the most important questions. On the contrary, the non linearities present such a rich variety of aspects and complications that a general theory does not appear to be conceivable. The existing results and the new investigations focus on more or less specific cases, especially interesting in the applied sciences.

To give the reader an idea of the wide range of applications we present a series of examples, suggesting one of the possible interpretations. Most of them are considered at various level of deepness in this book. In the examples, **x** represents a space variable (usually in dimension $n = 1, 2, 3$) and t is a time variable.

We start with **linear equations**. In particular, equations (1.2)–(1.5) are fundamental and their theory constitutes a starting point for many other equations.

1. *Transport equation* (first order):

$$u_t + \mathbf{v} \cdot \nabla u = 0 \qquad (1.2)$$

It describes for instance the transport of a solid polluting substance along a channel; here u is the concentration of the substance and **v** is the stream speed. We consider the one-dimensional version of (1.2) in Section 4.2

2. *Diffusion* or *heat equation* (second order):

$$u_t - D\Delta u = 0, \qquad (1.3)$$

where $\Delta = \partial_{x_1 x_1} + \partial_{x_2 x_2} + ... + \partial_{x_n x_n}$ is the *Laplace operator*. It describes the conduction of heat through a homogeneous and isotropic medium; u is the temperature and D encodes the thermal properties of the material. Chapter 2 is devoted to the heat equation and its variants.

3. *Wave equation* (second order):

$$u_{tt} - c^2 \Delta u = 0. \tag{1.4}$$

It describes for instance the propagation of transversal waves of small amplitude in a perfectly elastic chord (e.g. of a violin) if $n = 1$, or membrane (e.g. of a drum) if $n = 2$. If $n = 3$ it governs the propagation of electromagnetic waves in vacuum or of small amplitude sound waves (Section 5.8). Here u may represent the wave amplitude and c is the propagation speed.

4. *Laplace's or potential equation* (second order):

$$\Delta u = 0, \tag{1.5}$$

where $u = u(\mathbf{x})$. The diffusion and the wave equations model evolution phenomena. The Laplace equation describes the corresponding *steady state*, in which the solution does not depend on time anymore. Together with its nonhomogeneous version

$$\Delta u = f,$$

called *Poisson's* equation, it plays an important role in electrostatics as well. Chapter 3 is devoted to these equations.

5. *Black-Scholes equation* (second order):

$$u_t + \frac{1}{2}\sigma^2 x^2 u_{xx} + rxu_x - ru = 0.$$

Here $u = u(x,t)$, $x \geq 0$, $t \geq 0$. Fundamental in mathematical finance, this equation governs the evolution of the price u of a so called *derivative* (e.g. an *European option*), based on an underlying asset (a stock, a currency, etc.) whose price is x. We meet the Black-Scholes equation in Section 2.9.

6. *Vibrating plate* (fourth order):

$$u_{tt} - \Delta^2 u = 0,$$

where $\mathbf{x} \in \mathbb{R}^2$ and

$$\Delta^2 u = \Delta(\Delta u) = \frac{\partial^4 u}{\partial x_1^4} + 2\frac{\partial^4 u}{\partial x_1^2 \partial x_2^2} + \frac{\partial^4 u}{\partial x_2^4}$$

is the *biharmonic operator*. In the theory of linear elasticity, it models the transversal waves of small amplitude of a homogeneous isotropic plate (see Section 8.7).

7. *Schrödinger equation* (second order):

$$-iu_t = \Delta u + V(\mathbf{x})u$$

where i is the complex unit. This equation is fundamental in quantum mechanics and governs the evolution of a particle subject to a potential V. The function $|u|^2$

represents a *probability density*. We will briefly encounter the Schrödinger equation in Problem 6.6.

Let us list now some examples of **nonlinear equations**

8. *Burger's equation* (quasilinear, first order):

$$u_t + cuu_x = 0 \qquad (x \in \mathbb{R}).$$

It governs a one dimensional flux of a non viscous fluid but it is used to model traffic dynamics as well. Its viscous variant

$$u_t + cuu_x = \varepsilon u_{xx} \qquad (\varepsilon > 0)$$

constitutes a basic example of competition between *dissipation* (due to the term εu_{xx}) and *steepening* (shock formation due to the term cuu_x). We will discuss these topics in Section 4.4.

9. *Fisher's equation* (semilinear, second order):

$$u_t - D\Delta u = ru(M - u)$$

It governs the evolution of a population of density u, subject to diffusion and logistic growth (represented by the right hand side). We examine the one-dimensional version of Fisher's equation in Section 2.10.

10. *Porous medium equation* (quasilinear, second order):

$$u_t = k \operatorname{div}(u^\gamma \nabla u)$$

where $k > 0$, $\gamma > 1$ are constant. This equation appears in the description of filtration phenomena, e.g. of the motion of water through the ground. We briefly meet the one-dimensional version of the porous medium equation in Section 2.10.

11. *Minimal surface equation* (quasilinear, second order):,

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = 0 \qquad (\mathbf{x} \in \mathbb{R}^2)$$

The graph of a solution u minimizes the area among all surfaces $z = v(x_1, x_2)$ whose boundary is a given curve. For instance, soap balls are minimal surfaces. We will not examine this equation (see e.g. *R. Mc Owen*, 1996).

12. *Eikonal equation* (fully nonlinear, first order):

$$|\nabla u| = c(\mathbf{x})$$

It appears in geometrical optics: if u is a solution, its level surfaces $u(\mathbf{x}) = t$ describe the position of a light wave front at time t. A bidimensional version is examined in Chapter 4.

Let us now give some examples of **systems**.

13. *Navier's equation of linear elasticity*: (three scalar equations of second order):

$$\varrho\mathbf{u}_{tt} = \mu\Delta\mathbf{u} + (\mu + \lambda)\text{grad div }\mathbf{u}$$

where $\mathbf{u} = (u_1(\mathbf{x},t), u_2(\mathbf{x},t), u_3(\mathbf{x},t))$, $\mathbf{x} \in\mathbb{R}^3$. The vector $\mathbf{u}$ represents the displacement from equilibrium of a deformable continuum body of (constant) density ϱ. We will not examine this system (see e.g. *R. Dautray and J. L. Lions*, Vol. 1,6, 1985).

14. *Maxwell's equations in vacuum* (six scalar linear equations of first order):

$$\mathbf{E}_t - \text{curl }\mathbf{B} = \mathbf{0}, \qquad \mathbf{B}_t + \text{curl }\mathbf{E} = \mathbf{0} \qquad \text{(Ampère and Faraday laws)}$$

$$\text{div }\mathbf{E} = 0 \qquad \text{div }\mathbf{B} = 0 \qquad \text{(Gauss' law)}$$

where $\mathbf{E}$ is the electric field and $\mathbf{B}$ is the magnetic induction field. The unit measures are the "natural" ones, i.e. the light speed is $c = 1$ and the magnetic permeability is $\mu_0 = 1$. We will not examine this system (see e.g. *R. Dautray and J. L. Lions*, Vol. 1, 1985).

15. *Navier-Stokes equations* (three quasilinear scalar equations of second order and one linear equation of first order):

$$\begin{cases} \mathbf{u}_t + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu\Delta\mathbf{u} \\ \text{div }\mathbf{u} = 0 \end{cases}$$

where $\mathbf{u} = (u_1(\mathbf{x},t), u_2(\mathbf{x},t), u_3(\mathbf{x},t))$, $p = p(\mathbf{x},t)$, $\mathbf{x} \in\mathbb{R}^3$. This equation governs the motion of a viscous, homogeneous and incompressible fluid. Here $\mathbf{u}$ is the fluid speed, p its pressure, ρ its density (constant) and ν is the kinematic viscosity, given by the ratio between the fluid viscosity and its density. The term $(\mathbf{u}\cdot\nabla)\mathbf{u}$ represents the inertial acceleration due to fluid transport. We will briefly meet the Navier-Stokes equations in Section 3.4.

1.3 Well Posed Problems

Usually, in the construction of a mathematical model, only some of the general laws of continuum mechanics are relevant, while the others are eliminated through the constitutive laws or suitably simplified according to the current situation. In general, additional information is necessary to select or to predict the existence of a unique solution. This information is commonly supplied in the form of *initial and/or boundary data,* although other forms are possible. For instance, typical boundary conditions prescribe the value of the solution or of its normal derivative, or a combination of the two. A main goal of a theory is to establish suitable conditions on the data in order to have a problem with the following features:

a) *there exists at least one solution;*

b) *there exists at most one solution;*

c) *the solution depends continuously on the data.*

This last condition requires some explanations. Roughly speaking, property c) states that the correspondence

$$data \rightarrow solution \tag{1.6}$$

is *continuous* or, in other words, that a *small error on the data entails a small error on the solution.*

This property is extremely important and may be expressed as a **local stability of the solution with respect to the data**. Think for instance of using a computer to find an approximate solution: the insertion of the data and the computation algorithms entail approximation errors of various type. A significant sensitivity of the solution on small variations of the data would produce an unacceptable result.

The notion of continuity and the error measurements, both in the data and in the solution, are made precise by introducing a suitable notion of *distance*. In dealing with a numerical or a finite dimensional set of data, an appropriate distance may be the usual *euclidean distance:* if $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n)$ then

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}.$$

When dealing for instance with real functions, defined on a set A, common distances are:

$$\text{dist}(f, g) = \max_{\mathbf{x} \in A} |f(\mathbf{x}) - g(\mathbf{x})|$$

which measures the maximum difference between f and g over A, or

$$\text{dist}(f, g) = \sqrt{\int_A (f - g)^2}$$

which is the so called *least square distance between f and g.*

Once the notion of distance has been chosen, the continuity of the correspondence (1.6) is easy to understand: *if the distance of the data tends to zero then the distance of the corresponding solutions tends to zero.*

When a problem possesses the properties a), b) c) above it is said to be **well posed**. When using a mathematical model, it is extremely useful, sometimes essential, to deal with well posed problems: existence of the solution indicates that the model is coherent, uniqueness and stability increase the possibility of providing accurate numerical approximations.

As one can imagine, complex models lead to complicated problems which require rather sophisticated techniques of theoretical analysis. Often, these problems

become well posed and efficiently treatable by numerical methods if suitably re-formulated in the abstract framework of Functional Analysis, as we will see in Chapter 6.

On the other hand, not only well posed problems are interesting for the applications. There are problems that are intrinsically *ill posed* because of the lack of uniqueness or of stability, but still of great interest for the modern technology. We only mention an important class of ill posed problems, given by the so called **inverse problems**, closely related to *control theory*, of which we provide simple examples in Sections 8.8 and 9.2.

1.4 Basic Notations and Facts

We specify some of the symbols we will constantly use throughout the book and recall some basic notions about sets, topology and functions.

Sets and Topology. We denote by: $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ the sets of natural numbers, integers, rational, real and complex numbers, respectively. $\mathbb{R}^n$ is the $n-$dimensional vector space of the $n-$uples of real numbers. We denote by $\mathbf{e}^1,...,\mathbf{e}^n$ the unit vectors in the canonical base in $\mathbb{R}^n$. In $\mathbb{R}^2$ and $\mathbb{R}^3$ we may denote them by $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$.

The symbol $B_r(\mathbf{x})$ denotes the *open* ball in $\mathbb{R}^n$, with radius r and center at $\mathbf{x}$, that is

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n;\ |\mathbf{x} - \mathbf{y}| < r\}.$$

If there is no need to specify the radius, we write simply $B(\mathbf{x})$. The volume of $B_r(\mathbf{x})$ and the area of $\partial B_r(\mathbf{x})$ are given by

$$|B_r| = \frac{\omega_n}{n}r^n \qquad \text{and} \qquad |\partial B_r| = \omega_n r^{n-1}$$

where ω_n is the surface area of the unit sphere[1] ∂B_1 in $\mathbb{R}^n$; in particular $\omega_2 = 2\pi$ and $\omega_3 = 4\pi$.

Let $A \subseteq \mathbb{R}^n$. A point $\mathbf{x} \in A$ is:

- an *interior point* if there exists a ball $B_r(\mathbf{x}) \subset A$;
- a *boundary point* if any ball $B_r(\mathbf{x})$ contains points of A **and** of its complement $\mathbb{R}^n \backslash A$. The set of boundary points of A, the *boundary of* A, is denoted by ∂A;
- a *limit point* of A if there exists a sequence $\{\mathbf{x}_k\}_{k \geq 1} \subset A$ such that $\mathbf{x}_k \to \mathbf{x}$.

A is *open* if every point in A is an interior point; the set $\overline{A} = A \cup \partial A$ is the *closure of* A; A is *closed* if $A = \overline{A}$. A set is closed if and only if it contains all of its limit points.

An open set A is *connected* if for every couple of points $\mathbf{x}, \mathbf{y} \in A$ there exists a regular curve joining them entirely contained in A. By a *domain* we mean an *open connected* set. Domains are usually denoted by the letter Ω.

---

[1] In general, $\omega_n = n\pi^{n/2}/\Gamma\left(\frac{1}{2}n + 1\right)$ where $\Gamma(s) = \int_0^{+\infty} t^{s-1}e^{-t}dt$ is the *Euler gamma function*.

If $U \subset A$, we say that U is *dense in* A if $\overline{U} = A$. This means that any point $\mathbf{x} \in A$ is a limit point of U. For instance, $\mathbb{Q}$ is dense in $\mathbb{R}$.

A is *bounded* if it is contained in some ball $B_r(\mathbf{0})$; it is *compact* if it is *closed and bounded*. If $\overline{A}_0$ is compact and contained in A, we write $A_0 \subset\subset A$ and we say that A_0 is *compactly contained* in A.

Infimum and supremum of a set of real numbers. A set $A \subset \mathbb{R}$ is *bounded from below* if there exists a number K such that

$$K \leq x \quad \text{for every } x \in A. \tag{1.7}$$

The greatest among the numbers K with the property (1.7) is called the *infimum* or *the greatest lower bound* of A and denoted by $\inf A$.

More precisely, we say that $\lambda = \inf A$ if $\lambda \leq x$ for every $x \in A$ and if, for every $\varepsilon > 0$, we can find $\bar{x} \in A$ such that $\bar{x} < \lambda + \varepsilon$. If $\inf A \in A$, then $\inf A$ is actually called the *minimum of* A, and may be denoted by $\min A$.

Similarly, $A \subset \mathbb{R}$ is *bounded from above* if there exists a number K such that

$$x \leq K \quad \text{for every } x \in A. \tag{1.8}$$

The smallest among the numbers K with the property (1.8) is called the *supremum* or the *lowest upper bound of* A and denoted by $\sup A$.

Precisely, we say that $\Lambda = \sup A$ if $\Lambda \geq x$ for every $x \in A$ and if, for every $\varepsilon > 0$, we can find $\bar{x} \in A$ such that $\bar{x} > \Lambda - \varepsilon$. If $\sup A \in A$, then $\sup A$ is actually called the *maximum of* A, and may be denoted by $\max A$.

Functions. Let $A \subseteq \mathbb{R}$ and $u : A \to \mathbb{R}$ be a real valued function defined in A. We say that u is *continuous* at $\mathbf{x} \in A$ if $u(\mathbf{y}) \to u(\mathbf{x})$ as $\mathbf{y} \to \mathbf{x}$. If u is continuous at any point of A we say that u is continuous in A. The set of such functions is denoted by $C(A)$.

The **support** of a continuous function is the *closure of the set where it is different from zero*. A continuous function is *compactly supported* in A if it vanishes outside a compact set contained in A.

We say that u is *bounded from below* (resp. *above*) in A if the image

$$u(A) = \{y \in \mathbb{R}, \, y = u(\mathbf{x}) \text{ for some } \mathbf{x} \in A\}$$

is *bounded from below* (resp. *above*). The infimum (supremum) of $u(A)$ is called the *infimum* (*supremum*) *of* u and is denoted by

$$\inf_{\mathbf{x} \in A} u(\mathbf{x}) \quad (\text{resp. } \sup_{\mathbf{x} \in A} u(\mathbf{x})).$$

We will denote by χ_A the *characteristic function of* A: $\chi_A = 1$ on A and $\chi_A = 0$ in $\mathbb{R}^n \backslash A$.

We use one of the symbols u_{x_j}, $\partial_{x_j} u$, $\dfrac{\partial u}{\partial x_j}$ for the first partial derivatives of u, and ∇u or $\operatorname{grad} u$ for the *gradient* of u. Accordingly, for the higher order derivatives we use the notations $u_{x_j x_k}$, $\partial_{x_j x_k} u$, $\dfrac{\partial^2 u}{\partial x_j \partial x_k}$ and so on.

We say that u is of class $C^k(\Omega)$, $k \geq 1$, or that it is a C^k−function, if u has continuous partials up to the order k (included) in the domain Ω. The class of continuously differentiable functions of any order in Ω, is denoted by $C^\infty(\Omega)$.

If $u \in C^1(\Omega)$ then u is differentiable in Ω and we can write, for $\mathbf{x} \in \Omega$ and $\mathbf{h} \in \mathbb{R}^n$ small:

$$u(\mathbf{x} + \mathbf{h}) - u(\mathbf{x}) = \nabla u(\mathbf{x}) \cdot \mathbf{h} + o(\mathbf{h})$$

where the symbol $o(\mathbf{h})$, "*little o of* $\mathbf{h}$", denotes a quantity such that $o(\mathbf{h})/|\mathbf{h}| \to 0$ as $|\mathbf{h}| \to 0$.

The symbol $C^k(\overline{\Omega})$ will denote the set of functions in $C^k(\Omega)$ whose derivatives up to the order k included can be extended continuously up to $\partial\Omega$.

Integrals. Up to Chapter 5 included, the integrals can be considered in the Riemann sense (proper or improper). A brief introduction to Lebesgue measure and integral is provided in Appendix B. Let $1 \leq p < \infty$ and $q = p/(p-1)$, the *conjugate exponent of* p. The following Hölder's inequality holds

$$\left| \int_\Omega uv \right| \leq \left(\int_\Omega |u|^p \right)^{1/p} \left(\int_\Omega |v|^q \right)^{1/q}. \tag{1.9}$$

The case $p = q = 2$ is known as the Schwarz inequality.

Uniform convergence. A series $\sum_{m=1}^\infty u_m$, where $u_m : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$, is said to be *uniformly convergent in* Ω, with *sum* u if, setting $S_N = \sum_{m=1}^N u_m$, we have

$$\sup_{\mathbf{x} \in \Omega} |S_N(\mathbf{x}) - u(\mathbf{x})| \to 0 \text{ as } N \to \infty.$$

Weierstrass test: Let $|u_m(\mathbf{x})| \leq a_m$, for every $m \geq 1$ and $\mathbf{x} \in \Omega$. If the numerical series $\sum_{m=1}^\infty a_m$ is convergent, then $\sum_{m=1}^\infty u_m$ converges absolutely and uniformly in Ω.

Limit and series. Let $\sum_{m=1}^\infty u_m$ be uniformly convergent in Ω. If u_m is continuous at $\mathbf{x}_0$ for every $m \geq 1$, then u is continuous at $\mathbf{x}_0$ and

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \sum_{m=1}^\infty u_m(\mathbf{x}) = \sum_{m=1}^\infty u_m(\mathbf{x}_0).$$

Term by term integration. Let $\sum_{m=1}^\infty u_m$ be uniformly convergent in Ω. If Ω is bounded and u_m is integrable in Ω for every $m \geq 1$, then:

$$\int_\Omega \sum_{m=1}^\infty u_m = \sum_{m=1}^\infty \int_\Omega u_m.$$

Term by term differentiation. Let Ω be bounded and $u_m \in C^1(\overline{\Omega})$ for every $m \geq 0$. If the series $\sum_{m=1}^\infty u_m(\mathbf{x}_0)$ is convergent at some $\mathbf{x}_0 \in A$ and the series $\sum_{m=1}^\infty \partial_{x_j} u_m$ are uniformly convergent in $\overline{\Omega}$ for every $j = 1, ..., n$, then $\sum_{m=1}^\infty u_m$ converges uniformly in $\overline{\Omega}$, with sum in $C^1(\overline{\Omega})$ and

$$\partial_{x_j} \sum_{m=1}^\infty u_m(\mathbf{x}) = \sum_{m=1}^\infty \partial_{x_j} u_m(\mathbf{x}) \qquad (j = 1, ..., n).$$

1.5 Smooth and Lipschitz Domains

We will need, especially in Chapters 7, 8 and 9, to distinguish the domains Ω in $\mathbb{R}^n$ according to the degree of smoothness of their boundary (Fig. 1.2).

Definition 1.1. *We say that Ω is a C^1-domain if for every point $\mathbf{x} \in \partial\Omega$, there exist a system of coordinates $(y_1, y_2, ..., y_{n-1}, y_n) \equiv (\mathbf{y}', y_n)$ with origin at $\mathbf{x}$, a ball $B(\mathbf{x})$ and a function φ defined in a neighborhood $\mathcal{N} \subset \mathbb{R}^{n-1}$ of $\mathbf{y}' = \mathbf{0}'$, such that*

$$\varphi \in C^1(\mathcal{N}), \; \varphi(\mathbf{0}') = 0$$

and
   1. $\partial\Omega \cap B(\mathbf{x}) = \{(\mathbf{y}', y_n) : y_n = \varphi(\mathbf{y}'), \mathbf{y}' \in \mathcal{N}\}$,
   2. $\Omega \cap B(\mathbf{x}) = \{(\mathbf{y}', y_n) : y_n > \varphi(\mathbf{y}'), \mathbf{y}' \in \mathcal{N}\}$.

The first condition expresses the fact that $\partial\Omega$ locally coincides with the graph of a C^1-function. The second one requires that Ω be locally placed on one side of its boundary.

The boundary of a C^1-domain does not have corners or edges and for every point $\mathbf{p} \in \partial\Omega$, a tangent straight line $(n = 2)$ or plane $(n = 3)$ or hyperplane $(n > 3)$ is well defined, *together* with the *outward* and *inward* normal unit vectors. Moreover these vectors vary continuously on $\partial\Omega$.

The couples $(\varphi, \mathcal{N})$ appearing in the above definition are called *local charts*. If they are all C^k-functions, for some $k \geq 1$, Ω is said to be a $C^k-domain$. If Ω is a C^k-domain for every $k \geq 1$, it is said to be a $C^\infty-domain$. These are the domains we consider **smooth** domains.

Observe that the one-to-one transformation (*diffeomorfism*) $\mathbf{z} = \boldsymbol{\Phi}(\mathbf{y})$ given by

$$\begin{cases} \mathbf{z}' = \mathbf{y}' \\ z_n = y_n - \varphi(\mathbf{y}') \end{cases} \tag{1.10}$$

maps $\partial\Omega \cap B(\mathbf{x})$ into a subset of the hyperplane $z_n = 0$, so that $\partial\Omega \cap B(\mathbf{x})$ *straightens*, as shown in figure 1.1.
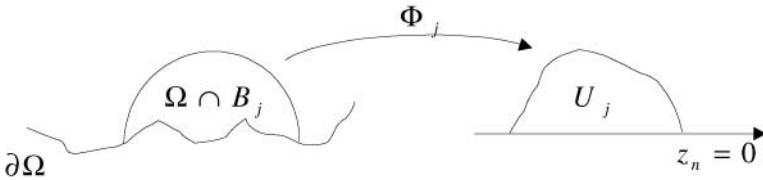
Fig. 1.1. Straightening the boundary $\partial\Omega$ by a diffeomorphism

In a great number of applications the relevant domains are rectangles, prisms, cones, cylinders or unions of them. Very important are polygonal domains obtained by *triangulation* procedures of smooth domains, for numerical approximations.
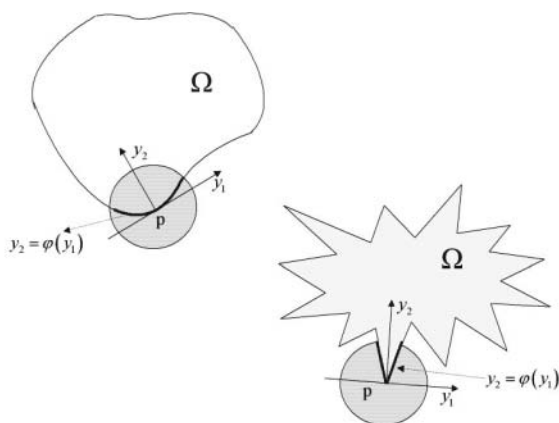
Fig. 1.2. A C^1 domain and a Lipschitz domain

These types of domains belong to the class of *Lipschitz domains*, whose boundary is locally described by the graph of a *Lipschitz function*.

Definition 1.2. *We say that* $u : \Omega \to \mathbb{R}^n$ *is Lipschitz if there exists* L *such that*

$$|u(\mathbf{x}) - u(\mathbf{y})| \leq L |\mathbf{x} - \mathbf{y}|$$

for every $\mathbf{x}, \mathbf{y} \in \Omega$. *The number* L *is called the Lipschitz constant of* u.

Roughly speaking, a function is Lipschitz in Ω if the increment quotients in every direction are bounded. In fact, Lipschitz functions are differentiable at all points of their domain with the exception of a negligible set of points. Precisely, we have (see e.g. *Evans and Gariepy*, 1997):

Theorem 1.1. (Rademacher). *Let* u *be a Lipschtz function in* $A \subseteq \mathbb{R}^n$. *Then* u *is differentiable at every point of* A, *except at a set points of Lebesgue measure zero.*

Typical real Lipschitz functions in $\mathbb{R}^n$ are $f(\mathbf{x}) = |\mathbf{x}|$ or, more generally, the *distance function from a closed set*, C, defined by

$$f(\mathbf{x}) = \text{dist}(\mathbf{x}, C) = \inf_{\mathbf{y} \in C} |\mathbf{x} - \mathbf{y}|.$$

We say that a **domain is Lipschitz** if in Definition 1.1 the functions φ *are Lipschitz* or, equivalently, if the map (1.10) is a *bi-Lipschitz* transformation, that is, both $\mathbf{\Phi}$ and $\mathbf{\Phi}^{-1}$ are Lipschitz.

1.6 Integration by Parts Formulas

Let $\Omega \subset \mathbb{R}^n$, be a C^1- domain. For vector fields

$$\mathbf{F} = (F_1, F_2, ..., F_n) : \Omega \to \mathbb{R}^n$$

with $\mathbf{F} \in C^1 \left(\overline{\Omega} \right)$, the **Gauss divergence formula** holds:

$$\int_\Omega \mathrm{div}\mathbf{F} \ d\mathbf{x} = \int_{\partial\Omega} \mathbf{F} \cdot \boldsymbol{\nu} \ d\sigma \qquad (1.11)$$

where $\mathrm{div}\mathbf{F} = \sum_{j=1}^n \partial_{x_j} F_j$, $\boldsymbol{\nu}$ denotes the *outward normal* unit vector to $\partial\Omega$, and $d\sigma$ is the "surface" measure on $\partial\Omega$, locally given in terms of local charts by

$$d\sigma = \sqrt{1 + |\nabla\varphi\left(\mathbf{y}'\right)|} \ d\mathbf{y}'.$$

A number of useful identities can be derived from (1.11). Applying (1.11) to $v\mathbf{F}$, with $v \in C^1 \left(\overline{\Omega} \right)$, and recalling the identity

$$\mathrm{div}(v\mathbf{F}) = v \ \mathrm{div}\mathbf{F} + \nabla v \cdot \mathbf{F}$$

we obtain the following **integration by parts** formula:

$$\int_\Omega v \ \mathrm{div}\mathbf{F} \ d\mathbf{x} = \int_\Omega v\mathbf{F} \cdot \boldsymbol{\nu} \ d\sigma - \int_\Omega \nabla v \cdot \mathbf{F} \ d\mathbf{x}. \qquad (1.12)$$

Choosing $\mathbf{F} = \nabla u$, $u \in C^2 \left(\Omega \right) \cap C^1 \left(\overline{\Omega} \right)$, since $\mathrm{div}\nabla u = \Delta u$ and $\nabla u \cdot \boldsymbol{\nu} = \partial_{\boldsymbol{\nu}} u$, the following **Green's identity** follows:

$$\int_\Omega v\Delta u \ d\mathbf{x} = \int_{\partial\Omega} v\partial_{\boldsymbol{\nu}} u \ d\sigma - \int_\Omega \nabla v \cdot \nabla u \ d\mathbf{x}. \qquad (1.13)$$

In particular, the choice $v \equiv 1$ yields

$$\int_\Omega \Delta u \ d\mathbf{x} = \int_{\partial\Omega} \partial_{\boldsymbol{\nu}} u \ d\sigma. \qquad (1.14)$$

If also $v \in C^2 \left(\Omega \right) \cap C^1 \left(\overline{\Omega} \right)$, interchanging the roles of u and v in (1.13) and subtracting, we derive a second **Green's identity**:

$$\int_\Omega (v\Delta u - u\Delta v) \ d\mathbf{x} = \int_{\partial\Omega} (v\partial_{\boldsymbol{\nu}} u - u\partial_{\boldsymbol{\nu}} v) \ d\sigma. \qquad (1.15)$$

Remark 1.1. All the above formulas hold for Lipschitz domains as well. In fact, the Rademacher theorem implies that at every point of the boundary of a Lipschitz domain, with the exception of a set of points of surface measure zero, there is a well defined tangent plane. This is enough for extending the formulas (1.12), (1.13) and (1.15) to Lipchitz domains.

2

Diffusion

The Diffusion Equation – Uniqueness – The Fundamental Solution – Symmetric Random Walk ($n = 1$) – Diffusion, Drift and Reaction – Multidimensional Random Walk – An Example of Reaction−Diffusion ($n = 3$) – The Global Cauchy Problem ($n = 1$) – An Application to Finance – Some Nonlinear Aspects

2.1 The Diffusion Equation

2.1.1 Introduction

The one-dimensional **diffusion equation** is the *linear second order partial differential equation*

$$u_t - Du_{xx} = f$$

where $u = u(x, t)$, x is a real space variable, t a time variable and D a positive constant, called **diffusion coefficient**. In space dimension $n > 1$, that is when $\mathbf{x} \in \mathbb{R}^n$, the diffusion equation reads

$$u_t - D\Delta u = f \qquad (2.1)$$

where Δ denotes the *Laplace operator*:

$$\Delta = \sum_{k=1}^{n} \frac{\partial^2}{\partial x_k^2}.$$

When $f \equiv 0$ the equation is said to be **homogeneous** and in this case the **superposition principle** holds: if u and v are solutions of (2.1) and a, b are real (or complex) numbers, $au + bv$ also is a solution of (2.1). More generally, if $u_k(\mathbf{x}, t)$ is a family of solutions depending on the parameter k (integer or real) and $g = g(k)$ is a function rapidly vanishing at infinity, then

$$\sum_{k=1}^{\infty} u_k(\mathbf{x}, t) g(k) \quad \text{and} \quad \int_{-\infty}^{+\infty} u_k(\mathbf{x}, t) g(k) \, dk$$

are still solutions.

A common example of diffusion is given by *heat conduction* in a solid body. Conduction comes from molecular collision, transferring heat by kinetic energy, without macroscopic material movement. If the medium is homogeneous and isotropic with respect to the heat propagation, the evolution of the temperature is described by equation (2.1); f represents the intensity of an external distributed source. For this reason equation (2.1) is also known as **the heat equation**.

On the other hand equation (2.1) constitutes a much more general diffusion model, where by **diffusion** we mean, for instance, the *transport of a substance due to the molecular motion of the surrounding medium.* In this case, u could represent the concentration of a polluting material or of a solute in a liquid or a gas (dye in a liquid, smoke in the atmosphere) or even a probability density. We may say that the diffusion equation unifies at a macroscopic scale a variety of phenomena, that look quite different when observed at a microscopic scale.

Through equation (2.1) and some of its variants we will explore the deep connection between probabilistic and deterministic models, according (roughly) to the scheme

diffusion processes ↔ probability density ↔ differential equations.

The *star* in this field is *Brownian motion,* derived from the name of the botanist Brown, who observed in the middle of the 19th century, the apparently chaotic behavior of certain particles on a water surface, due to the molecular motion. This irregular motion is now modeled as a *stochastic process* under the terminology of *Wiener process or Brownian motion.* The operator

$$\frac{1}{2}\Delta$$

is strictly related to Brownian motion[1] and indeed it captures and synthesizes the microscopic features of that process.

Under equilibrium conditions, that is when there is no time evolution, the solution u depends only on the space variable and satisfies the *stationary* version of the diffusion equation (letting $D = 1$)

$$-\Delta u = f \tag{2.2}$$

($-u_{xx} = f$, in dimension $n = 1$). Equation (2.2) is known as the *Poisson equation.* When $f = 0$, it is called *Laplace's equation* and its solutions are so important in so many fields that they have deserved the special name of **harmonic functions**. This equation will be considered in the next chapter.

2.1.2 The conduction of heat

Heat is a form of energy which it is frequently convenient to consider as separated from other forms. For historical reasons, *calories* instead of Joules are used as units of measurement, each *calorie* corresponding to 4.182 Joules.

---

[1] In the theory of stochastic processes, $\frac{1}{2}\Delta$ represents the *infinitesimal generator of the Brownian motion.*

We want to derive a mathematical model for the heat conduction in a solid body. We assume that the body is homogeneous and isotropic, with constant *mass density* ρ, and that it can receive energy from an external source (for instance, from an electrical current or a chemical reaction or from external absorption/radiation). Denote by r the time rate per unit mass at which heat is supplied[2] by the external source.

Since heat is a form of energy, it is natural to use the law of conservation of energy, that we can formulate in the following way:

Let V be an arbitrary control volume inside the body. *The time rate of change of thermal energy in V equals the net flux of heat through the boundary ∂V of V, due to the conduction, plus the time rate at which heat is supplied by the external sources.*

If we denote by $e = e(\mathbf{x}, t)$ the thermal energy per unit mass, the total quantity of thermal energy inside V is given by

$$\int_V e\rho \, d\mathbf{x}$$

so that its time rate of change is[3]

$$\frac{d}{dt} \int_V e\rho \, d\mathbf{x} = \int_V e_t \rho \, d\mathbf{x}.$$

Denote by $\mathbf{q}$ the *heat flux* vector[4], which specifies the heat flow direction and the magnitude of the rate of flow across a unit area. More precisely, if $d\sigma$ is an area element contained in ∂V with *outer* unit normal $\boldsymbol{\nu}$, then $\mathbf{q} \cdot \boldsymbol{\nu} d\sigma$ is the energy flow rate through $d\sigma$ and therefore the *total inner heat flux* through ∂V is given by

$$-\int_{\partial V} \mathbf{q} \cdot \boldsymbol{\nu} \, d\sigma \underset{\text{(divergence theorem)}}{=} -\int_V \operatorname{div}\mathbf{q} \, d\mathbf{x}.$$

Finally, the contribution due to the external source is given by

$$\int_V r\rho \, d\mathbf{x}.$$

Thus, conservation of energy requires:

$$\int_V e_t \rho \, d\mathbf{x} = -\int_V \operatorname{div}\mathbf{q} \, d\mathbf{x} + \int_V r\rho \, d\mathbf{x}. \tag{2.3}$$

The arbitrariness of V allows us to convert the integral equation (2.3) into the pointwise relation

$$e_t \rho = -\operatorname{div}\mathbf{q} + r\rho \tag{2.4}$$

---

[2] Dimensions of r: $[r] = [cal] \times [time]^{-1} \times [mass]^{-1}$.
[3] Assuming that the time derivative can be carried inside the integral.
[4] $[\mathbf{q}] = [cal] \times [lenght]^{-2} \times [time]^{-1}$.

that constitutes a basic law of heat conduction. However, e and $\mathbf{q}$ are unknown and we need additional information through *constitutive relations* for these quantities. We assume the following:

• **Fourier law** of heat conduction. Under "normal" conditions, for many solid materials, the heat flux is a linear function of the temperature gradient, that is:

$$\mathbf{q} = -\kappa \nabla u \qquad (2.5)$$

where u is the absolute temperature and $\kappa > 0$, the *thermal conductivity*[5], depends on the properties of the material. In general, κ may depend on u, $\mathbf{x}$ and t, but often varies so little in cases of interest that it is reasonable to neglect its variation. Here we consider κ *constant* so that

$$\mathrm{div}\mathbf{q} = -\kappa \Delta u. \qquad (2.6)$$

The minus sign in the law (2.5) reflects the tendency of heat to flow from hotter to cooler regions.

• The thermal energy is a linear function of the absolute temperature:

$$e = c_v u \qquad (2.7)$$

where c_v denotes the *specific heat*[6] (at constant volume) of the material. In many cases of interest c_v can be considered constant. The relation (2.7) is reasonably true over not too wide ranges of temperature.

Using (2.6) and (2.7), equation (2.4) becomes

$$u_t = \frac{\kappa}{c_v \varrho} \Delta u + \frac{1}{c_v} r \qquad (2.8)$$

which is the diffusion equation with $D = \kappa / (c_v \varrho)$ and $f = r/c_v$. As we will see, the coefficient D, called *thermal diffusivity*, encodes the thermal response time of the material.

2.1.3 Well posed problems ($n = 1$)

As we have mentioned at the end of chapter one, the governing equations in a mathematical model have to be supplemented by additional information in order to obtain a *well posed problem*, i.e. a problem that has exactly one solution, depending continuously on the data.

On physical grounds, it is not difficult to outline some typical well posed problems for the heat equation. Consider the evolution of the temperature u inside a cylindrical bar, whose lateral surface is *perfectly insulated* and whose length is much larger than its cross-sectional area A. Although the bar is three dimensional,

---

[5] $[\kappa] = [cal] \times [deg]^{-1} \times [time]^{-1} \times [length]^{-1}$ (deg stays for degree, Celsius or Kelvin).
[6] $[c_v] = [cal] \times [deg]^{-1} \times [mass]^{-1}$.

we may assume that heat moves only down the length of the bar and that the heat transfer intensity is uniformly distributed in each section of the bar. Thus we may assume that $e = e\left(x, t\right), r = r\left(x, t\right)$, with $0 \leq x \leq L$. Accordingly, the constitutive relations (2.5) and (2.7) read

$$e\left(x, t\right) = c_v u\left(x, t\right), \quad \mathbf{q} = -\kappa u_x \mathbf{i}.$$

By choosing $V = A \times \left[x, x + \Delta x\right]$ as the control volume in (2.3), the cross-sectional area A cancels out, and we obtain

$$\int_x^{x+\Delta x} c_v \rho u_t \ dx = \int_x^{x+\Delta x} \kappa u_{xx} \ dx + \int_x^{x+\Delta x} r\rho \ dx$$

that yields for u the one-dimensional heat equation

$$u_t - D u_{xx} = f.$$

We want to study the temperature evolution during an interval of time, say, from $t = 0$ until $t = T$. It is then reasonable to prescribe its initial distribution inside the bar: different initial configurations will correspond to different evolutions of the temperature along the bar. Thus we need to prescribe **the initial condition**

$$u\left(x, 0\right) = g\left(x\right)$$

where g models the initial temperature profile.

   This is not enough to determine a unique evolution; it is necessary to know how the bar interacts with the surroundings. Indeed, starting with a given initial temperature distribution, we can change the evolution of u by controlling the temperature or the heat flux at the two ends of the bar[7]; for instance, we could keep the temperature at a certain fixed level or let it vary in a certain way, depending on time. This amounts to prescribing

$$u\left(0, t\right) = h_1\left(t\right), \quad u\left(L, t\right) = h_2\left(t\right) \tag{2.9}$$

at any time $t \in \left(0, T\right]$. The (2.9) are called **Dirichlet boundary conditions** .

   We could also prescribe the heat flux at the end points. Since from Fourier law we have

$$\text{inward heat flow at } x = 0 : -\kappa u_x\left(0, t\right)$$

$$\text{inward heat flow at } x = L : \kappa u_x\left(L, t\right)$$

the heat flux is assigned through the **Neumann boundary conditions**

$$-u_x\left(0, t\right) = h_1\left(t\right), u_x\left(L, t\right) = h_2\left(t\right)$$

at any time $t \in \left(0, T\right]$.

---

[7] Remember that the bar has perfect lateral thermal insulation.

Another type of boundary condition is the **Robin** or **radiation condition**. Let the surroundings be kept at temperature U and assume that the *inward* heat flux from one end of the bar, say $x = L$, depends linearly on the difference $U - u$, that is[8]

$$\kappa u_x = \gamma(U - u) \qquad (\gamma > 0). \tag{2.10}$$

Letting $\alpha = \gamma/\kappa > 0$ e $h = \gamma U/\kappa$, the Robin condition at $x = L$ reads

$$u_x + \alpha u = h.$$

Clearly, it is possible to assign **mixed conditions**: for instance, at one end a Dirichlet condition and at the other one a Neumann condition.

The problems associated with the above boundary conditions have a corresponding nomenclature. Summarizing, we can state the most common problems for the one dimensional heat equation as follows: *given* $f = f(x,t)$ *(external source) and* $g = g(x)$ *(initial or Cauchy data), determine* $u = u(x,t)$ *such that:*

$$\begin{cases} u_t - Du_{xx} = f & 0 < x < L, 0 < t < T \\ u(x,0) = g(x) & 0 \leq x \leq L \\ + \text{ boundary conditions} & 0 < t \leq T \end{cases}$$

where the boundary conditions may be:

- *Dirichlet*:
$$u(0,t) = h_1(t), \ u(L,t) = h_2(t),$$

- *Neumann*:
$$-u_x(0,t) = h_1(t), \ u_x(L,t) = h_2(t),$$

- *Robin or radiation*:
$$-u_x(0,t) + \alpha u(0,t) = h_1(t), \ u_x(L,t) + \alpha u(L,t) = h_2(t) \qquad (\alpha > 0),$$

or *mixed* conditions. Accordingly, we have the initial-Dirichlet problem, the initial-Neumann problem and so on. When $h_1 = h_2 = 0$, we say that the boundary conditions are **homogeneous.**

Remark 2.1. Observe that only a special part of the boundary of the rectangle

$$Q_T = (0,L) \times (0,T),$$

called the *parabolic boundary of* Q_T, carries the data (see Fig. 2.1). *No final condition (for* $t = T, 0 < x < L$) *is required.*

---

[8] Formula (2.10) is based on *Newton's law of cooling*: the heat loss from the surface of a body is a linear function of the temperature drop $U - u$ from the surroundings to the surface. It represents a good approximation to the radiative loss from a body when $|U - u|/u \ll 1$.
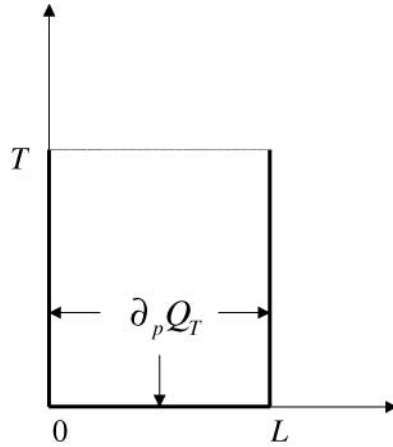
Fig. 2.1. The parabolic boundary of Q_T

In important applications, for instance in financial mathematics, x varies over unbounded intervals, typically $(0, \infty)$ or $\mathbb{R}$. In these cases one has to require that the solution do not grow too much at infinity. We will later consider the global Cauchy problem:

$$\begin{cases} u_t - Du_{xx} = f & x \in \mathbb{R}, 0 < t < T \\ u(x,0) = g(x) & x \in \mathbb{R} \\ + \text{ conditions as } x \to \pm\infty. \end{cases}$$

2.1.4 A solution by separation of variables

We will prove that under reasonable hypotheses the initial Dirichlet, Neumann or Robin problems are well posed. Sometimes this can be shown using elementary techniques like *the separation of variables method* that we describe below through a simple example of heat conduction. We will come back to this method from a more general point of view in Section 6.9.

As in the previous section, consider a bar (that we can consider one-dimensional) of length L, initially (at time $t = 0$) at constant temperature u_0. Thereafter, the end point $x = 0$ is kept at the same temperature while the other end $x = L$ is kept at a constant temperature $u_1 > u_0$. We want to know how the temperature evolves inside the bar.

Before making any computations, let us try to conjecture what could happen. Given that $u_1 > u_0$, heat starts flowing from the hotter end, raising the temperature inside the bar and causing a heat outflow into the cold boundary. On the other hand, the interior increase of temperature causes the hot inflow to decrease in time, while the ouflow increases. We expect that sooner or later the two fluxes balance each other and that the temperature eventually reaches a steady state

distribution. It would also be interesting to know how fast the steady state is reached.

We show that this is exactly the behavior predicted by our mathematical model, given by the heat equation

$$u_t - Du_{xx} = 0 \quad t > 0, 0 < x < L$$

with the initial-Dirichlet conditions

$$u(x, 0) = g(x) \qquad\qquad 0 \leq x \leq L$$
$$u(0, t) = u_0, \, u(L, t) = u_1 \qquad t > 0.$$

Since we are interested in the long term behavior of our solution, we leave t unlimited. Notice the *jump discontinuity* between the initial and the boundary data at $x = L$; we will take care of this little difficulty later.

• *Dimensionless variables.* First of all we introduce dimensionless variables, that is variables *independent of the units of measurement*. To do that we rescale space, time and temperature with respect to quantities that are characteristic of our problem. For the space variable we can use the length L of the bar as rescaling factor, setting

$$y = \frac{x}{L}$$

which is clearly dimensionless, being a ratio of lengths. Notice that

$$0 \leq y \leq 1.$$

How can we rescale time? Observe that the dimensions of the diffusion coefficient D are

$$[length]^2 \times [time]^{-1}.$$

Thus the constant $\tau = L^2/D$ gives a characteristic time scale for our diffusion problem. Therefore we introduce the dimensionless time

$$s = \frac{t}{\tau}. \tag{2.11}$$

Finally, we rescale the temperature by setting

$$z(y, s) = \frac{u(Ly, \tau s) - u_0}{u_1 - u_0}.$$

For the dimensionless temperature z we have:

$$z(y, 0) = \frac{u(Ly, 0) - u_0}{u_1 - u_0} = 0, \quad 0 \leq y \leq 1$$

$$z(0, s) = \frac{u(0, \tau s) - u_0}{u_1 - u_0} = 0, \quad z(1, s) = \frac{u(L, \tau s) - u_0}{u_1 - u_0} = 1.$$

Moreover

$$(u_1 - u_0)z_s = \frac{\partial t}{\partial s}u_t = \tau u_t = \frac{L^2}{D}u_t$$

$$(u_1 - u_0)z_{yy} = \left(\frac{\partial x}{\partial y}\right)^2 u_{xx} = L^2 u_{xx}.$$

Hence, since $u_t = Du_{xx}$,

$$(u_1 - u_0)(z_s - z_{yy}) = \frac{L^2}{D}u_t - L^2 u_{xx} = \frac{L^2}{D}Du_{xx} - L^2 u_{xx} = 0.$$

In conclusion, we find

$$z_s - z_{yy} = 0 \tag{2.12}$$

with the initial condition

$$z(y,0) = 0 \tag{2.13}$$

and the boundary conditions

$$z(0,s) = 0, \quad z(1,s) = 1. \tag{2.14}$$

We see that in the dimensionless formulation the parameters L and D have disappeared, emphasizing the mathematical essence of the problem. On the other hand, we will show later the relevance of the dimensionless variables in test modelling.

• *The steady state solution* . We start solving problem (2.12), (2.13), (2.14) by first determining the steady state solution z^{St}, that satisfies the equation $z_{yy} = 0$ and the boundary conditions (2.14). An elementary computation gives

$$z^{St}(y) = y.$$

In terms of the original variables the steady state solution is

$$u^{St}(x) = u_0 + (u_1 - u_0)\frac{x}{L}$$

corresponding to a uniform heat flux along the bar given by the Fourier law:

$$\text{heat flux} = -\kappa u_x = -\kappa \frac{(u_1 - u_0)}{L}.$$

• *The transient regime.* Knowing the steady state solution, it is convenient to introduce the function

$$U(y,s) = z^{St}(y,s) - z(y,s) = y - z(y,s).$$

Since we expect our solution to eventually reach the steady state, U represents a *transient regime* that should converge to zero as $s \to \infty$. Furthermore, the rate of convergence to zero of U gives information on how fast the temperature reaches its equilibrium distribution. U satisfies (2.12) with initial condition

$$U(y,0) = y \tag{2.15}$$

and *homogeneous* boundary conditions

$$U(0, s) = 0 \quad \text{and} \quad U(1, s) = 0. \tag{2.16}$$

• *The method of separation of variables.* We are now in a position to find an explicit formula for U using the method of separation of variables. The main idea is to exploit the linear nature of the problem constructing the solution by superposition of simpler solutions of the form $w(s) v(y)$ in which the variables s and y appear in *separated form.*

Step **1**. We look for non-trivial solutions of (2.12) of the form

$$U(y, s) = w(s) v(y)$$

with $v(0) = v(1) = 0$. By substitution in (2.12) we find

$$0 = U_s - U_{yy} = w'(s) v(y) - w(s) v''(y)$$

from which, separating the variables,

$$\frac{w'(s)}{w(s)} = \frac{v''(y)}{v(y)}. \tag{2.17}$$

Now, the left hand side in (2.17) is a function of s only, while the right hand side is a function of y only and the equality must hold for every $s > 0$ and every $y \in (0, L)$. This is possible only when both sides are equal to a common constant λ, say. Hence we have

$$v''(y) - \lambda v(y) = 0 \tag{2.18}$$

with

$$v(0) = v(1) = 0 \tag{2.19}$$

and

$$w'(s) - \lambda w(s) = 0. \tag{2.20}$$

Step **2**. We first solve problem (2.18), (2.19). There are three different possibilities for the general solution of (2.18):

a) If $\lambda = 0$,

$$v(y) = A + By \qquad (A, B \text{ arbitrary constants})$$

and the conditions (2.19) imply $A = B = 0$.

b) If λ is a positive real number, say $\lambda = \mu^2 > 0$, then

$$v(y) = Ae^{-\mu y} + Be^{\mu y}$$

and again it is easy to check that the conditions (2.19) imply $A = B = 0$.

c) Finally, if $\lambda = -\mu^2 < 0$, then

$$v(y) = A \sin \mu y + B \cos \mu y.$$

From (2.19) we get

$$v(0) = B = 0$$
$$v(1) = A\sin\mu + B\cos\mu = 0$$

from which

$$A \text{ arbitrary}, \ B = 0, \ \mu_m = m\pi, \ m = 1, 2, \ldots .$$

Thus, only in case c) we find non-trivial solutions

$$v_m(y) = A\sin m\pi y. \tag{2.21}$$

In this context, (2.18), (2.19) is called an *eigenvalue problem*; the special values μ_m are the *eigenvalues* and the solutions v_m are the corresponding *eigenfunctions*.
With $\lambda = -\mu_m^2 = -m^2\pi^2$, the general solution of (2.20) is

$$w_m(s) = Ce^{-m^2\pi^2 s} \qquad (C \text{ arbitrary constant}). \tag{2.22}$$

From (2.21) and (2.22) we obtain damped sinusoidal waves of the form

$$U_m(y, s) = A_m e^{-m^2\pi^2 s}\sin m\pi y.$$

Step **3**. Although the solutions U_m satisfy the homogeneous Dirichlet conditions, they do not match, in general, the initial condition $U(y, 0) = y$. As we already mentioned, we try to construct the correct solution superposing the U_m by setting

$$U(y, s) = \sum_{m=1}^{\infty} A_m e^{-m^2\pi^2 s}\sin m\pi y. \tag{2.23}$$

Some questions arise:

Q1. The initial condition requires

$$U(y, 0) = \sum_{m=1}^{\infty} A_m \sin m\pi y = y \qquad \text{for } 0 \le y \le 1. \tag{2.24}$$

Is it possible to choose the coefficients A_m in order to satisfy (2.24)? In which sense does U attain the initial data? For instance, is it true that

$$U(z, s) \to y \quad \text{if} \quad (z, s) \to (y, 0)?$$

Q2. Any finite linear combination of the U_m is a solution of the heat equation; can we make sure that the same is true for U? The answer is positive if we could differentiate term by term the infinite sum and get

$$(\partial_s - \partial_{yy}^2)U(y, s) = \sum_{m=1}^{\infty}(\partial_s - \partial_{yy})U_m(y, s) = 0. \tag{2.25}$$

What about the boundary conditions?

Q3. Even if we have a positive answer to questions 1 and 2, are we confident that U is the unique solution of our problem and therefore that it describes the correct evolution of the temperature?

Q1. Question 1 is rather general and concerns the *Fourier series expansion*[9] of a function, in particular of the initial data $f(y) = y$, in the interval $(0,1)$. Due to the homogeneous Dirichlet conditions it is convenient to expand $f(y) = y$ in a *sine Fourier series*, whose coefficients are given by the formulas

$$A_m = 2 \int_0^1 y \sin m\pi y \, dy = -\frac{2}{m\pi} [y \cos m\pi y]_0^1 + \frac{2}{m\pi} \int_0^1 \cos m\pi y \, dy =$$

$$= -2\frac{\cos m\pi}{m\pi} = (-1)^{m+1} \frac{2}{m\pi}.$$

The sine Fourier expansion of $f(y) = y$ is therefore

$$y = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{2}{m\pi} \sin m\pi y. \tag{2.26}$$

Where is the expansion (2.26) valid? It cannot be true at $y = 1$ since $\sin m\pi = 0$ for every m and we would obtain $1 = 0$. This clearly reflects the jump discontinuity of the data at $y = 1$.

The theory of Fourier series implies that (2.26) is true at every point $y \in [0, 1)$ and that the series converges uniformly in every interval $[0, a]$, $a < 1$. Moreover, equality (2.26) holds **in the leat square sense** (or $L^2(0,1)$ sense), that is

$$\int_0^1 [y - \sum_{m=1}^{N} (-1)^{m+1} \frac{2}{m\pi} \sin m\pi y]^2 dy \to 0 \qquad \text{as } N \to \infty.$$

From (2.23) and the expression of A_m, we obtain the *formal* solution

$$U(y,s) = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{2}{m\pi} e^{-m^2\pi^2 s} \sin m\pi y \tag{2.27}$$

that attains the initial data in the least squares sense, i.e.[10].

$$\lim_{s \to 0^+} \int_0^1 [U(y,s) - y]^2 dy = 0. \tag{2.28}$$

In fact, from Parseval's equality[11], we can write

$$\int_0^1 [U(y,s) - y]^2 \, dy = \frac{4}{\pi^2} \sum_{m=1}^{\infty} \frac{\left(e^{-m^2\pi^2 s} - 1\right)^2}{m^2}. \tag{2.29}$$

---

[9] Appendix A.
[10] It is also true that $U(z,s) \to y$ in the pointwise sense, when $y \neq 1$ and $(z,s) \to (y,0)$. We omit the proof.
[11] Appendix A.

Since for $s \geq 0$

$$\frac{\left(e^{-m^2\pi^2 s} - 1\right)^2}{m^2} \leq \frac{1}{m^2}$$

and the series $\sum 1/m^2$ converges, then the series (2.29) converges uniformly by the Weierstrass test (see Section 1.4) in $[0, \infty)$ and we can take the limit under the sum, obtaining (2.28).

Q2. The analytical expression of U is rather reassuring: it is a superposition of sinusoids of increasing frequency m and of strongly damped amplitude because of the negative exponential, at least when $s > 0$. Indeed, for $s > 0$, the rapid convergence to zero of each term and its derivatives in the series (2.27) allows us to differentiate term by term. Precisely, we have

$$\frac{\partial U_m}{\partial s} = (-1)^{m+2} \, 2m\pi e^{-m^2\pi^2 s} \sin m\pi y, \quad \frac{\partial^2 U_m}{\partial y^2} = (-1)^{m+2} \, 2 e^{-m^2\pi^2 s} \sin m\pi y$$

so that, if $s \geq s_0 > 0$,

$$\left|\frac{\partial U_m}{\partial s}\right| \leq 2m\pi e^{-m^2\pi^2 s_0}, \quad \left|\frac{\partial^2 U_m}{\partial y^2}\right| \leq 2m\pi e^{-m^2\pi^2 s_0}.$$

Since the numerical series

$$\sum_{m=1}^{\infty} m e^{-m^2\pi^2 s_0}$$

is convergent, we conclude by the Weierstrass test that the series

$$\sum_{m=1}^{\infty} \frac{\partial U_m}{\partial s} \quad \text{and} \quad \sum_{m=1}^{\infty} \frac{\partial^2 U_m}{\partial y^2}$$

converge uniformly in $[0, 1] \times [s_0, \infty)$ so that (2.25) is true and therefore U is a solution of (2.12).

It remains to check the Dirichlet conditions: if $s_0 > 0$,

$$U(z, s) \to 0 \quad \text{as} \quad (z, s) \to (0, s_0) \quad \text{or} \quad (z, s) \to (L, s_0).$$

This is true because we can take the two limits under the sum, due to the uniform convergence of the series (2.27) in any region $[0, L] \times (b, +\infty)$ with $b > 0$. For the same reason, U has continuous derivatives of any order, up to the lateral boundary of the strip $[0, L] \times (b, +\infty)$.

Note, in particular, that U *immediately* forgets the initial discontinuity and becomes smooth at any positive time.

Q3. To show that U is indeed the unique solution, we use the so-called *energy method*, that we will develop later in greater generality. Suppose W is another solution of problem (2.12), (2.15), (2.16). Then, by linearity,

$$v = U - W$$

satisfies

$$v_s - v_{yy} = 0 \tag{2.30}$$

and has zero initial-boundary data. Multiplying (2.30) by v, integrating in y over the interval $[0, 1]$ and keeping $s > 0$, fixed, we get

$$\int_0^1 vv_s \, dy - \int_0^1 vv_{yy} \, dy = 0. \tag{2.31}$$

Observe that

$$\int_0^1 vv_s \, dy = \frac{1}{2} \int_0^1 \partial_s \left(v^2 \right) dy = \frac{1}{2} \frac{d}{ds} \int_0^1 v^2 dy. \tag{2.32}$$

Moreover, integrating by parts we can write

$$\int_0^1 vv_{yy} \, dy = [v(1, s) v_y(1, s) - v(0, s) v_y(0, s)] - \int_0^1 (v_y)^2 \, dy \tag{2.33}$$

$$= - \int_0^1 (v_y)^2 \, dy$$

since $v(1, s) = v(0, s) = 0$. From (2.31), (2.32) and (2.33) we get

$$\frac{1}{2} \frac{d}{ds} \int_0^1 v^2 dy = - \int_0^1 (v_y)^2 \, dy \leq 0 \tag{2.34}$$

and therefore, the *nonnegative* function

$$E(s) = \int_0^1 v^2(y, s) \, dy$$

is non-increasing. On the other hand, using (2.28) for v instead of U, we get

$$E(s) \to 0 \qquad \text{as } s \to 0$$

which forces $E(s) = 0$, for every $s > 0$. But $v^2(y, s)$ is nonnegative and continuous in $[0, 1]$ if $s > 0$, so that it must be $v(y, s) = 0$ for every $s > 0$ or, equivalently, $U = W$.

• *Back to the original variables.* In terms of the original variables, our solution is expressed as

$$u(x, t) = u_0 + (u_1 - u_0) \frac{x}{L} - \sum_{m=1}^{\infty} (-1)^{m+1} \frac{2}{m\pi} e^{\frac{-m^2 \pi^2 D}{L^2} t} \sin \frac{m\pi}{L} x.$$

This formula confirms our initial guess about the evolution of the temperature towards the steady state. Indeed, each term of the series converges to zero exponentially as $t \to +\infty$ and it is not difficult to show[12] that

$$u(x, t) \to u_0 + (u_1 - u_0) \frac{x}{L} \qquad \text{as } t \to +\infty.$$

---

[12] The Weierstrass test works here for $t \geq t_0 > 0$.

Moreover, among the various terms of the series, the first one $(m = 1)$ decays much more slowly than the others and very soon it determines the main deviation of u from the equilibrium, *independently of the initial condition*. This leading term is the damped sinusoid

$$\frac{2}{\pi} e^{\frac{-\pi^2 D}{L^2} t} \sin \frac{\pi}{L} x.$$

In this mode there is a concentration of heat at $x = L/2$ where the temperature reaches its maximum amplitude $2 \exp(-\pi^2 Dt/L^2)/\pi$. At time $t = L^2/D$ the amplitude decays to $2 \exp(-\pi^2)/\pi \simeq 3.3 \times 10^{-5}$, about 0.005 per cent of its initial value. This simple calculation shows that to reach the steady state a time of order L^2/D is required, a fundamental fact in heat diffusion.

Not surprisingly, the scaling factor in (2.11) was exactly $\tau = L^2/D$. The dimensionless formulation is extremely useful in experimental modelling tests. To achieve reliable results, these models must reproduce the same characteristics at different scales. For instance, if our bar were an experimental model of a much bigger beam of length L_0 and diffusion coefficient D_0, to reproduce the same heat diffusion effects, we must choose material (D) and length (L) for our model bar such that

$$\frac{L^2}{D} = \frac{L_0^2}{D_0}.$$

Figure 2.2 shows the solution of the dimensionless problem (2.12), (2.15), (2.16) for $0 < t \le 1$.
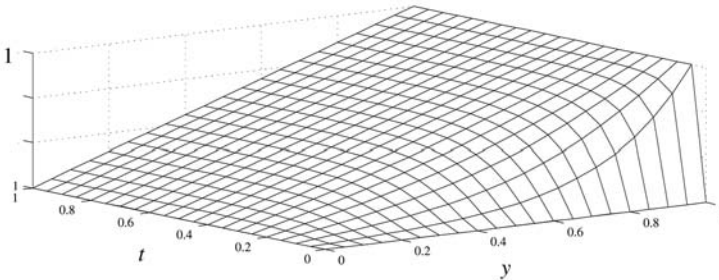
Fig. 2.2. The solution to the dimensionless problem (2.12), (2.13), (2.14)

2.1.5 Problems in dimension $n > 1$

The formulation of the well posed problems in subsection 2.1.3 can be easily generalized to any spatial dimension $n > 1$, in particular to $n = 2$ or $n = 3$. Suppose we want to determine the evolution of the temperature in a heat conducting body that occupies a bounded domain[13] $\Omega \subset \mathbb{R}^n$, during an interval of time $[0, T]$. Under the hypotheses of subsection 2.1.2, the temperature is a function $u = u(\mathbf{x}, t)$

---

[13] Recall that by *domain* we mean an *open connected set* in $\mathbb{R}^n$.

that satisfies the heat equation $u_t - D\Delta u = f$, in the *space-time cylinder*

$$Q_T = \Omega \times (0, T).$$

To select a unique solution we have to prescribe first of all the *initial distribution*

$$u(\mathbf{x}, 0) = g(\mathbf{x}) \qquad \mathbf{x} \in \overline{\Omega},$$

where $\overline{\Omega} = \Omega \cup \partial\Omega$ denotes the *closure* of Ω.

The control of the interaction of the body with the surroundings is modeled through *suitable conditions* on $\partial\Omega$. The most common ones are:

Dirichlet condition: the temperature is kept at a prescribed level on $\partial\Omega$; this amounts to assigning

$$u(\boldsymbol{\sigma}, t) = h(\boldsymbol{\sigma}, t) \qquad \boldsymbol{\sigma} \in \partial\Omega \ \text{ and } \ t \in (0, T].$$

Neumann condition: the heat flux through $\partial\Omega$ is assigned. To model this condition, we assume that the boundary $\partial\Omega$ is a smooth curve or surface, having a tangent line or plane at every point[14] with *outward* unit vector $\boldsymbol{\nu}$. From Fourier law we have

$$\mathbf{q} = \text{heat flux} = -\kappa\nabla u$$

so that the *inward heat flux* is

$$-\mathbf{q} \cdot \boldsymbol{\nu} = \kappa\nabla u \cdot \boldsymbol{\nu} = \kappa\partial_{\boldsymbol{\nu}} u.$$

Thus the Neumann condition reads

$$\partial_{\boldsymbol{\nu}} u(\boldsymbol{\sigma}, t) = h(\boldsymbol{\sigma}, t) \qquad \boldsymbol{\sigma} \in \partial\Omega \ \text{ and } \ t \in (0, T].$$

Radiation or Robin condition: the *inward* (say) heat flux through $\partial\Omega$ depends linearly on the difference[15] $U - u$:

$$-\mathbf{q} \cdot \boldsymbol{\nu} = \gamma(U - u) \qquad (\gamma > 0)$$

where U is the ambient temperature. From the Fourier law we obtain

$$\partial_{\boldsymbol{\nu}} u + \alpha u = h \qquad \text{on } \partial\Omega \times (0, T]$$

with $\alpha = \gamma/\kappa > 0$, $h = \gamma U/\kappa$.

Mixed conditions: the boundary of Ω is decomposed into various parts where different boundary conditions are prescribed. For instance, a formulation of a mixed Dirichlet-Neumann problem is obtained by writing

$$\partial\Omega = \partial_D\Omega \cup \partial_N\Omega \quad \text{with} \quad \partial_D\Omega \cap \partial_N\Omega = \varnothing$$

---

[14] We can also allow boundaries with corner points, like squares, cones, or edges, like cubes. It is enough that the set of points where the tangent plane does not exist has zero surface measure (zero length in two dimensions). Lipschitz domains have this property (see Section 1.4).

[15] Linear Newton law of cooling.

where $\partial_D \Omega$ and $\partial_N \Omega$ "reasonable" subsets of $\partial \Omega$. Typically $\partial_N \Omega = \partial \Omega \cap A$, where A is open in $\mathbb{R}^n$. In this case we say that $\partial_N \Omega$ is a *relatively open* set in $\partial \Omega$. Then we assign

$$u = h_1 \text{ on } \partial_D \Omega \times (0, T]$$
$$\partial_\nu u = h_2 \text{ on } \partial_N \Omega \times (0, T].$$

Summarizing, we have the following typical problems: *given $f = f(\mathbf{x}, t)$ and $g = g(\mathbf{x})$, determine $u = u(\mathbf{x}, t)$ such that:*

$$\begin{cases} u_t - D\Delta u = f & \text{in } Q_T \\ u(\mathbf{x}, 0) = g(\mathbf{x}) & \text{in } \overline{\Omega} \\ + \text{ boundary conditions on } \partial \Omega \times (0, T] \end{cases}$$

where the boundary conditions are:

- *Dirichlet:*
$$u = h,$$

- *Neumann:*
$$\partial_\nu u = h,$$

- *radiation or Robin:*
$$\partial_\nu u + \alpha u = h \qquad (\alpha > 0),$$

- *mixed:*
$$u = h_1 \text{ on } \partial_D \Omega, \qquad \partial_\nu u = h_2 \text{ on } \partial_N \Omega.$$
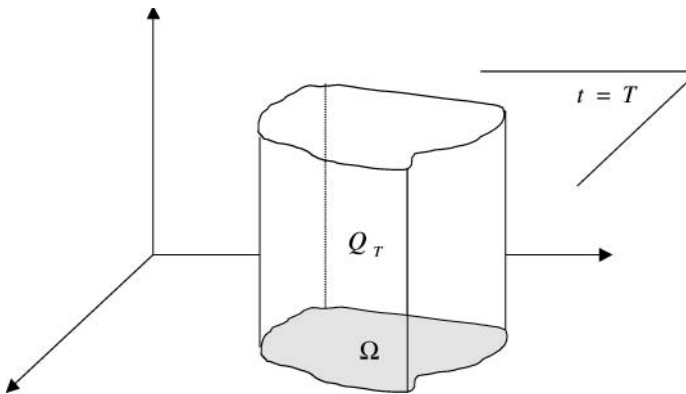
Fig. 2.3. The space-time cylinder Q_T

Also in dimension $n > 1$, the *global Cauchy problem* is important:

$$\begin{cases} u_t - D\Delta u = f & \mathbf{x} \in \mathbb{R}^n, 0 < t < T \\ u(\mathbf{x}, 0) = g(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \\ + \text{ condition as } |\mathbf{x}| \to \infty. \end{cases}$$

Remark 2.2. We again emphasize that no final condition (for $t = T, \mathbf{x} \in \Omega$) is required. The data is assigned on the *parabolic boundary* $\partial_p Q_T$ of Q_T, given by the union of the bottom points $\bar{\Omega} \times \{t = 0\}$ and the side points $\partial\Omega \times (0, T]$:

$$\partial_p Q_T = \left(\bar{\Omega} \times \{t = 0\} \right) \cup \left(\partial\Omega \times (0, T] \right).$$

2.2 Uniqueness

2.2.1 Integral method

Generalizing the energy method used in subsection 2.1.4, it is easy to show that all the problems we have formulated in the previous section have at most one solution under reasonable conditions on the data. Suppose u and v are solutions of one of those problems, sharing the same boundary conditions, and let $w = u - v$; we want to show that $w \equiv 0$. For the time being we do not worry about the precise hypotheses on u e v; we assume they are sufficiently smooth in Q_T up to $\partial_p Q_T$ and observe that w satisfies the homogeneous equation

$$w_t - D\Delta w = 0 \tag{2.35}$$

in $Q_T = \Omega \times (0, T)$, with initial condition

$$w(\mathbf{x}, 0) = 0$$

in $\overline{\Omega}$, and one of the following conditions on $\partial\Omega \times (0, T]$:

$$w = 0 \qquad \text{(Dirichlet)} \tag{2.36}$$

or

$$\partial_\nu w = 0 \qquad \text{(Neumann)} \tag{2.37}$$

or

$$\partial_\nu w + \alpha w = 0 \qquad \alpha > 0, \qquad \text{(Robin)} \tag{2.38}$$

or

$$w = 0 \text{ on } \partial_D \Omega, \quad \partial_\nu w = 0 \text{ on } \partial_N \Omega \qquad \text{(mixed)}. \tag{2.39}$$

Multiply equation (2.35) by w and integrate on Ω; we find

$$\int_\Omega w w_t \, d\mathbf{x} = D \int_\Omega w \Delta w \, d\mathbf{x}.$$

Now,

$$\int_\Omega w w_t \, d\mathbf{x} = \frac{1}{2} \frac{d}{dt} \int_\Omega w^2 d\mathbf{x} \tag{2.40}$$

and from Green's identity (1.13) with $u = v = w$,

$$\int_\Omega w \Delta w \, d\mathbf{x} = \int_{\partial\Omega} w \partial_\nu w \, d\sigma - \int_\Omega |\nabla w|^2 \, d\mathbf{x}. \tag{2.41}$$

Then, letting

$$E\left(t\right) = \int_{\Omega} w^2 d\mathbf{x},$$

(2.40) and (2.41) give

$$\frac{1}{2}E'\left(t\right) = D\int_{\partial\Omega} w\partial_{\nu}w \, d\sigma - D\int_{\Omega}\left|\nabla w\right|^2 d\mathbf{x}.$$

If Robin condition (2.38) holds,

$$\int_{\partial\Omega} w\partial_{\nu}w \, d\sigma = -\alpha\int_{\Omega} w^2 d\mathbf{x} \le 0.$$

If one of the (2.36), (2.37), (2.39) holds, then

$$\int_{\partial\Omega} w\partial_{\nu}w \, d\sigma = 0.$$

In any case it follows that
$$E'\left(t\right) \le 0$$

and therefore E is a nonincreasing function. Since

$$E\left(0\right) = \int_{\Omega} w^2\left(\mathbf{x},0\right) \, d\mathbf{x} = 0,$$

we must have $E\left(t\right) = 0$ for every $t \ge 0$ and this implies $w\left(\mathbf{x},t\right) \equiv 0$ in Ω for every $t > 0$. Thus $u = v$.

The above calculations are completely justified if Ω is a sufficiently smooth domain[16] and, for instance, we require that u and v are continuous in $\overline{Q}_T = \overline{\Omega} \times [0,T]$, together with their first and second spatial derivatives and their first order time derivatives. We denote the set of these functions by the symbol (not too appealing...)
$$C^{2,1}\left(\overline{Q}_T\right)$$

and synthesize everything in the following statement.

Theorem 2.1. *The initial Dirichlet, Neumann, Robin and mixed problems have at most one solution belonging to* $C^{2,1}\left(\overline{Q}_T\right)$.

2.2.2 Maximum principles

The fact that heat flows from higher to lower temperature regions implies that a solution of the homogeneous heat equation attains its maximum and minimum values on $\partial_p Q_T$. This result is known as the *maximum principle*. Moreover the equation reflects the time irreversibility of the phenomena that it describes, in the

---

[16] C^1 or even Lipschitz domains, for instance (see Section 1.4).

sense that the future cannot have an influence on the past (*causality principle*). In other words, the value of a solution u at time t is independent of any change of the data after t.

The following simple theorem translates these principles and holds for functions in the class $C^{2,1}(Q_T) \cap C(\overline{Q}_T)$. These functions are continuous up to the boundary of Q_T, with derivatives continuous in the interior of Q_T.

Theorem 2.2. *Let* $w \in C^{2,1}(Q_T) \cap C(\overline{Q}_T)$ *such that*

$$w_t - D\Delta w = q \leq 0 \qquad in \; Q_T. \tag{2.42}$$

Then w *attains its maximum on* $\partial_p Q_T$:

$$\max_{\overline{Q}_T} w = \max_{\partial_p Q_T} w. \tag{2.43}$$

In particular, if w *is negative on* $\partial_p Q_T$, *then is negative in all* Q_T.

Proof. We split the proof into two steps.
1. Let $\varepsilon > 0$ such that $T - \varepsilon > 0$. We prove that

$$\max_{\overline{Q}_{T-\varepsilon}} w \leq \max_{\partial_p Q_T} w + \varepsilon T. \tag{2.44}$$

Let $u = w - \varepsilon t$. Then

$$u_t - D\Delta u = q - \varepsilon < 0. \tag{2.45}$$

We claim that the maximum of u on $\overline{Q}_{T-\varepsilon}$ occurs on $\partial_p Q_{T-\varepsilon}$. Suppose not. Let $(\mathbf{x}_0, t_0)$, $\mathbf{x}_0 \in \Omega$, $0 < t_0 \leq T - \varepsilon$ be a maximum point for u on $\overline{Q}_{T-\varepsilon}$. From elementary calculus, we have

$$\Delta u(\mathbf{x}_0, t_0) \leq 0$$

and either

$$u_t(\mathbf{x}_0, t_0) = 0 \qquad if \; t_0 < T - \varepsilon$$

or

$$u_t(\mathbf{x}_0, T - \varepsilon) \geq 0.$$

In both cases

$$u_t(\mathbf{x}_0, t_0) - \Delta u(\mathbf{x}_0, t_0) \geq 0,$$

contradicting (2.45). Thus

$$\max_{\overline{Q}_{T-\varepsilon}} u \leq \max_{\partial_p Q_{T-\varepsilon}} u \leq \max_{\partial_p Q_T} w \tag{2.46}$$

since $u \leq w$. On the other hand, $w \leq u + \varepsilon T$, and therefore, from (2.46) we get

$$\max_{\overline{Q}_{T-\varepsilon}} w \leq \max_{\overline{Q}_{T-\varepsilon}} u + \varepsilon T \leq \max_{\partial_p Q_T} w + \varepsilon T$$

which is (2.44).

Step 2. Since w is continuous in $\overline{Q}_T$, we deduce that (why?)

$$\max_{\overline{Q}_{T-\varepsilon}} w \to \max_{\overline{Q}_T} w \quad \text{as } \varepsilon \to 0.$$

Hence, letting $\varepsilon \to 0$ in (2.44) we find $\max_{\overline{Q}_T} w \le \max_{\partial_p Q_T} w$ which concludes the proof. $\square$

As an immediate consequence of Theorem 2.2 (see Problem 2.4) we have that if

$$w_t - D\Delta w = 0 \quad \text{in } Q_T$$

then w attains its maximum and its minimum on $\partial_p Q_T$. In particular

$$\min_{\partial_p Q_T} w \le w\left(\mathbf{x},t\right) \le \max_{\partial_p Q_T} w \qquad \text{for every } \left(\mathbf{x},t\right) \in Q_T.$$

Moreover:

Corollary 2.1. *(Comparison and stability). Let v and w satisfy*

$$v_t - D\Delta v = f_1 \quad \text{and} \quad w_t - D\Delta w = f_2.$$

Then:

a) If $v \ge w$ on $\partial_p Q_T$ and $f_1 \ge f_2$ in Q_T then $v \ge w$ in all Q_T.

b) The following stability estimate holds

$$\max_{\overline{Q}_T} |v - w| \le \max_{\partial_p Q_T} |v - w| + T \max_{\overline{Q}_T} |f_1 - f_2|. \tag{2.47}$$

In particular the initial-Dirichlet problem has at most one solution that, moreover, depends continuously on the data.

For the proof see Problem 2.5.

Remark 2.3. Corollary 2.1 gives uniqueness for the initial-Dirichlet problem under much less restrictive hypotheses than Theorem 2.1: indeed it does not require the continuity of any derivatives of the solution up to $\partial_p Q_T$.

Inequality (2.47) is a *uniform pointwise stability* estimate, extremely useful in several applications. In fact if $v = g_1$, $w = g_2$ on $\partial_p Q_T$ and

$$\max_{\partial_p Q_T} |g_1 - g_2| \le \varepsilon \quad \text{and} \quad \max_{\overline{Q}_T} |f_1 - f_2| \le \varepsilon,$$

we deduce

$$\max_{\overline{Q}_T} |v - w| \le \varepsilon \left(1 + T\right).$$

Thus, in finite time, a small uniform distance between the data implies   small uniform distance between the corresponding solutions.

Remark 2.4. Strong maximum principle. Theorem 2.2 is a version of the so called weak maximum principle, weak because this result says nothing about the possibility that a solution achieves its maximum or minimum at an interior point as well. Actually a more precise result is known as *strong maximum principle* and states[17] that *if a solution of $u_t - D\Delta u = 0$ achieves its maximum M (minimum) at a point $(\mathbf{x}_1, t_1)$ with $x_1 \in V$, $0 < t_1 \le T$, then $u = M$ in $\bar{V} \times [0, t_1]$.*
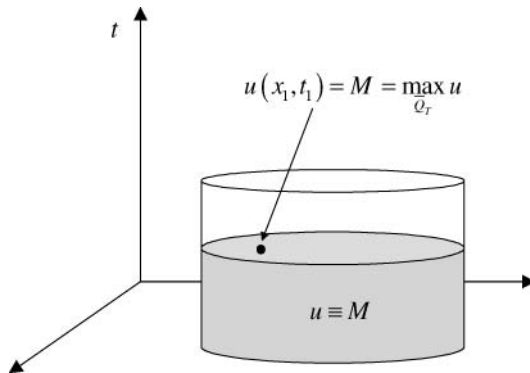
Fig. 2.4. The strong maximum principle

2.3 The Fundamental Solution

There are privileged solutions of the diffusion equation that can be used to construct many other solutions. In this section we are going to discover one of these special building blocks, the most important one.

2.3.1 Invariant transformations

The *homogeneous* diffusion equation has simple but important properties. Let $u = u(\mathbf{x}, t)$ be a solution of

$$u_t - D\Delta u = 0. \tag{2.48}$$

- *Time reversal.* The function

$$v(\mathbf{x}, t) = u(\mathbf{x}, -t),$$

obtained by the change of variable $t \longmapsto -t$, is a solution of the **adjoint** or **backward** equation.

$$v_t + D\Delta v = 0.$$

---

[17] We omit the rather long proof.

Coherently, the (2.48) is sometimes called the **forward** equation. The non-invariance of (2.48) with respect to a change of sign in time is another aspect of time irreversibility.

- *Space and time translations invariance.* For **y**,s fixed, the function

$$v(\mathbf{x},t) = u\left(\mathbf{x} - \mathbf{y}, t - s\right),$$

is still a solution of (2.48). Clearly, for $\mathbf{x}, t$ fixed the function $u\left(\mathbf{x} - \mathbf{y}, t - s\right)$ is a solution of the *backward* equation with respect to **y** and s.

- *Parabolic dilations* The transformation

$$\mathbf{x} \longmapsto a\mathbf{x}, \qquad t \longmapsto bt, \qquad u \longmapsto cu \qquad\qquad (a, b, c > 0)$$

represents a dilation (or contraction) of the graph of u. Let us check for which values of a, b, c the function

$$u^* \left(\mathbf{x},t\right) = cu\left(a\mathbf{x},bt\right)$$

is still a solution of (2.48). We have:

$$u_t^* \left(\mathbf{x},t\right) - D\Delta u^* \left(\mathbf{x},t\right) = cbu_t \left(a\mathbf{x},bt\right) - ca^2 D\Delta u \left(a\mathbf{x},bt\right)$$

and so u^* is a solution of (2.48) if

$$b = a^2. \qquad\qquad (2.49)$$

The relation (2.49) suggests the name of *parabolic dilation* for the transformation

$$\mathbf{x} \longmapsto a\mathbf{x} \qquad t \longmapsto a^2 t \qquad\qquad (a, b > 0).$$

Under this transformation the expressions

$$\frac{|\mathbf{x}|^2}{Dt} \qquad \text{or} \qquad \frac{\mathbf{x}}{\sqrt{Dt}}$$

are left unchanged. Moreover, we already observed that they are *dimensionless groups*. Thus it is not surprising that these combinations of the independent variables occur frequently in the study of diffusion phenomena.

- *Dilations and conservation of mass* (or *energy*). Let $u = u\left(\mathbf{x}, t\right)$ be a solution of (2.48) in the half-space $\mathbb{R}^n \times \left(0, +\infty\right)$. Then we just checked that the function

$$u^* \left(\mathbf{x},t\right) = cu\left(a\mathbf{x},a^2 t\right) \qquad\qquad (a > 0)$$

is also a solution in the same set. Suppose u satisfies the condition

$$\int_{\mathbb{R}^n} u\left(\mathbf{x}, t\right) d\mathbf{x} = q \qquad \text{for every } t > 0. \qquad\qquad (2.50)$$

If, for instance, u represents the concentration of a substance (density of mass), equation (2.50) states that the total mass is q at every time t. If u is a temperature, (2.50) says that the total internal energy is constant $(= q\rho c_v)$. We ask for which a, c the solution u^* still satisfies (2.50). We have

$$\int_{\mathbb{R}^n} u^* (\mathbf{x}, t) \, d\mathbf{x} = c \int_{\mathbb{R}^n} u \left(a\mathbf{x}, a^2 t \right) d\mathbf{x}.$$

Letting $\mathbf{y} = a\mathbf{x}$, so that $d\mathbf{y} = a^n d\mathbf{x}$, we find

$$\int_{\mathbb{R}^n} u^* (\mathbf{x}, t) \, d\mathbf{x} = c a^{-n} \int_{\mathbb{R}^n} u \left(\mathbf{y}, a^2 t \right) d\mathbf{y} = c a^{-n}$$

and for (2.50) to be satisfied we must have:

$$c = q a^n.$$

In conclusion, if $u = u(\mathbf{x}, t)$ is a solution of (2.48) in the half-space $\mathbb{R}^n \times (0, +\infty)$ satisfying (2.50), the same is true for

$$u^* (\mathbf{x}, t) = q a^n u \left(a\mathbf{x}, a^2 t \right). \tag{2.51}$$

2.3.2 Fundamental solution ($n = 1$)

We are now in position to construct our special solution, starting with dimension $n = 1$. To help intuition, think for instance of our solution as the concentration of a substance of total mass q and suppose we want to keep the total mass equal to q at any time.

We have seen that the combination of variables $x/\sqrt{Dt}$ is not only invariant with respect to parabolic dilations but also dimensionless. It is then natural to check if there are solutions of (2.48) involving such dimensionless group. Since $\sqrt{Dt}$ has the dimension of a length, the quantity $q/\sqrt{Dt}$ is a typical order of magnitude for the concentration, so that it makes sense to look for solutions of the form

$$u^* (x, t) = \frac{q}{\sqrt{Dt}} U \left(\frac{x}{\sqrt{Dt}} \right) \tag{2.52}$$

where U is a (dimensionless) function of a single variable.

Here is the main question: is it possible to determine $U = U(\xi)$ such that u^* is a solution of (2.48)? Solutions of the form (2.52) are called *similarity solutions*[18].

---

[18] A solution of a particular evolution problem is a *similarity* or *self-similar* solution if its spatial configuration (graph) remains similar to itself at all times during the evolution. In one space dimension, *self-similar* solutions have the general form

$$u(x, t) = a(t) F(x/b(t))$$

where, preferably, u/a and x/b are dimensionless quantity.

Moreover, since we are interpreting u^* as a concentration, we require $U \geq 0$ and the total mass condition yields

$$1 = \frac{1}{\sqrt{Dt}} \int_{\mathbb{R}} U\left(\frac{x}{\sqrt{Dt}}\right) dx \underset{\xi = x/\sqrt{Dt}}{=} \int_{\mathbb{R}} U(\xi)\, d\xi$$

so that we require that

$$\int_{\mathbb{R}} U(\xi)\, d\xi = 1. \tag{2.53}$$

Let us check if u^* is a solution to (2.48). We have

$$u_t^* = \frac{q}{\sqrt{D}}\left[-\frac{1}{2}t^{-\frac{3}{2}}U(\xi) - \frac{1}{2\sqrt{D}}xt^{-2}U'(\xi)\right]$$

$$= -\frac{q}{2t\sqrt{Dt}}\left[U(\xi) + \xi U'(\xi)\right]$$

$$u_{xx}^* = \frac{q}{(Dt)^{3/2}}U''(\xi),$$

hence

$$u_t^* - Du_{xx}^* = -\frac{q}{t\sqrt{Dt}}\left\{U''(\xi) + \frac{1}{2}\xi U'(\xi) + \frac{1}{2}U(\xi)\right\}.$$

We see that for u^* to be a solution of (2.48), U must be a solution in $\mathbb{R}$ of the ordinary differential equation

$$U''(\xi) + \frac{1}{2}\xi U'(\xi) + \frac{1}{2}U(\xi) = 0. \tag{2.54}$$

Since $U \geq 0$, (2.53) implies[19]:

$$U(-\infty) = U(+\infty) = 0.$$

On the other hand, (2.54) is invariant with respect to the change of variables

$$\xi \mapsto -\xi$$

and therefore we look for *even solutions*: $U(-\xi) = U(\xi)$. Then we can restrict ourselves to $\xi \geq 0$, asking

$$U'(0) = 0 \text{ and } U(+\infty) = 0. \tag{2.55}$$

---

[19] Rigorously, the precise conditions are:

$$\liminf_{x \to \pm\infty} U(x) = 0.$$

To solve (2.54) observe that it can be written in the form

$$\frac{d}{d\xi}\left\{U'\left(\xi\right)+\frac{1}{2}\xi U\left(\xi\right)\right\}=0$$

that yields

$$U'\left(\xi\right)+\frac{1}{2}\xi U\left(\xi\right)=C \qquad (C\in\mathbb{R}). \qquad (2.56)$$

Letting $\xi=0$ in (2.56) and recalling (2.55) we deduce that $C=0$ and therefore

$$U'\left(\xi\right)+\frac{1}{2}\xi U\left(\xi\right)=0. \qquad (2.57)$$

The general integral of (2.57) is

$$U\left(\xi\right)=c_0 e^{-\frac{\xi^2}{4}} \qquad (c_0\in\mathbb{R}).$$

This function is even, positive, integrable and vanishes at infinity. It only remains to choose c_0 in order to ensure (2.53). Since[20]

$$\int_{\mathbb{R}} e^{-\frac{\xi^2}{4}}d\xi \underset{\xi=2z}{=} 2\int_{\mathbb{R}} e^{-z^2}dz = 2\sqrt{\pi}$$

the choice is $c_0 = (4\pi)^{-1/2}$.

Going back to the original variables, we have found the following solution of (2.48)

$$u^*\left(x,t\right)=\frac{q}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}, \qquad x\in\mathbb{R}, t>0$$

positive, even in x, and such that

$$\int_{\mathbb{R}} u^*\left(x,t\right)dx = q \qquad \text{for every } t>0. \qquad (2.58)$$

The choice $q=1$ gives a family of *Gaussians*, parametrized with time, and it is natural to think of a *normal probability density*.

Definition 2.1. *The function*

$$\Gamma_D\left(x,t\right)=\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}, \qquad x\in\mathbb{R}, t>0 \qquad (2.59)$$

*is called the **fundamental solution** of equation (2.48).*

---

[20] Recall that

$$\int_{\mathbb{R}} e^{-z^2} = \sqrt{\pi}.$$
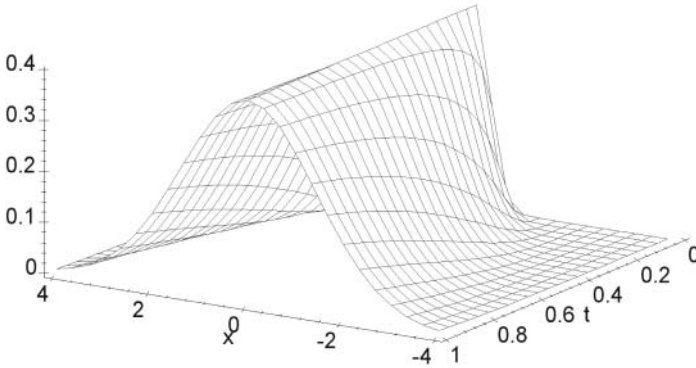
Fig. 2.5. The fundamental solution Γ_1 for $-4 < x < 4$, $0 < t < 1$

2.3.3 The Dirac distribution

It is worthwhile to examine the behavior of the fundamental solution. For every fixed $x \neq 0$,

$$\lim_{t \to 0^+} \Gamma_D\left(x, t\right) = \lim_{t \to 0^+} \frac{1}{\sqrt{4\pi D t}} e^{-\frac{x^2}{4Dt}} = 0 \tag{2.60}$$

while

$$\lim_{t \to 0^+} \Gamma_D\left(0, t\right) = \lim_{t \to 0^+} \frac{1}{\sqrt{4\pi D t}} = +\infty. \tag{2.61}$$

If we interpret Γ_D as a probability density, equations (2.60), (2.61) and (2.58) imply that when $t \to 0^+$ the fundamental solution tends to concentrate mass around the origin; eventually, the whole probability mass is concentrated at $x = 0$ (see Fig. 2.5).

The limiting density distribution can be mathematically modeled by the so called *Dirac distribution* (or *measure*) *at the origin,* denoted by the symbol δ_0 or simply by δ. The Dirac distribution is not a function in the usual sense of Analysis; if it were, it should have the following properties:

- $\delta\left(0\right) = \infty$, $\delta\left(x\right) = 0$ for $x \neq 0$
- $\int_{\mathbb{R}} \delta\left(x\right) dx = 1$,

clearly incompatible with any concept of classical function or integral. A rigorous definition of the Dirac measure requires the theory of *generalized functions* or *distributions of L. Schwartz,* that we will consider in Chapter 7. Here we restrict ourselves to some heuristic considerations.

Let

$$\mathcal{H}\left(x\right) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$
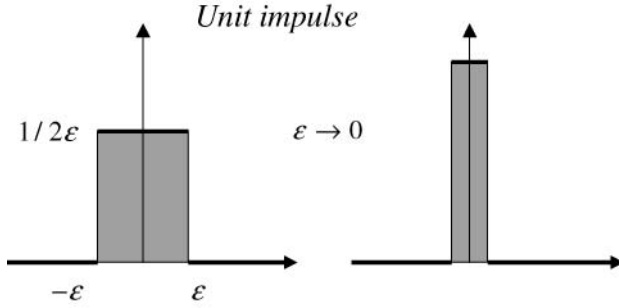
Fig. 2.6. Approximation of the Dirac measure

be the characteristic function of the interval $[0, \infty)$, known as the Heaviside function. Observe that

$$\frac{\mathcal{H}(x + \varepsilon) - \mathcal{H}(x - \varepsilon)}{2\varepsilon} = \begin{cases} \frac{1}{2\varepsilon} & \text{if } -\varepsilon \leq x < \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{2.62}$$

Denote by $I_\varepsilon(x)$ the quotient (2.62); the following properties hold:

i) For every $\varepsilon > 0$,

$$\int_{\mathbb{R}} I_\varepsilon(x)\, dx = \frac{1}{2\varepsilon} \times 2\varepsilon = 1.$$

We can interpret I_ε as a *unit impulse of extent* 2ε (Fig. 2.6).

ii)

$$\lim_{\varepsilon \downarrow 0} I_\varepsilon(x) = \begin{cases} 0 & \text{if } x \neq 0 \\ \infty & \text{if } x = 0. \end{cases}$$

iii) If $\varphi = \varphi(x)$ is a smooth function, vanishing outside a bounded interval, (a *test function*), we have

$$\int_{\mathbb{R}} I_\varepsilon(x)\,\varphi(x)\, dx = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} \varphi(x)\, dx \xrightarrow[\varepsilon \to 0]{} \varphi(0).$$

Properties i) e ii) say that I_ε tends to a mathematical object that has precisely the formal features of the Dirac distribution at the origin. In particular iii) suggests how to identify this object, that is *through its action on test functions*.

Definition 2.2. *We call Dirac measure at the origin the generalized function, denoted by δ, that acts on a test function φ as follows:*

$$\delta[\varphi] = \varphi(0). \tag{2.63}$$

Equation (2.63) is often written in the form $\langle \delta, \varphi \rangle = \varphi(0)$ or even

$$\int \delta(x)\,\varphi(x)\, dx = \varphi(0)$$

where the integral symbol is purely formal. Observe that property ii) shows that

$$\mathcal{H}' = \delta$$

whose meaning is given in the following computations, where an integration by parts is used and φ is a test function:

$$\int_{\mathbb{R}} \varphi \, d\mathcal{H} = -\int_{\mathbb{R}} \mathcal{H}\varphi' = -\int_0^\infty \varphi' = \varphi(0), \qquad (2.64)$$

since φ vanishes for large[21] x .

With the notion of Dirac measure at hand, we can say that Γ_D satisfies the initial conditions

$$\Gamma_D(x, 0) = \delta.$$

If the unit mass is concentrated at a point $y \neq 0$, we denote by δ_y or $\delta(x - y)$ the Dirac measure at y, defined through the formula

$$\int \delta(x - y)\,\varphi(x)\,dx = \varphi(y).$$

Then, by translation invariance, the fundamental solution $\Gamma_D(x - y, t)$ is a solution of the diffusion equation, that satisfies the initial condition

$$\Gamma_D(x - y, 0) = \delta(x - y).$$

Indeed it is the unique solution satisfying the total mass condition (2.58) with $q = 1$.

As any solution u of (2.48) has several interpretations (concentration of a substance, probability density, temperature in a bar) so the fundamental solution can have several meanings.

We can think of it as a **unit source solution**: $\Gamma_D(x, t)$ gives the concentration at the point x at time t, generated by the diffusion of **a unit mass initially** $(t = 0)$ **concentrated at the origin**. From another point of view, if we imagine a unit mass composed of a large number N of particles, $\Gamma_D(x, t)\,dx$ gives the probability that a single particle is placed between x and $x + dx$ at time t or equivalently, the percentage of particles inside the interval $(x, x + dx)$ at time t.

Initially Γ_D is zero outside the origin. As soon as $t > 0$, Γ_D becomes positive everywhere: this amounts to saying that the unit mass diffuses instantaneously all over the $x-$axis and therefore with *infinite speed of propagation*. This could be a problem in using (2.48) as a realistic model, although (see Fig. 2.5) for $t > 0$, small, Γ_D is practically zero outside an interval centered at the origin of length $4D$.

---

[21] The first integral in (2.64) is a Riemann-Stieltjes integral, that formally can be written as

$$\int \varphi(x)\,\mathcal{H}'(x)\,dx$$

and interpreted as *the action of the generalized function $\mathcal{H}'$ on the test function φ.*

2.3.4 Fundamental solution ($n > 1$)

In space dimension greater than 1, we can more or less repeat the same arguments. We look for positive, radial, self-similar solutions u^* to (2.48), with total mass equal to q at every time, that is

$$\int_{\mathbb{R}^n} u^*(\mathbf{x},t)\, d\mathbf{x} = q \qquad \text{for every } t > 0. \tag{2.65}$$

Since $q/(Dt)^{n/2}$ is a concentration per unit volume, we set

$$u^*(\mathbf{x}, t) = \frac{q}{(Dt)^{n/2}} U(\xi), \qquad \xi = |\mathbf{x}|/\sqrt{Dt}.$$

We have, recalling the expression of the Laplace operator for radial functions (see Appendix C),

$$u_t^* = -\frac{1}{2t(Dt)^{n/2}} \left[nU(\xi) + \xi U'(\xi) \right]$$

$$\Delta u^* = \frac{1}{(Dt)^{1+n/2}} \left\{ U''(\xi) + \frac{n-1}{\xi} U'(\xi) \right\}.$$

Therefore, for u^* to be a solution of (2.48), U must be a nonnegative solution in $(0, +\infty)$ of the ordinary differential equation

$$\xi U''(\xi) + (n-1)U'(\xi) + \frac{\xi^2}{2}U'(\xi) + \frac{n}{2}\xi U(\xi) = 0. \tag{2.66}$$

Multiplying by ξ^{n-2}, we can write (2.66) in the form

$$(\xi^{n-1}U')' + \frac{1}{2}(\xi^n U)' = 0$$

that gives

$$\xi^{n-1}U' + \frac{1}{2}\xi^n U = C \qquad (C \in \mathbb{R}). \tag{2.67}$$

Assuming that $\lim_{\xi \to 0+}$ of U and U' are finite, letting $\xi \to 0^+$ into (2.67), we deduce $C = 0$ and therefore

$$U' + \frac{1}{2}\xi U = 0.$$

Thus we obtain the family of solutions

$$U(\xi) = c_0 e^{-\frac{\xi^2}{4}}.$$

The total mass condition requires

$$1 = \frac{1}{(Dt)^{n/2}} \int_{\mathbb{R}^n} U\left(\frac{|\mathbf{x}|}{\sqrt{Dt}}\right) d\mathbf{x} = \frac{c_0}{(Dt)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{|\mathbf{x}|^2}{4Dt}\right) d\mathbf{x}$$

$$\underset{\mathbf{y} = \mathbf{x}/\sqrt{Dt}}{=} c_0 \int_{\mathbb{R}^n} e^{-|\mathbf{y}|^2} d\mathbf{y} = c_0 \left(\int_{\mathbb{R}} e^{-z^2} dz \right)^n = c_0 \left(4\pi \right)^{n/2}$$

and therefore $c_0 = (4\pi)^{-n/2}$. Thus, we have obtained solutions of the form

$$u^* \left(\mathbf{x}, t \right) = \frac{q}{\left(4\pi Dt \right)^{n/2}} \exp \left(-\frac{|\mathbf{x}|^2}{4Dt} \right), \qquad (t > 0).$$

Once more, the choice $q = 1$ is special.

Definition 2.3. *The function*

$$\Gamma_D \left(\mathbf{x}, t \right) = \frac{1}{\left(4\pi Dt \right)^{n/2}} \exp \left(-\frac{|\mathbf{x}|^2}{4Dt} \right) \qquad (t > 0)$$

is called the **fundamental solution** *of the diffusion equation (2.48).*

The remarks after Definition 2.2 can be easily generalized to the multidimensional case. In particular, it is possible to define the $n-$ dimensional Dirac measure at a point $\mathbf{y}$ through the formula[22]

$$\int \delta \left(\mathbf{x} - \mathbf{y} \right) \varphi \left(\mathbf{x} \right) dx = \varphi \left(\mathbf{y} \right) \tag{2.68}$$

that expresses the action on the *test function* φ, smooth in $\mathbb{R}^n$ and vanishing outside a *compact* set. For fixed $\mathbf{y}$, the fundamental solution $\Gamma_D \left(\mathbf{x} - \mathbf{y}, t \right)$ is the unique solution of the global Cauchy problem

$$\begin{cases} u_t - D\Delta u = 0 & \mathbf{x} \in \mathbb{R}^n, t > 0 \\ u \left(\mathbf{x}, 0 \right) = \delta \left(\mathbf{x} - \mathbf{y} \right) & \mathbf{x} \in \mathbb{R}^n \end{cases}$$

that satisfies (2.65) with $q = 1$.

2.4 Symmetric Random Walk $(n = 1)$

In this section we start exploring the connection between probabilistic and deterministic models, in dimension $n = 1$. The main purpose is to construct a Brownian motion, which is a **continuous model** (in both space and time), as a limit of a simple stochastic process, called *random walk,* which is instead a **discrete model** (in both space and time). During the realization of the limiting procedure we shall see how the diffusion equation can be approximated by a *difference equation.* Moreover, this new perspective will better clarify the nature of the diffusion coefficient.

---

[22] As in dimension $n = 1$, in (2.68) the integral has a symbolic meaning only.

2.4.1 Preliminary computations

Consider a unit mass particle[23] that moves randomly along the x axis, according to the following rules: fix

$$\cdot\ h > 0,\ \text{space step}$$
$$\cdot\ \tau > 0,\ \text{time step.}$$

1. During an interval of time τ, the particle takes one step of h unit length, starting from $x = 0$.
2. The particle moves to the left or to the right with probability $p = \frac{1}{2}$, independently of the previous step (Fig. 2.7).

At time $t = N\tau$, after N steps, the particle will be at a point $x = mh$, where $N \geq 0$ and m are integers, $-N \leq m \leq N$.

Our task is: *Compute the probability $p(x,t)$ of finding the particle at x at time t.*
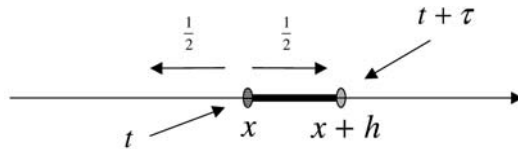
Fig. 2.7. Symmetric random walk

Random walks can be found in a wide variety of situations. To give an example, think of a gambling game in which a *fair coin* is thrown. If heads comes out, the particle moves to the right and the player gains 1 *dollar*; if tails comes out it moves to the left and the player loses 1 *dollar*: $p(x,t)$ is the probability to gain m dollars after N throws.

● *Computation of $p(x,t)$.*

Let $x = mh$ be the position of the particle after N steps. To reach x, the particle takes some number of steps to the right, say k, and $N - k$ steps to the left. Clearly, $0 \leq k \leq N$ and

$$m = k - (N - k) = 2k - N \tag{2.69}$$

so that N and m are both even or both odd integers and

$$k = \frac{1}{2}(N + m).$$

Thus, $p(x,t) = p_k$ where

$$p_k = \frac{\text{number of walks with } k \text{ steps to the right after } N \text{ steps}}{\text{number of possible walks after } N \text{ steps}}. \tag{2.70}$$

---

[23] One can also think of a large number of particles of total mass one.

Now, the number of possible walks with k steps to the right and $N - k$ to the left is given by the binomial coefficient[24]

$$C_{N,k} = \binom{N}{k} = \frac{N!}{k!\,(N-k)!}.$$

On the other hand, the number of possible walks after N steps is 2^N (why?); hence, from (2.70):

$$p_k = \frac{C_{N,k}}{2^N} \qquad x = mh,\ t = N\tau,\ k = \frac{1}{2}\,(N + m)\,. \tag{2.71}$$

- *Mean displacement and standard deviation of* x.

Our ultimate goal is to let h and τ go to zero in order to get a continuous walk, which incorporates the main features of the discrete random walk. This is a delicate point, since, if we want to obtain eventually a continuous faithful copy of the random walk, we need to isolate some quantitative parameters able to capture the essential features of the walk and maintain them unchanged. In our case there are two key parameters[25]:

(a) the **mean displacement** of x after N steps $= \langle x \rangle = \langle m \rangle\,h$

(b) the **second moment** of x after N steps $= \langle x^2 \rangle = \langle m^2 \rangle\,h^2$.

The quantity $\sqrt{\langle x^2 \rangle} = \sqrt{\langle m^2 \rangle}\,h$ is essentially the average distance from the origin after N steps.

First observe that, from (2.69), we have

$$\langle m \rangle = 2\,\langle k \rangle - N \tag{2.72}$$

and

$$\langle m^2 \rangle = 4\,\langle k^2 \rangle - 4\,\langle k \rangle\,N + N^2. \tag{2.73}$$

---

[24] The set of walks with k steps to the right and $N - k$ to the left is in one to one correspondence with the set of *sequences of* N binary digits, containing k "1" and $N - k$ "0", where 1 means *right* and 0 means *left*. There are exactly $C_{N,k}$ of these sequences.

[25] If a random variable x takes N possible outcomes $x_1,..., x_N$ with probability $p_1, ..., p_N$, its *moments of (integer) order* $q \geq 1$ are given by

$$E\,(x^q) = \langle x^q \rangle = \sum_{j=1}^{N} x_j^q p_j\,.$$

The first moment ($q = 1$) is the *mean or expected value of* x, while

$$var\,(x) = \langle x^2 \rangle - \langle x \rangle^2$$

is the *variance of* x. The square root of the variance is called *standard deviation*.

Thus, to compute $\langle m \rangle$ and $\langle m^2 \rangle$ it is enough to compute $\langle k \rangle$ and $\langle k^2 \rangle$. We have, by definition and from (2.71),

$$\langle k \rangle = \sum_{k=1}^{N} k p_k = \frac{1}{2^N} \sum_{k=1}^{N} k C_{N,k}, \qquad \langle k^2 \rangle = \sum_{k=1}^{N} k^2 p_k = \frac{1}{2^N} \sum_{k=1}^{N} k^2 C_{N,k}. \quad (2.74)$$

Although it is possible to make the calculations directly from (2.74), it is easier to use the *probability generating* function, defined by

$$G\left(s\right) = \sum_{k=0}^{N} p_k s^k = \frac{1}{2^N} \sum_{k=0}^{N} C_{N,k} s^k.$$

The function G contains in compact form all the information on the moments of k and works for all the discrete random variables taking integer values. In particular, we have

$$G'\left(s\right) = \frac{1}{2^N} \sum_{k=1}^{N} k C_{N,k} s^{k-1}, \qquad G''\left(s\right) = \frac{1}{2^N} \sum_{k=2}^{N} k\left(k-1\right) C_{N,k} s^{k-2}. \quad (2.75)$$

Letting $s = 1$ and using (2.74), we get

$$G'\left(1\right) = \frac{1}{2^N} \sum_{k=1}^{N} k C_{N,k} = \langle k \rangle \quad (2.76)$$

and

$$G''\left(1\right) = \frac{1}{2^N} \sum_{k=2}^{N} k\left(k-1\right) C_{N,k} = \langle k\left(k-1\right)\rangle = \langle k^2 \rangle - \langle k \rangle. \quad (2.77)$$

On the other hand, letting $a = 1$ and $b = s$ in the elementary formula

$$\left(a+b\right)^N = \sum_{k=0}^{N} C_{N,k} a^{N-k} b^k,$$

we deduce

$$G\left(s\right) = \frac{1}{2^N}\left(1+s\right)^N$$

and therefore

$$G'\left(1\right) = \frac{N}{2} \quad \text{and} \quad G''\left(1\right) = \frac{N\left(N-1\right)}{4}. \quad (2.78)$$

From (2.78), (2.76) and (2.77) we easily find

$$\langle k^2 \rangle = \frac{N}{2} \quad \text{and} \quad \langle k^2 \rangle = \frac{N\left(N+1\right)}{4}.$$

Finally, since $m = 2k - N$, we have

$$\langle m \rangle = 2\langle k \rangle - N = 2\frac{N}{2} - N = 0$$

and also $\langle x \rangle = \langle m \rangle h = 0$, which is not surprising, given the symmetry of the walk. Furthermore

$$\langle m^2 \rangle = 4 \langle k^2 \rangle - 4N \langle k \rangle + N^2 = N^2 + N - 2N^2 + N^2 = N$$

from which

$$\sqrt{\langle x^2 \rangle} = \sqrt{N} h \tag{2.79}$$

which is the *standard deviation of* x, since $\langle x \rangle = 0$. Formula (2.79) contains a key information: at time $N\tau$, the distance from the origin is of order $\sqrt{N} h$, that is **the order of the time scale is the square of the space scale**. In other words, if we want to leave the standard deviation unchanged in the limit process, we must rescale the time as the square of the space, that is we must use a *space-time parabolic dilation*!

But let us proceed step by step. The next one is to deduce a *difference equation* for the transition probability $p = p(x, t)$. It is on this equation that we will carry out the limit procedure.

2.4.2 The limit transition probability

The particle motion has no memory since each move is independent from the previous one. If the particle location at time $t + \tau$ is x, this means that at time t its location was at $x - h$ or at $x + h$, with equal probability. The total probability formula then gives

$$p(x, t + \tau) = \frac{1}{2} p(x - h, t) + \frac{1}{2} p(x + h, t) \tag{2.80}$$

with the initial conditions

$$p(0, 0) = 1 \qquad \text{and} \qquad p(x, 0) = 0 \quad \text{if } x \neq 0.$$

Keeping fixed x and t, let us examine what happens when $h \to 0, \tau \to 0$. It is convenient to think of p as a smooth function, defined in the whole half plane $\mathbb{R} \times (0, +\infty)$ and not only at the discrete set of points $(mh, N\tau)$. In addition, by passing to the limit, we will find a continuous probability distribution so that $p(x, t)$, being the probability to find the particle at (x, t), should be zero. If we interpret p as a *probability density*, this inconvenience disappears. Using Taylor's formula we can write[26]

$$p(x, t + \tau) = p(x, t) + p_t(x, t)\tau + o(\tau),$$

$$p(x \pm h, t) = p(x, t) \pm p_x(x, t) h + \frac{1}{2} p_{xx}(x, t) h^2 + o(h^2).$$

---

[26] The symbol $o(z)$, ("little o of z") denotes a quantity of lower order with respect to z; precisely

$$\frac{o(z)}{z} \to 0 \qquad \text{when } z \to 0.$$

Substituting into (2.80), after some simplifications, we find

$$p_t \tau + o\left(\tau\right) = \frac{1}{2}p_{xx}h^2 + o\left(h^2\right).$$

Dividing by τ,

$$p_t + o\left(1\right) = \frac{1}{2}\frac{h^2}{\tau}p_{xx} + o\left(\frac{h^2}{\tau}\right). \tag{2.81}$$

This is the crucial point; in the last equation we meet again the combination $\frac{h^2}{\tau}$!!

If we want to obtain something non trivial when $h, \tau \to 0$, **we must require that h^2/τ has a finite and positive limit**; the simplest choice is to keep

$$\frac{h^2}{\tau} = 2D \tag{2.82}$$

for some number $D > 0$ (the number 2 is there for aesthetic reasons only).

Passing to the limit in (2.81), we get for p the equation

$$p_t = Dp_{xx} \tag{2.83}$$

while the initial condition becomes

$$\lim_{t \to 0^+} p\left(x, t\right) = \delta. \tag{2.84}$$

We have already seen that the unique solution of (2.83), (2.84) is

$$p\left(x, t\right) = \Gamma_D\left(x, t\right)$$

since

$$\int_{\mathbb{R}} p\left(x, t\right) dx = 1.$$

Thus, the constant D in (2.82) is precisely the *diffusion coefficient*. Recalling that

$$h^2 = \frac{\langle x^2 \rangle}{N}, \qquad \tau = \frac{t}{N}$$

we have

$$\frac{h^2}{\tau} = \frac{\langle x^2 \rangle}{t} = 2D$$

that means: *in unit time, the particle diffuses an average distance of $\sqrt{2D}$.* It is worthwhile to recall that the dimensions of D are

$$[D] = [length]^2 \times [time]^{-1}$$

and that the combination x^2/Dt is *dimensionless*, not only invariant by parabolic dilations. Also, from (2.82) we deduce

$$\frac{h}{\tau} = \frac{2D}{h} \to +\infty. \tag{2.85}$$

This shows that the average speed h/τ of the particle at each step becomes unbounded. Therefore, the fact that the particle diffuses in unit time to a finite average distance is purely due to the rapid fluctuations of its motion.

2.4.3 From random walk to Brownian motion

What happened in the limit to the random walk? What kind of motion did it become? We can answer using some more tools from probability theory. Let $x_j = x(j\tau)$ the position of our particle after j steps and let, for $j \geq 1$,

$$h\xi_j = x_j - x_{j-1}.$$

The ξ_j are *independent, identically distributed random variables*: each one takes on value 1 or -1 with probability $\frac{1}{2}$. They have expectation $\langle \xi_j \rangle = 0$ and variance $\langle \xi_j^2 \rangle = 1$. The displacement of the particle after N steps is

$$x_N = h \sum_{j=1}^{N} \xi_j.$$

If we choose

$$h = \sqrt{\frac{2Dt}{N}},$$

that is $\frac{h^2}{\tau} = 2D$, and let $N \to \infty$, the *Central Limit Theorem* assures that x_N converges in law[27] to a random variable $X = X(t)$, normally distributed with mean 0 and variance $2Dt$, whose density is $\Gamma_D(x, t)$.

The random walk has become a continuous walk; if $D = 1/2$, it is called (1-dimensional) **Brownian motion** or **Wiener process**, that we will characterize later through its essential features.

Usually the symbol $B = B(t)$ is used to indicate the random position of a Brownian particle. The family of random variables $B(t)$ (where t plays the role of a parameter) is defined on a common probability space (Ω, F, P), where Ω is the set of elementary events, F a $\sigma-$algebra in Ω of measurable events, and P a suitable probability measure[28] in F; therefore the right notation should be $B(t, \omega)$, with $\omega \in \Omega$, but the dependence on ω is usually omitted and understood (for simplicity or laziness).

The family of random variables $B(t, \omega)$, with time t as a real parameter, **is a continuous stochastic process**. Keeping $\omega \in \Omega$ fixed, we get the real function

$$t \longmapsto B(t, \omega)$$

whose graph describes a Brownian path (see Fig. 2.8).

---

[27] That is, if $N \to +\infty$,

$$\text{Prob}\{a < x_N < b\} \to \int_a^b \Gamma_D(x, t) \ dx.$$

[28] See Appendix B.
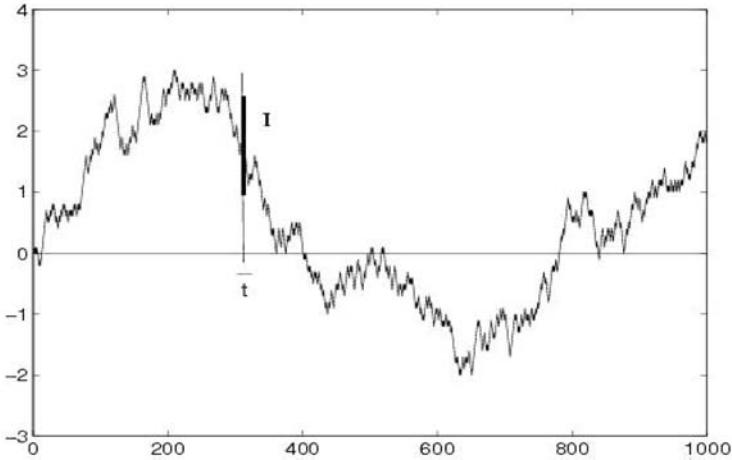
Fig. 2.8. A Brownian path

Keeping t fixed, we get the random variable

$$\omega \longmapsto B(t, \omega).$$

Without caring too much of what really is Ω, it is important to be able to compute the probability

$$P\{B(t) \in I\}$$

where $I \subseteq \mathbb{R}$ is a reasonable subset of $\mathbb{R}$, (*a so called Borel set*[29]). Figure 2.8 shows the meaning of this computation: fixing t amounts to fixing a vertical straight line, say $t = \bar{t}$. Let I be a subset of this line; in the picture I is an interval. $P\{B(t) \in I\}$ is the probability that the particle hits I at time t.

The main properties of Brownian motion are listed below. To be minimalistic we could synthesize everything in the formula[30]

$$dB \sim \sqrt{dt}N(0,1) = N(0, dt) \tag{2.86}$$

where $N(0,1)$ is a normal random variable, with zero mean and variance equal to one.

• *Path continuity.* With probability 1, the possible paths of a Brownian particle are continuous functions

$$t \longmapsto B(t), \quad t \geq 0.$$

Since from (2.85) the instantaneous speed of the particle is infinite, their graphs are nowhere differentiable!

---

[29] An interval or a set obtained by countable unions and intersections of intervals, for instance. See Appendix B.

[30] If X is a random variable, we write $X \sim N(\mu, \sigma^2)$ if X has normal distribution with mean μ and variance σ^2.

• *Gaussian law for increments.* We can allow the particle to start from a point $x \neq 0$, by considering the process

$$B^x(t) = x + B(t).$$

With every point x is associated a probability P^x, with the following properties (if $x = 0$, $P^0 = P$).

a) $P^x\{B^x(0) = x\} = P\{B(0) = 0\} = 1$.

b) For every $s \geq 0, t \geq 0$, the increment

$$B^x(t + s) - B^x(s) = B(t + s) - B(s)$$

has normal law **with zero mean and variance** t, whose density is

$$\Gamma(x, t) \equiv \Gamma_{\frac{1}{2}}(x, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

Moreover it is independent of any event occurred at a time $\leq s$. For instance, the two events

$$\{B^x(t_2) - B^x(t_1) \in I_2\} \qquad \{B^x(t_1) - B^x(t_0) \in I_1\}$$

$t_0 < t_1 < t_2$, are independent.

• *Transition probability.* For each Borel set $I \subseteq \mathbb{R}$, a *transition function*

$$P(x, t, I) = P^x\{B^x(t) \in I\}$$

is defined, assigning the probability that the particle, initially at x, belongs to I at time t. We can write:

$$P(x, t, I) = P\{B(t) \in I - x\} = \int_{I-x} \Gamma(y, t)\, dy = \int_I \Gamma(y - x, t)\, dy.$$

• *Invariance.* The motion is invariant with respect to translations.

• *Markov and strong Markov properties.* Let μ be a probability measure[31] on $\mathbb{R}$. If the initial position of the particle is random with a probability distribution μ, we can consider a *Brownian motion with initial distribution* μ, and for it we use the symbol B^μ. With this motion is associated a probability distribution P^μ such that, for every Borel set $F \subseteq \mathbb{R}$,

$$P^\mu\{B^\mu(0) \in F\} = \mu(F).$$

The probability that the particle belongs to I at time t can be computed through the formula

$$P^\mu\{B^\mu(t) \in I\} = \int_{\mathbb{R}} P^x\{B^x(t) \in I\}\, d\mu(x)$$

$$= \int_{\mathbb{R}} P(x, t, I)\, d\mu(x).$$

---

[31] See Appendix B for the definition of a probability measure μ and of the integral with respect to the measure μ.

The *Markov property* can be stated as follows: given any condition H, related to the behavior of the particle before time $s \geq 0$, the process $Y(t) = B^x(t + s)$ is a Brownian motion with initial distribution[32]

$$\mu(I) = P^x \{B^x(s) \in I \,|\, H\}.$$

This property establishes the independence of the *future* process $B^x(t + s)$ from the *past* (absence of memory) when the *present* $B^x(s)$ is known and reflects the *absence of memory* of the random walk.

In the strong *Markov property*, s is substituted by a random time τ, depending only on the behavior of the particle in the interval $[0, \tau]$. In other words, to decide whether or not the event $\{\tau \leq t\}$ is true, it is enough to know the behavior of the particle up to time t. These kinds of random times are called *stopping times*. An important example is the *first exit time* from a domain, that we will consider in the next chapter. Instead, the random time τ defined by

$$\tau = \inf\{t : B(t) > 10 \text{ and } B(t+1) < 10\}$$

is *not* a stopping time. Indeed (measuring time in *seconds*), τ is "the smallest" among the times t such that the Brownian path is above level 10 at time t, and after one second is below 10. Clearly, to decide whether $\tau \leq 3$, say, it is not enough to know the path up to time $t = 3$, since τ involves the behavior of the path up to the *future* time $t = 4$.

• *Expectation.* Given a sufficiently smooth function $g = g(y)$, $y \in \mathbb{R}$, we can define the random variable

$$Z(t) = (g \circ B^x)(t) = g(B^x(t)).$$

Its expected value is given by the formula

$$E^x[Z(t)] = \int_{\mathbb{R}} g(y) P(x, t, dy) = \int_{\mathbb{R}} g(y) \Gamma(y - x, t)\, dy.$$

We will meet this formula in a completely different situation later on.

2.5 Diffusion, Drift and Reaction

2.5.1 Random walk with drift

The hypothesis of symmetry of our random walk can be removed. Suppose our unit mass particle moves along the x axis with space step $h > 0$, every time interval of duration $\tau > 0$, according to the following rules (Fig. 2.9).

1. The particle starts from $x = 0$.
2. It moves to the right with probability $p_0 \neq \frac{1}{2}$ and to the left with probability $q_0 = 1 - p_0$, independently of the previous step.
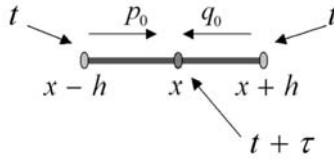
Fig. 2.9. Random walk with drift

Rule 2 breaks the symmetry of the walk and models a particle tendency to move to the right or to the left, according to the sign of $p_0 - q_0$ being positive or negative, respectively. Again we denote by $p = p(x, t)$ the probability that the particle location is $x = mh$ at time $t = N\tau$. From the total probability formula we have:

$$p(x, t + \tau) = p_0 p(x - h, t) + q_0 p(x + h, t) \tag{2.87}$$

with the usual initial conditions

$$p(0, 0) = 1 \quad \text{and} \quad p(x, 0) = 0 \quad \text{if } x \neq 0.$$

As in the symmetric case, keeping x and t fixed, we want to examine what happens when we pass to the limit for $h \to 0, \tau \to 0$. From Taylor formula, we have

$$p(x, t + \tau) = p(x, t) + p_t(x, t)\tau + o(\tau),$$

$$p(x \pm h, t) = p(x, t) \pm p_x(x, t)h + \frac{1}{2}p_{xx}(x, t)h^2 + o(h^2).$$

Substituting into (2.87), we get

$$p_t \tau + o(\tau) = \frac{1}{2}p_{xx}h^2 + (q_0 - p_0)hp_x + o(h^2). \tag{2.88}$$

A new term appears: $(q_0 - p_0)hp_x$. Dividing by τ, we obtain

$$p_t + o(1) = \frac{1}{2}\frac{h^2}{\tau}p_{xx} + \boxed{\frac{(q_0 - p_0)h}{\tau}p_x} + o\left(\frac{h^2}{\tau}\right). \tag{2.89}$$

Again, here is the crucial point. If we let $h, \tau \to 0$, we realize that the assumption

$$\frac{h^2}{\tau} = 2D \tag{2.90}$$

alone is not sufficient anymore to get something non trivial from (2.89): indeed, if we keep p_0 and q_0 constant, we have

$$\frac{(q_0 - p_0)h}{\tau} \to \infty$$

_____

[32] $P(A|H)$ denotes the conditional probability of A, given H.

and from (2.89) we get a contradiction. What else we have to require? Writing

$$\frac{(q_0 - p_0)\,h}{\tau} = \frac{(q_0 - p_0)\,h^2}{h}\,\frac{1}{\tau}$$

we see we must require, in addition to (2.90), that

$$\frac{q_0 - p_0}{h} \to \beta \qquad (2.91)$$

with β finite. Notice that, since $q_0 + p_0 = 1$, (2.91) is equivalent to

$$p_0 = \frac{1}{2} - \frac{\beta}{2}h + o\,(h) \quad \text{and} \quad q_0 = \frac{1}{2} + \frac{\beta}{2}h + o\,(h)\,, \qquad (2.92)$$

that could be interpreted as a *symmetry of the motion at a microscopic scale*.

With (2.91) at hand, we have

$$\frac{(q_0 - p_0)\,h^2}{h}\,\frac{1}{\tau} \to 2D\beta \equiv b$$

and (2.89) becomes in the limit,

$$p_t = Dp_{xx} + bp_x. \qquad (2.93)$$

We already know that Dp_{xx} models a diffusion phenomenon. Let us *unmask* the term bp_x, by first examining the dimensions of b. Since $q_0 - p_0$ is dimensionless, being a difference of probabilities, the dimensions of b are those of h/τ, namely of a **velocity**.

Thus the coefficient b codifies the tendency of the limiting continuous motion, to move towards a privileged direction with speed $|b|$: to the right if $b < 0$, to the left if $b > 0$. In other words, there exists a *current of intensity* $|b|$ driving the particle. *The random walk has become a* **diffusion process with drift**.

The last point of view calls for an analogy with the diffusion of a substance transported along a channel.

2.5.2 Pollution in a channel

In this section we examine a simple convection-diffusion model of a pollutant on the surface of a narrow channel. A water stream of constant speed v transports the pollutant along the positive direction of the x axis. We can neglect the depth of the water (thinking to a floating pollutant) and the transverse dimension (thinking of a very narrow channel).

Our purpose is to derive a mathematical model capable of describing the evolution of the concentration[33] $c = c\,(x, t)$ of the pollutant. Accordingly, the integral

$$\int_x^{x+\Delta x} c\,(y, t)\,dy \qquad (2.94)$$

---

[33] $[c] = [mass] \times [length]^{-1}$.

gives the mass inside the interval $[x, x + \Delta x]$ at time t (Fig. 2.10). In the present case there are neither sources nor sinks of pollutant, therefore to construct a model we use the **law of mass conservation**: *the growth rate of the mass contained in an interval* $[x, x + \Delta x]$ *equals the net mass flux into* $[x, x + \Delta x]$ *through the end points.*
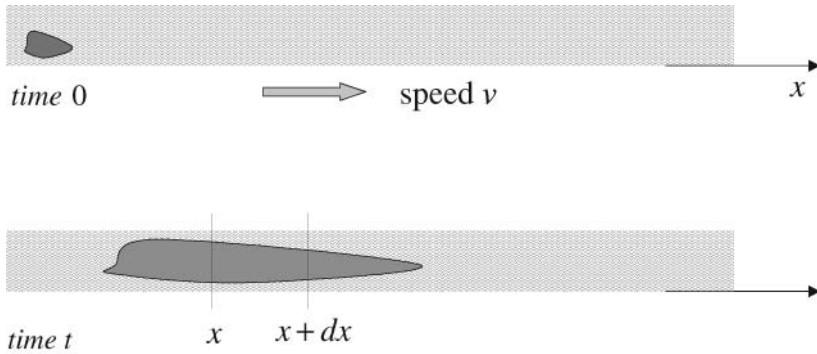
Fig. 2.10. Pollution in a narrow channel

From (2.94), the growth rate of the mass contained in an interval $[x, x + \Delta x]$ is given by [34]

$$\frac{d}{dt} \int_x^{x+\Delta x} c(y, t)\, dy = \int_x^{x+\Delta x} c_t(y, t)\, dy. \tag{2.95}$$

Denote by $q = q(x, t)$ the mass flux[35] *entering* the interval $[x, x + \Delta x]$, through the point x at time t. The net mass flux into $[x, x + \Delta x]$ through the end points is

$$q(x, t) - q(x + \Delta x, t). \tag{2.96}$$

Equating (2.95) and (2.96), the law of mass conservation reads

$$\int_x^{x+\Delta x} c_t(y, t)\, dy = q(x, t) - q(x + \Delta x, t).$$

Dividing by Δx and letting $\Delta x \to 0$, we find the basic law

$$c_t = -q_x. \tag{2.97}$$

At this point we have to decide which kind of mass flux we are dealing with. In other words, we need a *constitutive relation for* q. There are several possibilities, for instance:

[34] Assuming we can take the derivative inside the integral.
[35] $[q] = [mass] \times [time]^{-1}$

a) *Convection.* The flux is determined by the water stream only. This case corresponds to a bulk of pollutant that is driven by the stream, without deformation or expansion. Translating into mathematical terms we find

$$q(x,t) = vc(x,t)$$

where, we recall, v denotes the stream speed.

b) *Diffusion.* The pollutant expands from higher concentration regions to lower ones. We have seen something like that in heat conduction, where, according to the Fourier law, the heat flux is proportional and opposite to the temperature gradient. Here we can adopt a similar law, that in this setting is known as the *Fick's law* and reads

$$q(x,t) = -Dc_x(x,t)$$

where the constant D depends on the polluting and has the usual dimensions ($[D] = [length]^2 \times [time]^{-1}$).

In our case, convection and diffusion are both present and therefore we superpose the two effects, by writing

$$q(x,t) = vc(x,t) - Dc_x(x,t).$$

From (2.97) we deduce

$$c_t = Dc_{xx} - vc_x \tag{2.98}$$

which constitutes our mathematical model and turns out to be identical to (2.93).

Since D and v are constant, it is easy to determine the evolution of a mass Q of pollutant, initially located at the origin (say). Its concentration is the solution of (2.98) with initial condition

$$c(x,0) = Q\delta(x)$$

where δ is the Dirac measure at the origin. To find an explicit formula, we can get rid of the drift term $-vc_x$ by setting

$$w(x,t) = c(x,t)e^{hx+kt}$$

with h, k to be chosen suitably. We have:

$$w_t = [c_t + kc]e^{hx+kt}$$
$$w_x = [c_x + hc]e^{hx+kt}, \qquad w_{xx} = [c_{xx} + 2hc_x + h^2c]e^{hx+kt}.$$

Using the equation $c_t = Du_{xx} - vc_x$, we can write

$$w_t - Dw_{xx} = e^{hx+kt}[c_t - Dc_{xx} - 2Dhc_x + (k - Dh^2)c] =$$
$$= e^{hx+kt}[(-v - 2Dh)c_x + (k - Dh^2)c].$$

Thus if we choose

$$h = -\frac{v}{2D} \qquad \text{and} \qquad k = \frac{v^2}{4D},$$

w is a solution of the diffusion equation $w_t - Dw_{xx} = 0$, with the initial condition

$$w(x, 0) = c(x, 0) e^{-\frac{v}{2D}x} = Q\delta(x) e^{-\frac{v}{2D}x}.$$

In chapter 7 we show that $\delta(x) e^{-\frac{v}{2D}x} = \delta(x)$, so that $w(x, t) = Q\Gamma_D(x, t)$ and finally

$$c(x, t) = Qe^{\frac{v}{2D}\left(x - \frac{v}{2}t\right)}\Gamma_D(x, t). \qquad (2.99)$$

The concentration c is thus given by the fundamental solution Γ_D, "carried" by the travelling wave $\exp\left\{\frac{v}{2D}\left(x - \frac{v}{2}t\right)\right\}$, in motion to the right with speed $v/2$.

In realistic situations, the pollutant undergoes some sort of decay, for instance by biological decomposition. The resulting equation for the concentration becomes

$$c_t = Dc_{xx} - vc_x - \gamma c$$

where γ is a rate of decay[36]. We deal with this case in the next section via a suitable variant of our random walk.

2.5.3 Random walk with drift and reaction

We go back to our $1-$ dimensional random walk, assuming that the particle loses mass at the constant rate $\gamma > 0$. This means that in an interval of time from t to $t + \tau$ a percentage of mass

$$Q(x, t) = \tau\gamma p(x, t)$$

disappears. The difference equation (2.87) for p becomes

$$p(x, t + \tau) = p_0[p(x - h, t) - Q(x - h, t)] + q_0[p(x + h, t) - Q(x + h, t)]$$

Since[37]

$$p_0 Q(x - h, t) + q_0 Q(x + h, t) = Q(x, t) + (q_0 - p_0)hQ_x(x, t) + \dots$$
$$= \tau\gamma p(x, t) + O(\tau h),$$

equation (2.88) modifies into

$$p_t \tau + o(\tau) = \frac{1}{2}p_{xx}h^2 + (q_0 - p_0)hp_x - \tau\gamma p + O(\tau h) + o(h^2).$$

Dividing by τ, letting $h, \tau \to 0$ and assuming

$$\frac{h^2}{\tau} = 2D, \qquad \frac{q_0 - p_0}{h} \to \beta,$$

we get

$$p_t = Dp_{xx} + bp_x - \gamma p \qquad (b = 2D\beta). \qquad (2.100)$$

---

[36] $[\gamma] = [time]^{-1}$.

[37] The symbol "$O(k)$" ("big O of k") denotes a quantity of order k.

The term $-\gamma p$ appears in (2.100) as a decaying term. On the other hand, in important situations, γ could be *negative*, meaning that this time we have a *creation of mass* at the rate $|\gamma|$. For this reason the last term is called generically a *reaction term* and (2.100) is a *diffusion equation with drift and reaction*.

Going back to equation (2.100), it is useful to look separately at the effect of the three terms in its right hand side.

- $p_t = Dp_{xx}$ models pure diffusion. The typical effects are spreading and smoothing, as shown by the typical behavior of the fundamental solution Γ_D.

- $p_t = bp_x$ is a pure transport equation, that we will consider in detail in chapter 3. The solutions are travelling waves of the form $g(x + bt)$.

- $p_t = -\gamma p$ models pure reaction. The solutions are multiples of $e^{-\gamma t}$, exponentially decaying (increasing) if $\gamma > 0$ ($\gamma < 0$).

So far we have given a probabilistic interpretation for a motion in all $\mathbb{R}$, where no boundary condition is present. The problems 7 and 8 give a probabilistic interpretation of the Dirichlet and Neumann condition in terms of *absorbing* and *reflecting* boundaries, respectively.

2.6 Multidimensional Random Walk

2.6.1 The symmetric case

What we have done in dimension $n = 1$ can be extended without much effort to dimension $n > 1$, in particular $n = 2, 3$. To define a symmetric random walk, we introduce the *lattice* $\mathbb{Z}^n$ given by the set of points $\mathbf{x} \in \mathbb{R}^n$, whose coordinates are signed integers. Given the *space step* $h > 0$, the symbol $h\mathbb{Z}^n$ denotes the lattice of points whose coordinates are signed integers *multiplied by h*.

Every point $\mathbf{x} \in h\mathbb{Z}^n$, has a "discrete neighborhood" of $2n$ points at distance h, given by

$$\mathbf{x} + h\mathbf{e}_j \qquad \text{and} \qquad \mathbf{x} - h\mathbf{e}_j \qquad (j = 1, ..., n),$$

where $\mathbf{e}_1, ..., \mathbf{e}_n$ is the canonical basis in $\mathbb{R}^n$. Our particle moves in $h\mathbb{Z}^n$ according to the following rules (Fig. 2.11).

1. It starts from $\mathbf{x} = \mathbf{0}$.
2. If it is located in $\mathbf{x}$ at time t, at time $t + \tau$ the particle location is at one of the $2n$ points $\mathbf{x} \pm h\mathbf{e}_j$, with probability $p = \frac{1}{2n}$.
3. Each step is independent of the previous one.

As in the $1-dimensional\ case$, our task is to *compute the probability $p(\mathbf{x}, t)$ of finding the particle at $\mathbf{x}$ at time t*.

Clearly the initial conditions for p are

$$p(\mathbf{0}, 0) = 1 \quad \text{and} \quad p(\mathbf{x}, 0) = 0 \quad \text{if } \mathbf{x} \neq \mathbf{0}.$$
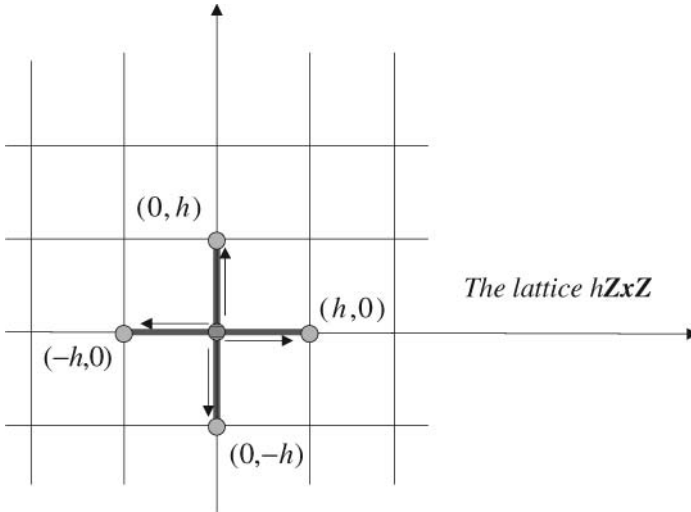
Fig. 2.11. Bidimensional random walk

The total probability formula gives

$$p\left(\mathbf{x}, t + \tau\right) = \frac{1}{2n} \sum_{j=1}^{n} \left\{p\left(\mathbf{x} + h\mathbf{e}_j, t\right) + p\left(\mathbf{x} - h\mathbf{e}_j, t\right)\right\}. \tag{2.101}$$

Indeed, to reach the point $\mathbf{x}$ at time $t + \tau$, at time t the particle must have been located at one of the points in the discrete neighborhood of $\mathbf{x}$ and moved from there towards $\mathbf{x}$ with probability $1/2n$. For fixed $\mathbf{x}$ and t, we want to examine what happens when we let $h \to 0, \tau \to 0$. Assuming p defined and smooth in all of $\mathbb{R}^n \times (0, +\infty)$, we use Taylor's formula to write

$$p\left(\mathbf{x}, t + \tau\right) = p\left(\mathbf{x}, t\right) + p_t\left(\mathbf{x}, t\right)\tau + o\left(\tau\right)$$

$$p\left(\mathbf{x} \pm h\mathbf{e}_j, t\right) = p\left(\mathbf{x}, t\right) \pm p_{x_j}\left(\mathbf{x}, t\right)h + \frac{1}{2}p_{x_j x_j}\left(\mathbf{x}, t\right)h^2 + o\left(h^2\right).$$

Substituting into (2.101), after some simplifications, we get

$$p_t \tau + o\left(\tau\right) = \frac{h^2}{2n}\Delta p + o\left(h^2\right).$$

Dividing by τ we obtain the equation

$$p_t + o\left(1\right) = \frac{1}{2n}\frac{h^2}{\tau}\Delta p + o\left(\frac{h^2}{\tau}\right). \tag{2.102}$$

The situation is quite similar to the $1-$ dimensional case: still, to obtain eventually something non trivial, we must require that the ratio h^2/τ has a finite and positive limit. The simplest choice is

$$\frac{h^2}{\tau} = 2nD \tag{2.103}$$

with $D > 0$. From (2.103), we deduce that *in unit time, the particle diffuses at an average distance of* $\sqrt{2nD}$. The physical dimensions of D have not changed. Letting $h \to 0, \tau \to 0$ in (2.102), we find for p the diffusion equation

$$p_t = D\Delta p \tag{2.104}$$

with the initial condition

$$\lim_{t \to 0^+} p(\mathbf{x}, t) = \delta. \tag{2.105}$$

Since $\int_{\mathbb{R}^n} p(\mathbf{x}, t)\, d\mathbf{x} = 1$ for every t, the unique solution is given by

$$p(\mathbf{x}, t) = \Gamma_D(\mathbf{x}, t) = \frac{1}{(4\pi Dt)^{n/2}} e^{-\frac{|\mathbf{x}|^2}{4Dt}}, \qquad t > 0.$$

The $n-$dimensional random walk has become a *continuous walk*; when $D = \frac{1}{2}$, it is called $n-$*dimensional Brownian motion*. Denote by $\mathbf{B}(t) = \mathbf{B}(t, \omega)$ the random position of a Brownian particle, defined for every $t > 0$ on a probability space $(\Omega, \mathcal{F}, P)^{[38]}$.

The family of random variables $\mathbf{B}(t, \omega)$, with time t as a real parameter, **is a vector valued continuous stochastic process**. For $\omega \in \Omega$ fixed, the *vector function*

$$t \longmapsto \mathbf{B}(t, \omega)$$

describes an $n-$dimensional Brownian path, whose main features are listed below.

• *Path continuity.* With probability 1, the Brownian paths are continuous for $t \geq 0$.

• *Gaussian law for increments.* The process $\mathbf{B}^{\mathbf{x}}(t) = \mathbf{x} + \mathbf{B}(t)$ defines a Brownian motion with start at $\mathbf{x}$. With every point $\mathbf{x}$ is associated a probability $P^{\mathbf{x}}$, with the following properties (if $\mathbf{x} = \mathbf{0}$, $P^{\mathbf{0}} = P$).

a) $P^{\mathbf{x}}\{\mathbf{B}^{\mathbf{x}}(0) = \mathbf{x}\} = P\{\mathbf{B}(0) = \mathbf{0}\} = 1$.

b) For every $s \geq 0, t \geq 0$, the increment

$$\mathbf{B}^{\mathbf{x}}(t + s) - \mathbf{B}^{\mathbf{x}}(s) = \mathbf{B}(t + s) - \mathbf{B}(s) \tag{2.106}$$

follows a *normal law with zero mean value and covariance matrix equal to* $t\mathbf{I}_n$, whose density is

$$\Gamma(\mathbf{x}, t) = \Gamma_{\frac{1}{2}}(\mathbf{x}, t) = \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|\mathbf{x}|^2}{2t}}.$$

Moreover, (2.106) is independent of any event occurred at any time less than s. For instance, the two events

$$\{\mathbf{B}(t_2) - \mathbf{B}(t_1) \in A_1\} \qquad \{\mathbf{B}(t_1) - \mathbf{B}(t_0) \in A_2\}$$

are independent if $t_0 < t_1 < t_2$.

---

[38] See Appendix B.

- *Transition function.* For each Borel set $A \subseteq \mathbb{R}^n$ a *transition function*

$$P(\mathbf{x}, t, A) = P^{\mathbf{x}} \{\mathbf{B}^{\mathbf{x}}(t) \in A\}$$

is defined, representing the probability that the particle, initially located at $\mathbf{x}$, belongs to A at time t. We have:

$$P(\mathbf{x}, t, A) = P\{\mathbf{B}(t) \in A - \mathbf{x}\} = \int_{A-\mathbf{x}} \Gamma(\mathbf{y}, t) \, d\mathbf{y} = \int_{A} \Gamma(\mathbf{y} - \mathbf{x}, t) \, d\mathbf{y}.$$

- *Invariance.* The motion is invariant with respect to rotations and translations.

- *Markov and strong Markov properties.* Let μ be a probability measure[39] on $\mathbb{R}^n$. If the particle has a random initial position with probability distribution μ, we can consider a *Brownian motion with initial distribution* μ, and for it we use the symbol $\mathbf{B}^\mu$. To $\mathbf{B}^\mu$ is associated a probability distribution P^μ such that

$$P^\mu \{\mathbf{B}^\mu(0) \in A\} = \mu(A).$$

The probability that the particle belongs to A at time t can be computed through the formula

$$P^\mu \{\mathbf{B}^\mu(t) \in A\} = \int_{\mathbf{R}^n} P(\mathbf{x}, t, A) \mu(d\mathbf{x}). \tag{2.107}$$

The *Markov property* can be stated as follows: given any condition H, related to the behavior of the particle before time $s \geq 0$, the process $\mathbf{Y}(t) = \mathbf{B}^{\mathbf{x}}(t + s)$, is a Brownian motion with initial distribution

$$\mu(A) = P^{\mathbf{x}} \{\mathbf{B}^{\mathbf{x}}(s) \in A \,| H\}.$$

Again, this property establishes the independence of the *future* process $\mathbf{B}^{\mathbf{x}}(t + s)$ from the *past* when the *present* $\mathbf{B}^{\mathbf{x}}(s)$ is known and encodes the *absence of memory* of the process. In the strong Markov property, a stopping time τ takes the place of s.

- *Expectation.* Given any sufficiently smooth real function $g = g(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^n$, we can define the real random variable

$$Z(t) = (g \circ \mathbf{B}^{\mathbf{x}})(t) = g(\mathbf{B}^{\mathbf{x}}(t)).$$

Its *expectation* is given by the formula

$$E[Z(t)] = \int_{\mathbb{R}^n} g(\mathbf{y}) P(\mathbf{x}, t, d\mathbf{y}) = \int_{\mathbb{R}^n} g(\mathbf{y}) \Gamma(\mathbf{y} - \mathbf{x}, t) \, d\mathbf{y}.$$

---

[39] See Appendix B for the definition of a probability measure μ and of the integral with respect to the measure μ .

2.6.2 Walks with drift and reaction

As in the $1-$dimensional case, we can construct several variants of the symmetric random walk. For instance, we can allow a different behavior along each direction, by choosing the space step h_j depending on $\mathbf{e}_j$. As a consequence the limit process models an anisotropic motion, codified in the matrix

$$
\mathbf{D} = \begin{pmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & D_n \end{pmatrix}
$$

where $D_j = h_j^2/2n\tau$ is the diffusion coefficient in the direction $\mathbf{e}_j$. The resulting equation for the transition probability $p(\mathbf{x}, t)$ is

$$
p_t = \sum_{j=1}^{n} D_j p_{x_j x_j}. \tag{2.108}
$$

We may also break the symmetry by asking that along the direction $\mathbf{e}_j$ the probability to go to the left (right) is q_j (resp. p_j). If

$$
\frac{q_j - p_j}{h_j} \to \beta_j \qquad \text{and} \qquad b_j = 2D_j\beta_j,
$$

the vector $\mathbf{b} = (b_1, ..., b_n)$ plays a role of a *drift vector*, reflecting the tendency of motion to move asymmetrically along each coordinate axis. Adding a reaction term of the form cp, the resulting *drift-diffusion-reaction* equation is

$$
p_t = \sum_{j=1}^{n} D_j p_{x_j x_j} + \sum_{j=1}^{n} b_j u_{x_j} + cp. \tag{2.109}
$$

In problem 2.17 we ask the reader to fill in all the details in the argument leading to equations (2.108) and (2.109). We will deal with general equations of these type in Chapter 9.

2.7 An Example of Reaction$-$Diffusion ($n = 3$)

In this section we examine a model of reaction-diffusion in a fissionable material. Although we deal with a greatly simplified model, some interesting implications can be drawn.

By shooting neutrons into an uranium nucleus it may happen that the nucleus breaks into two parts, releasing other neutrons already present in the nucleus and causing a chain reaction. Some macroscopic aspects of this phenomenon can be described by means of an elementary model.

Suppose a cylinder with height h and radius R is made of a fissionable material of constant density ρ, with total mass

$$M = \pi \varrho R^2 h.$$

At a macroscopic level, the free neutrons diffuse like a chemical in a porous medium, with a flux proportional and opposite to the density gradient. In other terms, if $N = N(x, y, z, t)$ is the *neutron density* and no fission occurs, the *flux of neutrons is equal to* $-k\nabla N$, where k is a positive constant depending on the material. The mass conservation then gives

$$N_t = k\Delta N.$$

When fission occurs at a constant rate $\gamma > 0$, we get the equation

$$N_t = D\Delta N + \gamma N, \tag{2.110}$$

where reaction and diffusion are competing: diffusion tends to slow down N, while, clearly, the reaction term tends to exponentially increase N. A crucial question is to examine the behavior of N in the long run (i.e. as $t \to +\infty$).

We look for *bounded* solutions satisfying a homogeneous Dirichlet condition on the boundary of the cylinder, with the idea that the density is higher at the center of the cylinder and very low near the boundary. Then it is reasonable to assume that N has a radial distribution with respect to the axis of the cylinder. More precisely, using the cylindrical coordinates (r, θ, z) with

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

we can write $N = N(r, z, t)$ and the homogeneous Dirichlet condition on the boundary of the cylinder translates into

$$\begin{aligned} N(R, z, t) &= 0 & 0 < z < h \\ N(r, 0, t) &= N(r, h, t) = 0 & 0 < r < R \end{aligned} \tag{2.111}$$

for every $t > 0$. Accordingly we prescribe an initial condition

$$N(r, z, 0) = N_0(r, z) \tag{2.112}$$

such that

$$N_0(R, z) = 0 \text{ for } 0 < z < h, \text{ and } N_0(r, 0) = N_0(r, h) = 0. \tag{2.113}$$

To solve problem (2.110), (2.111), (2.112), let us first get rid of the reaction term by setting

$$N(r, z, t) = \mathcal{N}(r, z, t)\, e^{\gamma t}. \tag{2.114}$$

Then, writing the Laplace operator in cylindrical coordinates[40], $\mathcal{N}$ solves

$$\mathcal{N}_t = k \left[\mathcal{N}_{rr} + \frac{1}{r}\mathcal{N}_r + \mathcal{N}_{zz} \right] \tag{2.115}$$

---

[40] Appendix C.

with the same initial and boundary conditions of N. By maximum principle, we know that there exists only one solution, continuous up to the boundary of the cylinder. To find an explicit formula for the solution, we use the method of separation of variables, first searching for bounded solutions of the form

$$N\left(r, z, t\right) = u\left(r\right) v\left(z\right) w\left(t\right), \tag{2.116}$$

satisfying the homogeneous Dirichlet conditions $u\left(R\right) = 0$ and $v\left(0\right) = v\left(h\right) = 0$.

Substituting (2.116) into (2.115), we find

$$u\left(r\right) v\left(z\right) w'\left(t\right) = k[u''\left(r\right) v\left(z\right) w\left(t\right) + \frac{1}{r} u'\left(r\right) v\left(z\right) w\left(t\right) + u\left(r\right) v''\left(z\right) w\left(t\right)].$$

Dividing by N and rearranging the terms, we get,

$$\frac{w'\left(t\right)}{kw\left(t\right)} - \left[\frac{u''\left(r\right)}{u\left(r\right)} + \frac{1}{r} \frac{u'\left(r\right)}{u\left(r\right)}\right] = \frac{v''\left(z\right)}{v\left(z\right)}. \tag{2.117}$$

The two sides of (2.117) depend on different variables so that they must be equal to a common constant b. Then for v we have the eigenvalue problem

$$v''\left(z\right) - bv\left(z\right) = 0$$

$$v\left(0\right) = v\left(h\right) = 0.$$

The eigenvalues are $b_m \equiv -\nu_m^2 = -\frac{m^2 \pi^2}{h^2}$, $m \geq 1$ integer, with corresponding eigenfunctions

$$\nu_m\left(z\right) = \sin \nu_m z.$$

The equation for w and u can be written in the form:

$$\frac{w'\left(t\right)}{kw\left(t\right)} + \nu_m^2 = \frac{u''\left(r\right)}{u\left(r\right)} + \frac{1}{r} \frac{u'\left(r\right)}{u\left(r\right)} \tag{2.118}$$

where the variables r and t are again separated. This forces the two sides of (2.118) to be equal to a common constant μ. Therefore, for w we have the equation

$$w'\left(t\right) = k(\mu - \nu_m^2)w\left(t\right)$$

that gives

$$w\left(t\right) = c \exp\left[k\left(\mu - \nu_m^2\right) t\right] \qquad c \in \mathbb{R}. \tag{2.119}$$

Then the equation for u is

$$u''\left(r\right) + \frac{1}{r} u'\left(r\right) - \mu u\left(r\right) = 0 \tag{2.120}$$

with

$$u\left(R\right) = 0 \quad \text{and} \quad u \text{ bounded in } [0, R]. \tag{2.121}$$

The (2.120) is a *Bessel equation of order zero with parameter* $-\mu$; conditions (2.121) force[41] $\mu = -\lambda^2 < 0$. Then the only bounded solution of (2.120), (2.121) is $J_0(\lambda r)$, where

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k!)^2} \left(\frac{x}{2}\right)^{2k}$$

is the *Bessel function of first kind and order zero.* To match the boundary condition $u(R) = 0$ we require $J_0(\lambda R) = 0$. Now, J_0 has an infinite number of positive simple zeros[42] λ_n, $n \geq 1$:

$$0 < \lambda_1 < \lambda_2 < ... < \lambda_n < ...$$

Thus, if $\lambda R = \lambda_n$, we find infinitely many solutions of (2.120), given by

$$u_n(r) = J_0\left(\frac{\lambda_n r}{R}\right).$$

Thus

$$\mu = \mu_n = -\frac{\lambda_n^2}{R^2}.$$

To summarize, we have determined so far a countable number of solutions

$$\mathcal{N}_{mn}(r, z, t) = u_n(r)\, v_m(z)\, w_{m,n}(t) =$$

$$= J_0\left(\frac{\lambda_n r}{R}\right) \sin \nu_m z \,\exp\left[-k\left(\nu_m^2 + \frac{\lambda_n^2}{R^2}\right)t\right]$$

satisfying the homogeneous Dirichlet conditions. It remains to satisfy the initial condition. Due to the linearity of the problem, we look for a solution obtained by superposition of the $\mathcal{N}_{m,n}$, that is

$$\mathcal{N}(r, z, t) = \sum_{n,m=1}^{\infty} c_{mn} \mathcal{N}_{mn}(r, z, t).$$

---

[41] In fact, write Bessel's equation (2.120) in the form

$$(ru')' - \mu r u = 0.$$

Multiplying by u and integrating over $(0, R)$, we have

$$\int_0^R (ru')' u\, dr = \mu \int_0^R u^2 dr. \qquad (2.122)$$

Integrating by parts and using (2.121), we get

$$\int_0^R (ru')' u\, dr = \left[(ru')u\right]_0^R - \int_0^R (u')^2 dr = -\int_0^R (u')^2 dr < 0$$

and from (2.122) we get $\mu < 0$.

[42] The zeros of the Bessel functions are known with a considerable degree of accuracy. The first five zeros of J_0 are: 2.4048..., 5.5201..., 8.6537..., 11.7915..., 14.9309....

Then, we choose the coefficients c_{mn} in order to have

$$\sum_{n,m=1}^{\infty} c_{mn} \mathcal{N}_{mn}(r,z,0) = \sum_{n,m=1}^{\infty} c_{mn} J_0\left(\frac{\lambda_n r}{R}\right) \sin\frac{m\pi}{h}z = N_0(r,z). \qquad (2.123)$$

The second of (2.113) and (2.123) suggest an expansion of N_0 in sine Fourier series with respect to z. Let

$$c_m(r) = \frac{2}{h}\int_0^h N(r,z)\sin\frac{m\pi}{h}z, \qquad m \geq 1,$$

and

$$N_0(r,z) = \sum_{m=1}^{\infty} c_m(r)\sin\frac{m\pi}{h}z.$$

Then (2.123) shows that, for fixed $m \geq 1$, the c_{mn} are the coefficients of the expansion of $c_m(r)$ in the *Fourier-Bessel series*

$$\sum_{n=1}^{\infty} c_{mn} J_0\left(\frac{\lambda_n r}{R}\right) = c_m(r).$$

We are not really interested in the exact formula for the c_{mn}, however we will come back to this point in Remark 2.5 below.

In conclusion, recalling (2.114), the analytic expression of the solution of our original problem is the following:

$$N(r,z,t) = \sum_{n,m=1}^{\infty} c_{mn} J_0\left(\frac{\lambda_n r}{R}\right)\exp\left\{\left(\gamma - k\nu_m^2 - k\frac{\lambda_n^2}{R^2}\right)t\right\}\sin\nu_m z. \qquad (2.124)$$

Of course, (2.124) is only a formal solution, since we should check in which sense the boundary and initial condition are attained and that term by term differentiation can be performed. This can be done under reasonable smoothness properties of N_0 and we do not pursue the calculations here.

Rather, we notice that from (2.124) we can draw an interesting conclusion on the long range behavior of N. Consider for instance the value of N at the center of the cylinder, that is at the point $r = 0$ and $z = h/2$; we have, since $J_0(0) = 1$ and $\nu_m^2 = \frac{m^2\pi^2}{h^2}$,

$$N\left(0,\frac{h}{2},t\right) = \sum_{n,m=1}^{\infty} c_{mn}\ \exp\left\{\left(\gamma - k\frac{m^2\pi^2}{h^2} - k\frac{\lambda_n^2}{R^2}\right)t\right\}\sin\frac{m\pi}{2}.$$

The exponential factor is maximized for $m = n = 1$, so the leading term in the sum is

$$c_{11}\exp\left\{\left(\gamma - k\frac{\pi^2}{h^2} - k\frac{\lambda_1^2}{R^2}\right)t\right\}.$$

If now

$$\gamma - k\left(\frac{\pi^2}{h^2} + \frac{\lambda_1^2}{R^2}\right) < 0,$$

each term in the series goes to zero as $t \to +\infty$ and the reaction dies out. On the opposite, if

$$\gamma - k\left(\frac{\pi^2}{h^2} + \frac{\lambda_1^2}{R^2}\right) > 0,$$

that is

$$\frac{\gamma}{k} > \frac{\pi^2}{h^2} + \frac{\lambda_1^2}{R^2}, \tag{2.125}$$

the leading term increases exponentially with time. To be true, (2.125) requires that the following relations be *both* satisfied:

$$h^2 > \frac{k\pi^2}{\gamma} \quad \text{and} \quad R^2 > \frac{k\lambda_1^2}{\gamma}. \tag{2.126}$$

The (2.126) gives a lower bound for the height and the radius of the cylinder. Thus, we deduce that *there exists a critical mass of material, below which the reaction cannot be sustained.*
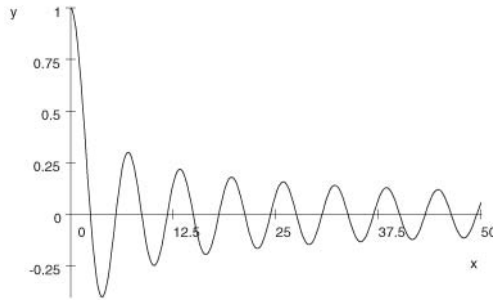
Fig. 2.12. The Bessel function J_0

Remark 2.5. A sufficiently smooth function f, for instance of class $C^1([0, R])$, can be expanded in a Fourier-Bessel series, where the Bessel functions $J_0\left(\frac{\lambda_n r}{R}\right)$, $n \geq 1$, play the same role of the trigonometric functions. More precisely, the functions $J_0(\lambda_n r)$ satisfy the following orthogonality relations:

$$\int_0^R x J_0(\lambda_m x) J_0(\lambda_n x) dx = \begin{cases} 0 & m \neq n \\ \frac{R^2}{2} c_n^2 & m = n \end{cases}$$

where

$$c_n = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,(k+1)!} \left(\frac{\lambda_n}{2R}\right)^{2k+1}.$$

Then

$$f(x) = \sum_{n=0}^{\infty} f_n J_0(\lambda_n x) \qquad (2.127)$$

with the coefficients f_n assigned by the formula

$$f_n = \frac{2}{R^2 c_n^2} \int_0^R x f(x) J_0(\lambda_n x)\, dx.$$

The series (2.127) converges in the following least square sense: if

$$S_N(x) = \sum_{n=0}^{N} f_n J_0(\lambda_n x)$$

then

$$\lim_{N \to +\infty} \int_0^R [f(x) - S_N(x)]^2 x\, dx = 0. \qquad (2.128)$$

In Chapter 6, we will interpret (2.128) from the point of view of Hilbert space theory.

2.8 The Global Cauchy Problem ($n = 1$)

2.8.1 The homogeneous case

In this section we consider the global Cauchy problem

$$\begin{cases} u_t - D u_{xx} = 0 & \text{in } \mathbb{R} \times (0, \infty) \\ u(x,0) = g(x) & \text{in } \mathbb{R} \end{cases} \qquad (2.129)$$

where g, the *initial data,* is given. We will limit ourselves to the one dimensional case; techniques, ideas and formulas can be extended without too much effort to the $n-$dimensional case.

The problem (2.129) models the evolution of the temperature or of the concentration of a substance along a very long (infinite) bar or channel, respectively, given the initial ($t = 0$) distribution.

By heuristic considerations, we can guess what could be a candidate solution. Consider a unit mass composed of a large number $M \gg 1$ of particles and interpret the solution u as their concentration (or percentage). Then, $u(x,t)\, dx$ gives the mass inside the interval $(x, x + dx)$ at time t.

We want to determine the concentration $u(x, y)$, due to the diffusion of a mass whose initial concentration is given by g.

Thus, the quantity $g(y)\, dy$ represents the mass concentrated in the interval $(y, y + dy)$ at time $t = 0$. As we have seen, $\Gamma(x - y, t)$ is a *unit source solution,*

representing the concentration at x at time t, due to the diffusion of a unit mass, initially concentrated in the same interval. Accordingly,

$$\Gamma_D (x - y, t)\, g(y)\, dy$$

gives the concentration at x at time t, due to the diffusion of the mass $g(y)\, dy$.

Thanks to the linearity of the diffusion equation, we can use the *superposition principle* and compute the solution as the sum of all contributions. In this way, we get the formula

$$u(x, t) = \int_{\mathbb{R}} g(y)\, \Gamma_D (x - y, t)\, dy = \frac{1}{\sqrt{4\pi D t}} \int_{\mathbb{R}} g(y)\, e^{-\frac{(x-y)^2}{4Dt}}\, dy. \qquad (2.130)$$

Clearly, one has to check rigorously that, under reasonable hypotheses on the initial data g, formula (2.130) really gives the unique solution of the Cauchy problem. This is not a negligible question. First of all, if g grows too much at infinity, more than an exponential of the type $e^{ax^2}, a > 0$, in spite of the rapid convergence to zero of the Gaussian, the integral in (2.130) could be divergent and formula (2.130) loses any meaning. Even more delicate is the question of the uniqueness of the solution, as we will see later.

Remark 2.6. Formula (2.130) has a probabilistic interpretation. Let $D = \frac{1}{2}$ and let $B^x(t)$ be the position of a Brownian particle, started at x. Let $g(y)$ be the *gain* obtained when the particle crosses y. Then, we can write:

$$u(x, t) = E^x [g(B^x(t))]$$

where E^x denotes *the expected value* with respect to the probability P^x, with density $\Gamma(x - y, t)$[43].

In other words: *to compute u at the point (x, t), consider a Brownian particle starting at x, compute its position $B^x(t)$ at time t, and finally compute the expected value of $g(B^x(t))$.*

2.8.2 Existence of a solution

The following theorem states that (2.130) is indeed a solution of the global Cauchy problem under rather general hypotheses on g, satisfied in most of the interesting applications[44].

Theorem 2.3. *Assume that g is a function with a finite number of jump discontinuities in $\mathbb{R}$ and there exist positive numbers a and c such that*

$$|g(x)| \leq c e^{ax^2} \qquad \forall x \in \mathbb{R}. \qquad (2.131)$$

---

[43] Appendix B.
[44] We omit the long and technical proof.

Let u be given by formula (2.130). Then:

i) $u \in C^{\infty}(\mathbb{R} \times (0, T))$ *for* $T < \dfrac{1}{4Da}$, *and in the strip* $\mathbb{R} \times (0, T)$

$$u_t - Du_{xx} = 0.$$

ii) *If* x_0 *is a continuity point of g, then*

$$u(y, t) \to g(x_0) \qquad if \ \ (y, t) \to (x_0, 0), t > 0.$$

iii) *There are positive numbers* c_1 *and A such that*

$$|u(x, t)| \le Ce^{Ax^2} \qquad \forall (x, t) \in \mathbb{R} \times (0, \infty).$$

Remark 2.7. The theorem says that, if we allow an initial data with a controlled exponential growth at infinity expressed by (2.131), then (2.130) is a solution in the strip $\mathbb{R} \times (0, T)$. We will see that, under the stated conditions, (2.130) is actually *the unique solution.*

In some applications (e.g. to Finance), the initial data grows at infinity no more than $c_1 e^{a_1 |x|}$. In this case (2.131) is satisfied by choosing any positive number a and a suitable c. This means that there is really no limitation on T, since

$$T < \frac{1}{4Da}$$

and a can be chosen as small as we like.

Remark 2.8. The property *i)* shows a typical and important phenomenon connected with the diffusion equation: even if the initial data is discontinuous at some point, immediately after the solution is smooth. The diffusion is therefore a **smoothing process**. In figure 2.13, this phenomenon is shown for the initial data $g(x) = \chi_{(-2,0)}(x) + \chi_{(1,4)}(x)$, where $\chi_{(a,b)}$ denotes the characteristic function of the interval (a, b). By *ii)*, if the initial data g is continuous in all of $\mathbb{R}$, then the solution is continuous up to $t = 0$, that is in $\mathbb{R} \times [0, T)$.
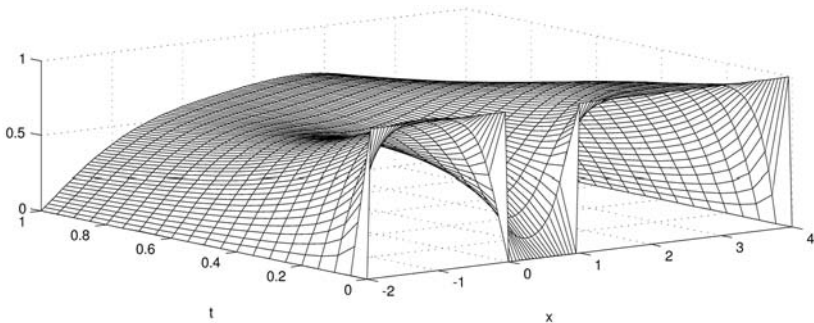
Fig. 2.13. Smoothing effect of the diffusion equation

2.8.3 The non homogeneous case. Duhamel's method

The difference equation (or the total probability formula)

$$p(x, t + \tau) = \frac{1}{2} p(x - h, t) + \frac{1}{2} p(x + h, t)$$

that we found in subsection 2.4.2 during the analysis of the symmetric random walk could be considered a probabilistic version of the *mass conservation principle*: the density of the mass located at x at time $t + \tau$ is the sum of the densities diffused from $x + h$ and $x - h$ at time t; no mass has been lost or added over the time interval $[t, t + \tau]$. Accordingly, the expression

$$p(x, t + \tau) - [\frac{1}{2} p(x - h, t) + \frac{1}{2} p(x + h, t)] \tag{2.132}$$

could be considered as a measure of the lost/added mass density over the time interval from t to $t + \tau$. Expanding with Taylor's formula as we did in Section 4.2, keeping $h^2 / \tau = 2D$, dividing by τ and letting h, $\tau \to 0$ in (2.132), we find

$$p_t - D p_{xx}.$$

Thus the differential operator $\partial_t - D \partial_{xx}$ **measures the instantaneous density production rate.**

Suppose now that from time $t = 0$ until a certain time $t = s > 0$ no mass is present and that at time s a unit mass at the point y (infinite density) appears. We know we can model this kind of source by means of a Dirac measure at y, that has to be time dependent since the mass appears only at time s. We can write it in the form

$$\delta(x - y, t - s).$$

Thus, we are lead to the non homogeneous equation

$$p_t - D p_{xx} = \delta(x - y, t - s)$$

with $p(x, 0) = 0$ as initial condition. What could be the solution? Until $t = s$ nothing happens and *after* s we have $\delta(x - y, t - s) = 0$. Therefore it is like starting from time $t = s$ and solving the problem

$$p_t - D p_{xx} = 0, \quad x \in \mathbb{R}, \, t > s$$

with initial condition

$$p(x, s) = \delta(x - y, t - s).$$

We have solved this problem when $s = 0$; the solution is $\Gamma_D(x - y, t)$. By the time translation invariance of the diffusion equation, we deduce that the solution for any $s > 0$ is given by

$$p(x, t) = \Gamma_D(x - y, t - s). \tag{2.133}$$

Consider now a distributed source on the half-plane $t > 0$, capable to produce mass density at the time rate $f(x,t)$. Precisely, $f(x,t)\,dxdt$ is the mass produced[45] between x and $x+dx$, over the time interval $(t,t+dt)$. If initially no mass is present, we are lead to the *non homogeneous Cauchy problem*

$$\begin{cases} v_t - Dv_{xx} = f(x,t) & \text{in } \mathbb{R}\times(0,T) \\ v(x,0) = 0 & \text{in } \mathbb{R}. \end{cases} \tag{2.134}$$

As in subsection 2.8.1, we motivate the form of the solution at the point (x,t) using heuristic considerations. Let us compute the contribution dv to $v(x,t)$ of a mass $f(y,s)\,dyds$. It is like having a source term of the form

$$f^*(x,t) = f(x,t)\,\delta(x-y,t-s)$$

and therefore, recalling (2.133), we have

$$dv(x,t) = \Gamma_D(x-y,t-s)\,f(y,s)\,dyds. \tag{2.135}$$

We obtain the solution $v(x,t)$ by superposition, summing all the contributions (2.135). We split it into the following two steps:

• we sum over y the contributions for fixed s, to get the total density at (x,t), due to the diffusion of mass produced at time s. The result is $w(x,t,s)\,ds$, where

$$w(x,t,s) = \int_{\mathbb{R}} \Gamma_D(x-y,t-s)\,f(y,s)\,dy. \tag{2.136}$$

• we sum the above contributions for s ranging from 0 to t:

$$v(x,t) = \int_0^t \int_{\mathbb{R}} \Gamma_D(x-y,t-s)\,f(y,s)\,dyds.$$

The above construction is an example of application of the *Duhamel method*, that we state below:

Duhamel's method. *The procedure to solve problem* (2.134) *consists in the following two steps:*

1. *Construct a family of solutions of homogeneous Cauchy problems, with variable initial time $s > 0$, and initial data $f(x,s)$.*

2. *Integrate the above family with respect to s, over $(0,t)$.*

Indeed, let us examine the two steps.

1. Consider the homogeneous Cauchy problems

$$\begin{cases} w_t - Dw_{xx} = 0 & x \in \mathbb{R}, \ t > s \\ w(x,s,s) = f(x,s) & x \in \mathbb{R} \end{cases} \tag{2.137}$$

where the initial time s plays the role of a parameter.

---
[45] Negative production ($f < 0$) means removal.

The function $\Gamma^{y,s}(x, t) = \Gamma_D(x - y, t - s)$ is the fundamental solution of the diffusion equation that satisfies for $t = s$, the initial condition

$$\Gamma^{y,s}(x, s) = \delta(x - y).$$

Hence, the solution of (2.137) is given by the function (2.136):

$$w(x, t, s) = \int_{\mathbb{R}} \Gamma_D(x - y, t - s) f(y, s)\, dy.$$

Thus, $w(x, t, s)$ is the required family.

2. Integrating w over $(0, t)$ with respect to s, we find

$$v(x, t) = \int_0^t w(x, t, s)\, ds = \int_0^t \int_{\mathbb{R}} \Gamma_D(x - y, t - s) f(y, s)\, dy ds. \qquad (2.138)$$

Using (2.137) we have

$$v_t - D v_{xx} = w(x, t, t) + \int_0^t [w_t(x, t, s) - D w_{xx}(x, t, s)] = f(x, t).$$

Moreover, $v(x, 0) = 0$ and therefore v is a solution to (2.134).

Everything works under rather mild hypotheses on f. More precisely:

Theorem 2.4. *If f and its derivatives f_t, f_x, f_{xx} are continuous and bounded in $\mathbb{R} \times [0, T)$, then (2.138) gives a solution v of problem (2.134) in $\mathbb{R} \times (0, T)$, continuous up to $t = 0$, with derivatives v_t, v_x, v_{xx} continuous in $\mathbb{R} \times (0, T)$.*

The formula for the general Cauchy problem

$$\begin{cases} u_t - D u_{xx} = f(x, t) & \text{in } \mathbb{R} \times (0, T) \\ u(x, 0) = g(x) & \text{in } \mathbb{R} \end{cases} \qquad (2.139)$$

is obtained by superposition of (2.130) and (2.134):

$$u(x, t) = \int_{\mathbb{R}} \Gamma_D(x - y, t) g(y)\, dy + \int_0^t \int_{\mathbb{R}} \Gamma(x - y, t - s) f(y, s)\, dy ds \qquad (2.140)$$

Under the hypotheses on f and g stated in Theorems 2.3 and 2.4, (2.140) is a solution of (2.139) in $\mathbb{R} \times (0, T)$,

$$T < \frac{1}{4Da},$$

continuous with its derivatives u_t, u_x, u_{xx}.

The initial condition means that $u(x, t) \to g(x_0)$ as $(x, t) \to (x_0, 0)$ at any point x_0 of continuity of g. In particular, if g is continuous in $\mathbb{R}$ then u is continuous in $\mathbb{R} \times [0, T)$.

2.8.4 Maximum principles and uniqueness

The uniqueness of the solution to the global Cauchy problem is still to be discussed. This is not a trivial question since the following counterexample of Tychonov shows that there could be several solutions of the homogeneous problem. Let

$$h\left(t\right) = \begin{cases} e^{-t^2} & \text{for } t > 0 \\ 0 & \text{for } t \le 0. \end{cases}$$

It can be checked[46] that the function

$$\mathcal{T}\left(x,t\right) = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \frac{d^k}{dt^k} h\left(t\right)$$

is a solution of

$$u_t - u_{xx} = 0 \quad \text{in } \mathbb{R} \times (0, +\infty)$$

with

$$u\left(x, 0\right) = 0 \quad \text{in } \mathbb{R}.$$

Since also $u\left(x,t\right) \equiv 0$ is a solution of the same problem, we conclude that, in general, the Cauchy problem *is not well posed.*

What is wrong with $\mathcal{T}$? It grows too much at infinity for small times. Indeed the best estimate available for $\mathcal{T}$ is the following:

$$|\mathcal{T}\left(x,t\right)| \le C \exp\left\{\frac{x^2}{\theta t}\right\} \qquad (\theta > 0)$$

that quickly deteriorates when $t \to 0^+$, due to the factor $1/\theta t$.

If instead of $1/\theta t$ we had a constant A, as in condition *iii*) of Theorem 2.3, then we can assure uniqueness.

In other words, among the class of functions with growth at infinity controlled by an exponential of the type Ce^{Ax^2} for any $t \ge 0$ (the so called *Tychonov class*), the solution to the homogeneous Cauchy problem is unique.

This is a consequence of the following maximum principle.

Theorem 2.5 (Global maximum principle). *Let z be continuous in $\mathbb{R} \times [0, T]$, with derivatives z_x, z_{xx}, z_t continuous in $\mathbb{R} \times (0, T)$, such that, in $\mathbb{R} \times (0, T)$:*

$$z_t - Dz_{xx} \le 0 \qquad (resp. \ \ge 0)$$

and

$$z\left(x,t\right) \le Ce^{Ax^2}, \qquad \left(resp. \ \ge -Ce^{Ax^2}\right) \tag{2.141}$$

where $C > 0$. Then

$$\sup_{\mathbb{R} \times [0,T]} z\left(x,t\right) \le \sup_{\mathbb{R}} z\left(x,0\right) \qquad \left(resp. \ \inf_{\mathbb{R} \times [0,T]} z\left(x,t\right) \ge \inf_{\mathbb{R}} z\left(x,0\right)\right).$$

---

[46] Not an easy task! See *John's* book in the references.

The proof is rather difficult, but if we assume that z is bounded from above or below ($A = 0$ in (2.141)), then the proof relies on a simple application of the weak maximum principle, Theorem 2.2.

In Problem 2.13 we ask the reader to fill in the details of the proof.

We now are in position to prove the following uniqueness result.

Corollary 2.2. Uniqueness I. *Suppose u is a solution of*

$$\begin{cases} u_t - Du_{xx} = 0 \ in \ \mathbb{R} \times (0, T) \\ u(x, 0) = 0 \quad in \ \mathbb{R}, \end{cases}$$

continuous in $\mathbb{R} \times [0, T]$, with derivatives u_x, u_{xx}, u_t continuous in $\mathbb{R} \times (0, T)$. If $|u|$ satisfies (2.141) then $u \equiv 0$.

Proof. From Theorem 2.5 we have

$$0 = \inf_{\mathbb{R}} u(x, 0) \leq \inf_{\mathbb{R} \times [0,T]} u(x, t) \leq \sup_{\mathbb{R} \times [0,T]} u(x, t) \leq \sup_{\mathbb{R}} u(x, 0) = 0$$

so that $u \equiv 0$. $\square$

Notice that if

$$|g(x)| \leq ce^{ax^2} \quad \text{for every } x \in \mathbb{R} \quad (c, a \quad \text{positive}), \tag{2.142}$$

we know from Theorem 2.3 that

$$u(x, t) = \int_{\mathbb{R}} \Gamma_D(x - y, t) \, g(y) \, dy$$

satisfies the estimate

$$|u(x, t)| \leq Ce^{Ax^2} \quad \text{in } \mathbb{R} \times (0, T) \tag{2.143}$$

and therefore it belongs to the Tychonov class in $\mathbb{R} \times (0, T)$, for $T < 1/4Da$.

Moreover, if f is as in Theorem 2.4 and

$$v(x, t) = \int_0^t \int_{\mathbb{R}} \Gamma_D(x - y, t - s) \, f(y, s) \, dyds,$$

we easily get the estimate

$$t \inf_{\mathbb{R}} f \leq v(x, t) \leq t \sup_{\mathbb{R}} f, \tag{2.144}$$

for every $x \in \mathbb{R}$, $0 \leq t \leq T$. In fact:

$$v(x, t) \leq \sup_{\mathbb{R}} f \int_0^t \int_{\mathbb{R}} \Gamma_D(x - y, t - s) \, dyds = t \sup_{\mathbb{R}} f$$

since

$$\int_{\mathbb{R}} \Gamma_D(x - y, t - s) \, dy = 1$$

for every $x, t, s, t > s$. In the same way it can be shown that $v(x, t) \geq t \inf_{\mathbb{R}} f$. As a consequence, we have:

Corollary 2.3. Uniqueness II. *Let g be continuous in $\mathbb{R}$, satisfying (2.143), and let f be as in Theorem 2.4. Then the Cauchy problem (2.139) has a unique solution u in $\mathbb{R} \times (0,T)$ for $T < 1/4Da$, belonging to the Tychonov class. This solution is given by (2.140) and moreover*

$$\inf_{\mathbb{R}} g + t \inf_{\mathbb{R}} f \leq u(x,t) \leq \sup_{\mathbb{R}} g + t \sup_{\mathbb{R}} f. \qquad (2.145)$$

Proof. If u and v are solutions of the same Cauchy problem (2.139), then $w = u - v$ is a solution of (2.139) with $f = g = 0$ and satisfies the hypotheses of Corollary 2.2. It follows that $w(x,t) \equiv 0$. $\square$

• *Stability and comparison.* As in Corollary 2.1, inequality (2.145) is a stability estimate for the correspondence

$$data \longmapsto solution.$$

Indeed, let u_1 and u_2 be solutions of (2.139) with data g_1, f_1 and g_2, f_2, respectively. Under the hypotheses of Corollary 2.2, from (2.145) we can write

$$\sup_{\mathbb{R} \times [0,T]} |u_1 - u_2| \leq \sup_{\mathbb{R}} |g_1 - g_2| + T \sup_{\mathbb{R} \times [0,T]} |f_1 - f_2|.$$

Therefore if

$$\sup_{\mathbb{R} \times [0,T]} |f_1 - f_2| \leq \varepsilon, \quad \sup_{\mathbb{R}} |g_1 - g_2| \leq \varepsilon$$

also

$$\sup_{\mathbb{R} \times [0,T]} |u_1 - u_2| \leq \varepsilon (1 + T)$$

that means *uniform pointwise stability.*

This is not the only consequence of (2.145). We can use it to compare two solutions. For instance, from the left inequality we immediately deduce that if $f \geq 0$ and $g \geq 0$, also $u \geq 0$.

Similarly, if $f_1 \geq f_2$ and $g_1 \geq g_2$, then

$$u_1 \geq u_2.$$

• *Backward equations* arise in several applied contexts, from *control theory* and *dynamic programming* to *probability* and *finance*. An example is the celebrated *Black–Scholes equation* we will present in the next section.

Due to the time irreversibility, to have a well posed problem for the backward equation in the time interval $[0,T]$ we must prescribe a *final condition*, that is for $t = T$, rather than an initial one. On the other hand, the change of variable $t \longmapsto T - t$ transforms the backward into the forward equation, so that, from the mathematical point of view, the two equations are equivalent. Except for this remark the theory we have developed so far remains valid.

2.9 An Application to Finance

2.9.1 European options

In this section we apply the above theory to determine the price of some financial products, in particular of some *derivative* products, called *European options.*

A financial product is a *derivative* if its payoff depends on the price behavior of an asset, in jargon *the underlying*, for instance a *stock,* a *currency* or a *commodity.*

Among the simplest derivatives are the **European call** and **put options**, that are contracts on a prescribed asset between a *holder* and a *subscriber*, with the following rules.

At the drawing up time of the contract (say at time $t = 0$) an **exercise** or **strike price** E is fixed.

<div style="text-align:center">

At an **expiry date** T, fixed in the future,
</div>

- the holder of a call option **can (but is not obliged to)** exercise the option by **purchasing** the asset at the price E. If the holder decides to buy the asset, the subscriber **must** sell it;

- the holder of a put option **can (but is not obliged to)** exercise the option by **selling** at the price E. If the holder decides to sell the asset, the subscriber **must** buy it.

Since an option gives to the holder a right without any obligation, the option has a price and the basic question is: **what is the "right" price that must be paid** at $t = 0$?

This price certainly depends on the evolution of the price S of the underlying, on the strike price E, on the expiring time T and on the current riskless interest rate $r > 0$.

For instance, for a call, to a lower E corresponds a greater price; the opposite holds for a put. The price fluctuations of the underlying affect in crucial way the value of an option, since they incorporate the amount of risk.

To answer our basic question, we introduce the **value function** $V = V(S, t)$, giving the proper price of the option if at time t the price of the underlying is S. What we need to know is $V(S(0), 0)$. When we like to distinguish between **call** and **put**, we use the notations $C(S, t)$ and $P(S, t)$, respectively.

The problem is then to determine V in agreement with the financial market, where both the underlying and the option are exchanged. We shall use the Black-Scholes method, based on the assumption of a reasonable evolution model for S and on the fundamental principle of *no arbitrage possibilities*.

2.9.2 An evolution model for the price S

Since S depends on more or less foreseeable factors, it is clear that we cannot expect a deterministic model for the evolution of S. To construct it we assume a *market efficiency* in the following sense:

a) The market responds instantaneously to new information on the asset.

b) The price has no memory: its past history is fully stored in the present price, without further information.

Condition a) implies the adoption of a continuous model. Condition b) basically requires that a change dS of the underlying price has the Markov property, like Brownian motion.

Consider now a time interval from t to $t + dt$, during which S undergoes a change from S to $S + dS$. One of the most common models assumes that the **return** dS/S is given by the sum of two terms.

One is a deterministic term, which gives a contribution μdt due to a constant *drift* μ, representing the average growth rate of S. With this term alone, we would have

$$\frac{dS}{S} = \mu dt$$

and therefore $d \log S = \mu dt$, that gives the exponential growth $S(t) = S(0) e^{\mu t}$.

The other term is stochastic and takes into account the random aspects of the evolution. It gives the contribution

$$\sigma dB$$

where dB is an increment of a Brownian motion and has zero mean and variance dt. The coefficient σ, that we assume to be constant, is called the **volatility** and measures the standard deviation of the return.

Summing the contributions we have

$$\frac{dS}{S} = \mu dt + \sigma dB. \tag{2.146}$$

Note the physical dimensions of μ and σ: $[\mu] = [time]^{-1}$, $[\sigma] = [time]^{-\frac{1}{2}}$.

The (2.146) is a **stochastic differential equation** (*s.d.e.*). To solve it one is tempted to write

$$d \log S = \mu dt + \sigma dB,$$

to integrate between 0 e t, and to obtain

$$\log \frac{S(t)}{S(0)} = \mu t + \sigma (B(t) - B(0)) = \mu t + \sigma B(t)$$

since $B(0) = 0$. However, this is not correct. The diffusion term σdB requires the use of the **Itô formula**, a stochastic version of the chain rule. Let us make a few intuitive remarks on this important formula.

Digression on Itô's formula. Let $B = B(t)$ the usual Brownian motion. An Itô process $X = X(t)$ is a solution of a *s.d.e.* of the type

$$dX = a(X, t) dt + \sigma (X, t) dB \tag{2.147}$$

where a is the *drift term* and σ is the *volatility coefficient*.

When $\sigma = 0$, the equation is deterministic and the trajectories can be computed with the usual analytic methods. Moreover, given a smooth function $F = F(x, t)$, we can easily compute the variation of F along those trajectories. It is enough to compute

$$dF = F_t dt + F_x dX = \{F_t + aF_x\} dt.$$

Let now be σ non zero; the preceding computation would give

$$dF = F_t dt + F_x dX = \{F_t + aF_x\} dt + \sigma F_x dB$$

but **this formula does not give the complete differential of** F. Indeed, using Taylor's formula, one has, letting $X(0) = X_0$:

$$F(X, t) = F(X_0, 0) + F_t dt + F_x dX + \frac{1}{2} \left\{ F_{xx} (dX)^2 + 2F_{xt} dX dt + F_{tt} (dt)^2 \right\} + ...$$

The differential of F along the trajectories of (2.147) is obtained by selecting in the right hand side of the preceding formula the terms which are **linear** with respect to dt or dX. We first find the terms

$$F_t dt + F_x dX = \{F_t + aF_x\} dt + \sigma F_x dB.$$

The terms $2F_{xt} dX dt$ and $F_{tt} (dt)^2$ are non linear with respect to dt and dX and therefore they are not in the differential. Let us now check the term $(dX)^2$. We have

$$(dX)^2 = [adt + \sigma dB]^2 = a^2 (dt)^2 + 2a\sigma dB dt + \boxed{\sigma^2 (dB)^2}.$$

While $a^2 (dt)^2$ and $2a\sigma dB dt$ are non linear with respect to dt and dX, the framed term **turns out to be exactly**

$$\sigma^2 dt.$$

Formally, this is a consequence of the basic formula[47] $dB \sim \sqrt{dt} N(0, 1)$ that assigns $\sqrt{dt}$ for the standard deviation of dB.

Thus the differential of F along the trajectories of (2.147) is given by the following **Itô formula**:

$$dF = \left\{ F_t + aF_x + \frac{1}{2}\sigma^2 F_{xx} \right\} dt + \sigma F_x dB. \tag{2.148}$$

We are now ready to solve (2.146), that we write in the form

$$dS = \mu S dt + \sigma S dB.$$

Let $F(S) = \log S$. Since

$$F_t = 0, \qquad F_S = \frac{1}{S}, \qquad F_{ss} = -\frac{1}{S^2}$$

---

[47] See (2.86), subsection 2.4.3.

Itô's formula gives, with $X = S$, $a(S, t) = \mu S$, $\sigma(S, t) = \sigma S$,

$$d \log S = \left(\mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dB.$$

We can now integrate between 0 and t, obtaining

$$\log S(t) = \log S_0 + \left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma B(t). \tag{2.149}$$

The (2.149) shows that the random variable $Y = \log S$ has a normal distribution, with mean $\log S_0 + \left(\mu - \frac{1}{2} \sigma^2 \right) t$ and variance $\sigma^2 t$. Its probability density is therefore

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp \left\{ - \frac{\left(y - \log S_0 - \left(\mu - \frac{1}{2}\sigma^2 \right) t \right)^2}{2\sigma^2 t} \right\}.$$

and the density of S is given by

$$p(s) = \frac{1}{s} f(\log s) = \frac{1}{s\sqrt{2\pi\sigma^2 t}} \left\{ - \frac{\left(\log s - \log S_0 - \left(\mu - \frac{1}{2}\sigma^2 \right) t \right)^2}{2\sigma^2 t} \right\}$$

which is called a **lognormal density**.

2.9.3 The Black-Scholes equation

We now construct a differential equation able to describe the evolution of $V(S, t)$. We work under the following hypotheses:

- S follows a **lognormal law.**
- The volatility σ is constant and known.
- There are no transaction costs or dividends.
- It is possible to buy or sell any number of the underlying asset.
- There is an interest rate $r > 0$, for a riskless investment. This means that 1 dollar in a bank at time $t = 0$ becomes e^{rT} dollars at time T.
- The market is **arbitrage free.**

The last hypothesis is crucial in the construction of the model and means that *there is no opportunity for instantaneous risk-free profit.* It could be considered as a sort of conservation law for money!

    The translation of this principle into mathematical terms is linked with the notion of *hedging* and the existence of *self-financing portfolios*[48]. The basic idea is first to compute the return of V through Itô formula and then to construct a riskless portfolio Π, consisting of shares of S and the option. By the arbitrage free hypothesis, Π must grow at the current interest rate r, i.e. $d\Pi = r\Pi dt$, which turns out to coincide with the fundamental Black-Scholes equation.

---

[48] A *portfolio* is a collection of *securities* (e.g. stocks) holdings.

Let us then use the Itô formula to compute the differential of V. Since

$$dS = \mu S dt + \sigma S dB,$$

we find

$$dV = \left\{ V_t + \mu S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS} \right\} dt + \sigma S V_S dB. \qquad (2.150)$$

Now we try to get rid of the risk term $\sigma S V_S dB$ by constructing a portfolio Π, consisting of the option and a quantity[49] $-\Delta$ of underlying:

$$\Pi = V - S\Delta.$$

This is an important financial operation called *hedging*. Consider now the interval of time $(t, t+dt)$ during which Π undergoes a variation $d\Pi$. If we manage to keep Δ equal to its value at t during the interval $(t, t+dt)$, the variation of Π is given by

$$d\Pi = dV - \Delta dS.$$

This is a key point in the whole construction, that needs to be carefully justified[50]. Although we content ourselves with an intuitive level, we will come back to this question in the last section of this chapter.

Using (2.150) we find

$$d\Pi = dV - \Delta dS = \qquad\qquad (2.151)$$
$$= \left\{ V_t + \mu S V_s + \frac{1}{2}\sigma^2 S^2 V_{SS} - \mu S\Delta \right\} dt + \sigma S(V_S - \Delta) dB.$$

Thus, if we choose

$$\Delta = V_S, \qquad\qquad (2.152)$$

meaning that Δ is the value of V_S at t, we eliminate the stochastic component in (2.151). The evolution of the portfolio Π is now entirely deterministic and its dynamics is given by the following equation:

$$d\Pi = \left\{ V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} \right\} dt. \qquad\qquad (2.153)$$

The choice (2.152) appears almost .... miraculous, but it is partly justified by the fact that V and S are dependent and the random component in their dynamics is proportional to S. Thus, in a suitable linear combination of V and S such component should disappear.

It is the moment to use the no-arbitrage principle. Investing Π at the riskless rate r, after a time dt we have an increment $r\Pi dt$. Compare $r\Pi dt$ with $d\Pi$ given by (2.153).

---

[49] We borrow from *finance* the use of the greek letter Δ in this context. Clearly here it has nothing to do with the Laplace operator.

[50] In fact, saying that we keep Δ constant for an infinitesimal time interval so that we can cancel $Sd\Delta$ from the differential $d\Pi$ requires a certain amount of impudence....

· If $d\Pi > r\Pi dt$, we borrow an amount Π to invest in the portfolio. The return $d\Pi$ would be greater of the cost $r\Pi dt$, so that we make an instantaneous riskless profit

$$d\Pi - r\Pi dt.$$

· If $d\Pi < r\Pi dt$, we sell the portfolio Π investing it in a bank at the rate r. This time we would make an instantaneous risk free profit

$$r\Pi dt - d\Pi.$$

Therefore, the arbitrage free hypothesis forces

$$d\Pi = \left\{ V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} \right\} dt = r\Pi dt. \tag{2.154}$$

Substituting

$$\Pi = V - S\Delta = V - V_S S$$

into (2.154), we obtain the celebrated **Black-Scholes equation**:

$$\mathcal{L}V = V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = 0. \tag{2.155}$$

Note that the coefficient μ, the drift of S, does not appear in (2.155). This fact is apparently counter-intuitive and shows an interesting aspect of the model. The financial meaning of the Black-Scholes equation is emphasized from the following decomposition of its right hand side:

$$\mathcal{L}V = \underbrace{V_t + \frac{1}{2}\sigma^2 S^2 V_{SS}}_{\text{portfolio return}} - \underbrace{r(V - SV_S)}_{\text{bank investment}}.$$

The Black-Scholes equation is a little more general than the equations we have seen so far. Indeed, the diffusion and the drift coefficients are both depending on S. However, as we shall see below, we can transform it into the diffusion equation $u_t = u_{xx}$.

Observe that the coefficient of V_{SS} is positive, so that (2.155) is a **backward equation**. To get a well posed problem, we need a **final condition** (at $t = T$), a side condition at $S = 0$ and one condition for $S \to +\infty$.

• *Final conditions.* We examine what conditions we have to impose at $t = T$.

Call. If at time T we have $S > E$ then we exercise the option, with a profit $S - E$. If $S \le E$, we do not exercise the option with no profit. The *final payoff* of the option is therefore

$$C(S,T) = \max\{S - E, 0\} = (S - E)^+, \qquad S > 0.$$

Put. If at time T we have $S \ge E$, we do not exercise the option, while we exercise the option if $S < E$. The *final payoff* of the option is therefore

$$P(S,T) = \max\{E - S, 0\} = (E - S)^+, \qquad S > 0.$$

• *Boundary conditions.* We now examine the conditions to be imposed at $S = 0$ and for $S \to +\infty$.

Call. If $S = 0$ at a time t, (2.146) implies $S = 0$ thereafter, and the option has no value; therefore

$$C(0, t) = 0 \qquad t \geq 0.$$

As $S \to +\infty$, at time t, the option will be exercised and its value becomes practically equal to S minus the discounted exercise price, that is

$$C(S, t) - (S - e^{-r(T-t)}E) \to 0 \qquad \text{as } S \to \infty.$$

Put. If at a certain time is $S = 0$, so that $S = 0$ thereafter, the final profit is E. Thus, to determine $P(0, t)$ we need to determine the present value of E at time T, that is

$$P(0, t) = Ee^{-r(T-t)}.$$

If $S \to +\infty$, we do not exercise the option, hence

$$P(S, t) = 0 \qquad \text{as } S \to +\infty.$$

2.9.4 The solutions

Let us summarize our model in the two cases.

Black-Scholes equation

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + rSV_S - rV = 0. \tag{2.156}$$

Final payoffs

$$
\begin{aligned}
C(S, T) &= (S - E)^+ &\text{(call)} \\
P(S, T) &= (E - S)^+ &\text{(put)}.
\end{aligned}
$$

Boundary conditions

$$
\begin{aligned}
&C(0, t) = 0, \quad C(S, t) - (S - e^{-r(T-t)}E) \to 0 \quad \text{as } S \to \infty &\text{(call)} \\
&P(0, t) = Ee^{-r(T-t)}, \quad P(S, T) = 0 \quad \text{as } S \to \infty &\text{(put)}.
\end{aligned}
$$

It turns out that the above problems can be reduced to a global Cauchy problem for the heat equation. In this way it is possible to find explicit formulas for the solutions. First of all we make a change of variables to reduce the Black-Scholes equation to constant coefficients and to pass from backward to forward in time. Also note that $1/\sigma^2$ can be considered an intrinsic reference time while the exercise price E gives a characteristic order of magnitude for S and V. Thus, $1/\sigma^2$ and E can be used as rescaling factors to introduce dimensionless variables.

Let us set

$$x = \log \frac{S}{E}, \quad \tau = \frac{1}{2}\sigma^2(T - t), \quad w(x, \tau) = \frac{1}{E}V\left(Ee^x, T - \frac{2\tau}{\sigma^2}\right).$$

When S goes from 0 to $+\infty$, x varies from $-\infty$ to $+\infty$. When $t = T$ we have $\tau = 0$. Moreover:

$$V_t = -\frac{1}{2}\sigma^2 E w_\tau$$

$$V_S = \frac{E}{S}w_x, \quad V_{SS} = -\frac{E}{S^2}w_x + \frac{E}{S^2}w_{xx}.$$

Substituting into (2.156), after some simplifications, we get

$$-\frac{1}{2}\sigma^2 w_\tau + \frac{1}{2}\sigma^2\left(-w_x + w_{xx}\right) + rw_x - rw = 0$$

or

$$w_\tau = w_{xx} + (k-1)w_x - kw$$

where $k = \frac{2r}{\sigma^2}$ is a dimensionless parameter. By further setting[51]

$$w(x,\tau) = e^{-\frac{k-1}{2}x - \frac{(k+1)^2}{4}\tau}v(x,\tau)$$

we find that v satisfies

$$v_\tau - v_{xx} = 0, \qquad -\infty < x < +\infty, \ 0 \le \tau \le T.$$

The final condition for V becomes an initial condition for v. Precisely, after some manipulations, we have

$$v(x,0) = g(x) = \begin{cases} e^{\frac{1}{2}(k+1)x} - e^{\frac{1}{2}(k-1)x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

for the call option, and

$$v(x,0) = g(x) = \begin{cases} e^{\frac{1}{2}(k-1)x} - e^{\frac{1}{2}(k+1)x} & x < 0 \\ 0 & x \ge 0 \end{cases}$$

for the put option.

Now we can use the preceding theory and in particular Theorem 2.3 and Corollary 2.3. The solution is unique and it is given by formula

$$v(x,\tau) = \frac{1}{\sqrt{4\pi\tau}}\int_{\mathbb{R}}g(y)\,e^{-\frac{(x-y)^2}{4\tau}}\,dy.$$

To have a more significant formula, let $y = \sqrt{2\tau}z + x$; then, focusing on the **call** option:

$$v(x,\tau) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}g\left(\sqrt{2\tau}z + x\right)e^{-\frac{z^2}{2}}\,dy =$$

$$= \frac{1}{\sqrt{2\pi}}\left\{\int_{-x/\sqrt{2\tau}}^{\infty}e^{\frac{1}{2}(k+1)\left(\sqrt{2\tau}z+x\right)-\frac{1}{2}z^2}\,dz - \int_{-x/\sqrt{2\tau}}^{\infty}e^{\frac{1}{2}(k-1)\left(\sqrt{2\tau}z+x\right)-\frac{1}{2}z^2}\,dz\right\}.$$

---

[51] See Problem 2.14.

After some manipulations in the two integrals[52], we obtain

$$v\left(x,\tau\right) = e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}N\left(d_+\right) - e^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau}N\left(d_-\right)$$

where

$$N\left(z\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{1}{2}y^2}dy$$

is the distribution of a standard normal random variable and

$$d_\pm = \frac{x}{\sqrt{2\tau}} + \frac{1}{2}\left(k\pm 1\right)\sqrt{2\tau}.$$

Going back to the original variables we have, for the **call:**

$$C\left(S,t\right) = SN\left(d_+\right) - Ee^{-r(T-t)}N\left(d_-\right)$$

with

$$d_\pm = \frac{\log\left(S/E\right) + \left(r\pm\frac{1}{2}\sigma^2\right)\left(T-t\right)}{\sigma\sqrt{T-t}}.$$

The formula for the **put** is

$$P\left(S,t\right) = Ee^{-r(T-t)}N\left(-d_-\right) - SN\left(-d_+\right).$$

It can be shown that[53]

$$\Delta = C_S = N\left(d_+\right) > 0 \qquad \text{for the \textbf{call}}$$
$$\Delta = P_S = N\left(d_+\right) - 1 < 0 \qquad \text{for the \textbf{put.}}$$

Note that C_S e P_S are strictly increasing with respect to S, since N is a strictly increasing function and d_+ is strictly increasing with S. The functions C, P are therefore *strictly convex functions* of S, for every t, namely $C_{ss} > 0$ and $P_{ss} > 0$.

• *Put-call parity*. Put and call options with the same exercise price and expiry time can be connected by forming the following portfolio:

$$\Pi = S + P - C$$

---

[52] For instance, to evaluate the first integral, complete the square at the exponent, writing

$$\frac{1}{2}\left(k+1\right)\left(\sqrt{2\tau}z+x\right) - \frac{1}{2}z^2 = \frac{1}{2}\left(k+1\right)x + \frac{1}{4}\left(k+1\right)^2\tau - \frac{1}{2}\left[z - \frac{1}{2}\left(k+1\right)\sqrt{2\tau}\right]^2.$$

Then, setting $y = \frac{1}{2}\left(k+1\right)\sqrt{2\tau}$,

$$\int_{-x/\sqrt{2\tau}}^{\infty}e^{\frac{1}{2}(k+1)(\sqrt{2\tau}z+x)-\frac{1}{2}z^2}dz = e^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\int_{-x/\sqrt{2\tau}-(k+1)\sqrt{\tau}/\sqrt{2}}^{\infty}e^{-\frac{1}{2}y^2}dz.$$

[53] The calculations are rather ... painful.

where the minus in front of C shows a so called *short position* (negative holding). For this portfolio the final payoff is

$$\Pi\left(S, T\right) = S + \left(E - S\right)^{+} - \left(S - E\right)^{+}.$$

If $E \geq S$, we have

$$\Pi\left(S, T\right) = S + \left(E - S\right) - 0 = E$$

while if $E \leq S$,

$$\Pi\left(S, T\right) = S + 0 - \left(S - E\right) = E.$$

Thus at expiry the payoff is always equal to E and it constitutes a riskless profit, whose value at t must be equal to the discounted value of E, because of the no arbitrage condition. Hence we find the following relation (*put–call parity*)

$$S + P - C = Ee^{-r(T-t)}. \tag{2.157}$$

Formula (2.157) also shows that, given the value of C (or P), we can find the value of P (or C).

From (2.157), since $Ee^{-r(T-t)} \leq E$ and $P \geq 0$, we get

$$C\left(S, t\right) = S + P - Ee^{-r(T-t)} \geq S - E$$

and therefore, since $C \geq 0$,

$$C\left(S, t\right) \geq \left(S - E\right)^{+}.$$

It follows that the value of C *is always greater than the final payoff*. It is not so for a put. In fact

$$P\left(0, t\right) = Ee^{-r(T-t)} \leq E$$

so that the value of P is below the final payoff when S is near 0, while it is above just before expiring. The figures 2.14 and 2.15 show the behavior of C and P versus S, for some values of $T - t$ up to expiry.
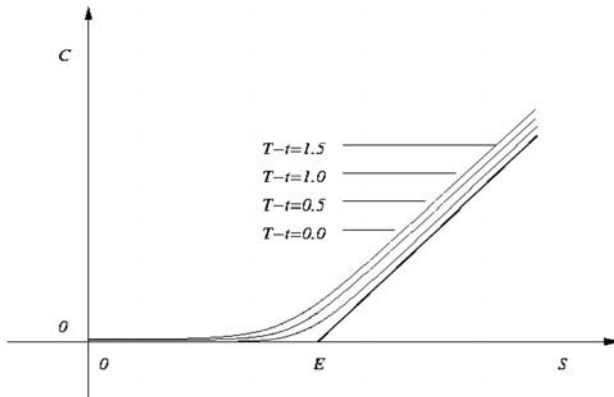
Fig. 2.14. The value function for an European call option
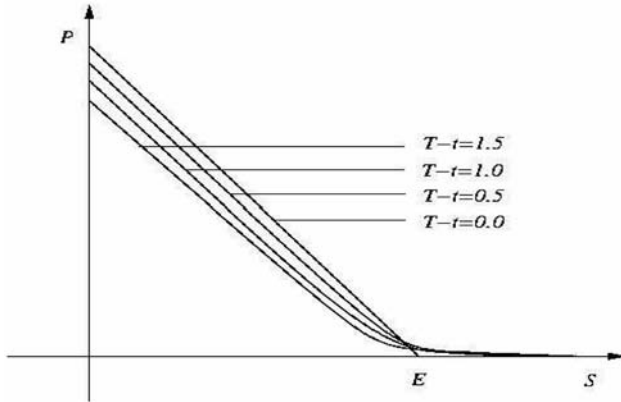
Fig. 2.15. The value function of an European put option

• *Different volatilities.* The maximum principle arguments in subsection 2.8.3 can be used to compare the value of two options with different volatilities σ_1 and σ_2, having the same exercise price E and the same strike time T. Assume that $\sigma_1 > \sigma_2$ and denote by $C^{(1)}$, $C^{(2)}$ the value of the corresponding call options. Diminishing the amount of risk the value of the option should decrease and indeed we want to confirm that

$$C^{(1)} > C^{(2)} \qquad S > 0, 0 \leq t < T.$$

Let $W = C^{(1)} - C^{(2)}$. Then

$$W_t + \frac{1}{2}\sigma_2^2 S^2 W_{SS} + rSW_S - rW = \frac{1}{2}(\sigma_2^2 - \sigma_1^2)S^2 C_{SS}^{(1)} \qquad (2.158)$$

with $W(S,T) = 0$, $W(0,t) = 0$ and $W \to 0$ as $S \to +\infty$.

The (2.158) is a nonhomogeneous equation, whose right hand side is *negative* for $S > 0$, because $C_{SS}^{(1)} > 0$. Since W is continuous in the half strip $[0, +\infty) \times [0, T]$ and vanishes at infinity, it attains its global minimum at a point (S_0, t_0).

We claim that the minimum is zero and cannot be attained at a point in $(0, +\infty) \times [0, T)$. Since the equation is backward, $t_0 = 0$ is excluded. Suppose $W(S_0, t_0) \leq 0$ with $S_0 > 0$ and $0 < t_0 < T$. We have

$$W_t(S_0, t_0) = 0$$

and

$$W_S(S_0, t_0) = 0, \qquad W_{SS}(S_0, t_0) \geq 0.$$

Substituting $S = S_0, t = t_0$ into (2.158) we get a contradiction. Therefore $W = C^{(1)} - C^{(2)} > 0$ for $S > 0, 0 < t < T$.

2.9.5 Hedging and self-financing strategy

The mathematical translation of the *no arbitrage* principle can be made more rigorously than we did in subsection 2.9.2, by introducing the concept of *self-financing* portfolio. The idea is to "duplicate" V by means of a portfolio consisting of a number of shares of S and a bond Z, a free risk investment growing at the rate r, e.g. $Z(t) = e^{rt}$.

To this purpose let us try to determine two processes $\phi = \phi(t)$ e $\psi = \psi(t)$ such that

$$V = \phi S + \psi Z \qquad (0 \leq t \leq T) \tag{2.159}$$

in order to eliminate any risk factor. In fact, playing the part of the subscriber (that has to sell), the risk is that at time T the price $S(T)$ is greater than E, so that the holder will exercise the option. If in the meantime the subscriber has constructed the portfolio (2.159), the profit from it exactly meets the funds necessary to pay the holder. On the other hand, if the option has zero value at time T, the portfolio has no value as well.

For the operation to make sense, it is necessary that the subscriber *does not put extra money in this strategy* (*hedging*). This can be assured by requiring that the portfolio (2.159) be *self-financing* that is, **its changes in value be dependent from variations of S and Z alone**.

In formulas, this amount to requiring

$$dV = \phi dS + \psi dZ \qquad (0 \leq t \leq T). \tag{2.160}$$

Actually, we have already met something like (2.160), when we have constructed the portfolio $\Pi = V - S\Delta$ or

$$V = \Pi + S\Delta,$$

asking that $dV = d\Pi + \Delta dS$. This construction is nothing else that a duplication of V by means of a *self-financing portfolio,* with Π playing the role of Z and choosing $\psi = 1$.

But, what is the real meaning of (2.160)? We see it better in a discrete setting. Consider a sequence of times

$$t_0 < t_1 < ... < t_N$$

and suppose that the intervals $(t_j - t_{j-1})$ are very small. Denote by S_j e Z_j the values at t_j of S and Z. Consequently, look for two sequences

$$\phi_j \text{ and } \psi_j$$

corresponding to the quantity of S and Z to be used in the construction of the portfolio (2.159) from t_{j-1} to t_j. Notice that ϕ_j and ψ_j are chosen at time t_{j-1}.

Thus, given the interval (t_{j-1}, t_j),

$$V_j = \phi_j S_j + \psi_j Z_j$$

represents the closing value of the portfolio while

$$\phi_{j+1}S_j + \psi_{j+1}Z_j$$

is the opening value, the amount of money necessary to buy the new one. The **self-financing condition means** that the value V_j of the portfolio at time t_j, determined by the couple (ϕ_j, ψ_j), exactly meets the purchasing cost of the portfolio in the interval (t_j, t_{j+1}), determined by (ϕ_{j+1}, ψ_{j+1}). This means

$$\phi_{j+1}S_j + \psi_{j+1}Z_j = \phi_j S_j + \psi_j Z_j \tag{2.161}$$

or that **the financial gap**

$$D_j = \phi_{j+1}S_j + \psi_{j+1}Z_j - V_j$$

must be zero, otherwise an amount of cash D_j has to be injected to substain the strategy $(D_j > 0)$ or the same amount of money can be drawn from it $(D_j < 0)$. From (2.161) we deduce that

$$\begin{aligned} V_{j+1} - V_j &= (\phi_{j+1}S_{j+1} + \psi_{j+1}Z_{j+1}) - (\phi_j S_j + \psi_j Z_j) \\ &= (\phi_{j+1}S_{j+1} + \psi_{j+1}Z_{j+1}) - (\phi_{j+1}S_j + \psi_{j+1}Z_j) \\ &= \phi_{j+1}(S_{j+1} - S_j) + \psi_{j+1}(Z_{j+1} - Z_j) \end{aligned}$$

or

$$\Delta V_j = \phi_{j+1}\Delta S_j + \psi_{j+1}\Delta Z_j$$

whose continuous version is exactly (2.160).

Going back to the continuous case, by combining formulas (2.150) and (2.160) for dV, we get

$$\left\{ V_t + \mu S V_S + \frac{1}{2}\sigma^2 S^2 V_{SS} \right\} dt + \sigma S V_S dB = \phi\left(\mu S dt + \sigma S dB\right) + \psi r Z dt.$$

Choosing $\phi = V_S$, we rediscover the Black and Scholes equation

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + r S V_S - r V = 0. \tag{2.162}$$

On the other hand, if V satisfies (2.162) and

$$\phi = V_S, \quad \psi = Z^{-1}(V - V_S S) = e^{-rt}(V - V_S S),$$

it can be proved that the self financing condition (2.160) is satisfied for the portfolio $\phi S + \psi Z$.

2.10 Some Nonlinear Aspects

All the mathematical models we have examined so far are *linear*. On the other hand, the nature of most real problems is nonlinear. For example, *nonlinear diffusion* has to be taken into account in filtration problems, *non linear drift* terms are quite important in fluid dynamics while *nonlinear reaction* terms occur frequently in population dynamics and kinetics chemistry.

The presence of a nonlinearity in a mathematical model gives rise to many interesting phenomena that cannot occur in the linear case; typical instances are finite speed of diffusion, finite time blow-up or existence of travelling wave solutions of certain special profiles, each one with its own characteristic velocity.

In this section we try to convey some intuition of what could happen in two typical and important examples from filtration through a porous medium and population dynamics. In Chapter 4, we shall deal with nonlinear transport models.

2.10.1 Nonlinear diffusion. The porous medium equation

Consider a gas of density $\rho = \rho(\mathbf{x}, t)$ flowing through a porous medium. Denote by $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$ the velocity of the gas and by κ the *porosity* of the medium, representing the volume fraction filled with gas. Conservation of mass reads, in this case:

$$\kappa \rho_t + \operatorname{div}(\rho \mathbf{v}) = 0. \tag{2.163}$$

Besides (2.163), the flow is governed by the two following constitutive (empirical) laws.

• **Darcy's law**:

$$\mathbf{v} = -\frac{\mu}{\nu} \nabla p \tag{2.164}$$

where $p = p(\mathbf{x}, t)$ is the pressure, μ is the *permeability* of the medium and ν is the *viscosity* of the gas. We assume μ and ν are positive constants.

• **Equation of state**:

$$p = p_0 \rho^\alpha \qquad p_0 > 0, \alpha > 0. \tag{2.165}$$

From (2.164) and (2.165) we have, since $p^{1/\alpha} \nabla p = (1 + 1/\alpha)^{-1} \Delta(p^{1+1/\alpha})$,

$$\operatorname{div}(\rho \mathbf{v}) = -\frac{\mu}{(1 + 1/\alpha)\nu p_0^{1/\alpha}} \Delta(p^{1+1/\alpha}) = -\frac{(m-1)\,\mu p_0}{m\nu} \Delta(\rho^m)$$

where $m = 1 + \alpha > 1$. From (2.163) we obtain

$$\rho_t = \frac{(m-1)\,\mu p_0}{\kappa m \nu} \Delta(\rho^m).$$

Rescaling time $(t \mapsto \dfrac{(m-1)\,\mu p_0}{\kappa m \nu} t)$ we finally get the **porous medium equation**

$$\rho_t = \Delta(\rho^m). \tag{2.166}$$

Since

$$\Delta(\rho^m) = \operatorname{div}\left(m\rho^{m-1}\nabla\rho\right)$$

we see that the diffusion coefficient is $D(\rho) = m\rho^{m-1}$, showing that the diffusive effect increases with the density.

The porous medium equation can be written in terms of the pressure variable

$$u = p/p_0 = \rho^{m-1}.$$

It is not difficult to check that the equation for u is given by

$$u_t = u\Delta u + \frac{m}{m-1}\,|\nabla u|^2 \tag{2.167}$$

showing once more the dependence on u of the diffusion coefficient.

One of the basic questions related to the equation (2.166) or (2.167) is to understand how an initial data ρ_0, confined in a small region Ω, evolves with time. The key object to examine is therefore the unknown boundary $\partial\Omega$, or *free boundary* of the gas, whose speed of expansion we expect to be proportional to $|\nabla u|$ (from (2.164)). This means that we expect a *finite speed of propagation,* in contrast with the classical case $m = 1$.

The porous media equation cannot be treated by elementary means, since at very low density the diffusion has a very low effect and the equation degenerates. However we can get some clue of what happens by examining a sort of fundamental solutions, the so called *Barenblatt solutions*, in spatial dimension 1.

The equation is

$$\rho_t = (\rho^m)_{xx}. \tag{2.168}$$

We look for *nonnegative self-similar* solutions of the form

$$\rho(x,t) = t^{-\alpha}U\left(xt^{-\beta}\right) \equiv t^{-\alpha}U(\xi)$$

satisfying

$$\int_{-\infty}^{+\infty} \rho(x,t)\,dx = 1.$$

This condition requires

$$1 = \int_{-\infty}^{+\infty} t^{-\alpha}U\left(xt^{-\beta}\right)dx = t^{\beta-\alpha}\int_{-\infty}^{+\infty}U(\xi)\,d\xi$$

so that we must have $\alpha = \beta$ and $\int_{-\infty}^{+\infty} U(\xi)\,d\xi = 1$. Substituting into (2.168), we find

$$\alpha t^{-\alpha-1}(-U - \xi U') = t^{-m\alpha-2\alpha}(U^m)''.$$

Thus, if we choose $\alpha = 1/(m+1)$, we get for U the differential equation

$$(m+1)(U^m)'' + \xi U' + U = 0$$

that can be written in the form

$$\frac{d}{d\xi}\left[(m+1)(U^m)' + \xi U\right] = 0.$$

Thus, we have

$$(m+1)(U^m)' + \xi U = \text{constant}.$$

Choosing the constant equal to zero, we get

$$(m+1)(U^m)' = (m+1)mU^{m-1}U' = -\xi U$$

or

$$(m+1)mU^{m-2}U' = -\xi.$$

This in turn is equivalent to

$$\frac{(m+1)m}{m-1}\left(U^{m-1}\right)' = -\xi$$

whose solution is

$$U(\xi) = \left[A - B_m\xi^2\right]^{1/(m-1)}$$

where A is an arbitrary constant and $B_m = (m-1)/2m(m+1)$. Clearly, to have a physical meaning, we must have $A > 0$ and $A - B_m\xi^2 \geq 0$.

In conclusion we have found solutions of the porous medium equation of the form

$$\rho(x,t) = \begin{cases} \dfrac{1}{t^\alpha}\left[A - B_m\dfrac{x^2}{t^{2\alpha}}\right]^{1/(m-1)} & \text{if } x^2 \leq At^{2\alpha}/B_m \\ 0 & \text{if } x^2 > At^{2\alpha}/B_m. \end{cases} \qquad (\alpha = 1/(m+1)).$$

known as *Barenblatt solutions*. The points

$$x = \pm\sqrt{A/B_m}\,t^\alpha \equiv \pm r(t)$$

represent the gas interface between the part filled by gas and the empty part. Its speed of propagation is therefore

$$\dot{r}(t) = \alpha\sqrt{A/B_m}\,t^{\alpha-1}.$$
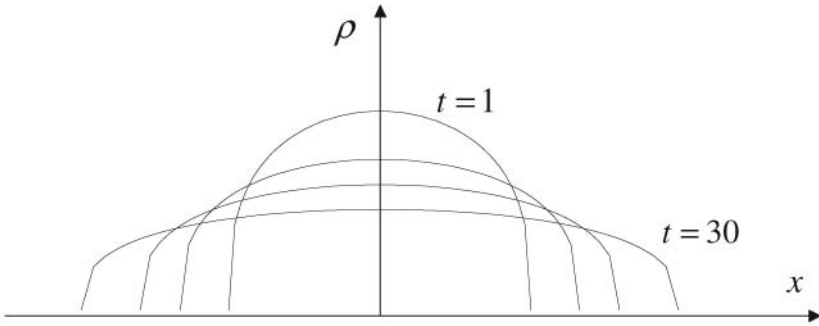
Fig. 2.16. The Barenblatt solution

$$\rho\left(x,t\right) = t^{-1/5}\left[1 - x^2 t^{-2/5}\right]_{+}^{1/3}$$

for $t = 1, 4, 10, 30$

2.10.2 Nonlinear reaction. Fischer's equation

In 1937 Fisher[54] introduced a model for the spatial spread of a so called *favoured*[55] (or *advantageous*) *gene in a population,* over an infinitely long one dimensional habitat. Denoting by v the gene concentration, Fisher's equation reads

$$v_\tau = Dv_{yy} + rv\left(1 - \frac{v}{M}\right) \qquad \tau > 0,\, y \in \mathbb{R}, \tag{2.169}$$

where $D, r,$ and M are positive parameters. An important question is to determine whether the gene has a typical speed of propagation.

Accordingly to the terminology in the introduction, (2.169) is a *semilinear equation* where diffusion is coupled with *logistic growth* through the reaction term

$$f\left(v\right) = rv\left(1 - \frac{v}{M}\right).$$

The parameter r represents a *biological potential* (net birth-death rate, with dimension $[time]^{-1}$), while M is the *carrying capacity* of the habitat. If we rescale time, space and concentration in the following way

$$t = r\tau, \quad x = \sqrt{r/D}y, \quad u = v/M,$$

(2.169) takes the dimensionless form

$$u_t = u_{xx} + u\left(1 - u\right), \qquad t > 0. \tag{2.170}$$

Note the two equilibria $u \equiv 0$ and $u \equiv 1$. In absence of diffusion, 0 is unstable, and 1 is asymptotically stable. A trajectory with initial data $u\left(0\right) = u_0$ between 0 and 1 has the typical behavior shown in figure 2.17:

----

[54] *Fisher, R. A. (1937), The wave of advance of advantageous gene.* Ann. Eugenics, **7**, 355-69.

[55] That is a *gene* that has an advantage in the struggle for life.
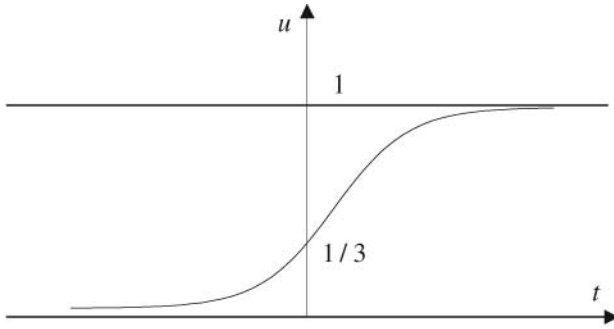
Fig. 2.17. Logistic curve $(r = 0.1, u_0 = 1/3)$

Therefore, if

$$u(x, 0) = u_0(x), \qquad x \in \mathbb{R}, \tag{2.171}$$

is an initial data for the equation (2.169), with $0 < u_0(x) < 1$, we expect a competitive action between diffusion and reaction, with diffusion trying to spread and lower u_0 against the reaction tendency to increase u towards the equilibrium solution 1.

What we intend to show here is the existence of permanent travelling waves solutions connecting the two equilibrium states, that is solutions of the form

$$u(x, t) = U(z), \qquad z = x - ct,$$

with c denoting the propagation speed, satisfying the conditions

$$0 < u < 1, \qquad t > 0, x \in \mathbb{R}$$

and

$$\lim_{x \to -\infty} u(x, t) = 1 \quad \text{and} \quad \lim_{x \to +\infty} u(x, t) = 0. \tag{2.172}$$

The first condition in (2.172), states that the gene concentration is saturated at the far left end while the second condition denotes zero concentration at the far right end. Clearly, this kind of solutions realize a balance between diffusion and reaction.

Since the equation (2.169) is invariant under the transformation $x \mapsto -x$, it suffices to consider $c > 0$, that is right-moving waves only.

Since

$$u_t = -cU', \quad u_x = U', \quad u_{xx} = U'', \qquad (' = d/dz)$$

substituting $u(x, t) = U(z)$ into (2.170), we find for U the ordinary differential equation

$$U'' + cU' + U - U^2 = 0 \tag{2.173}$$

with

$$\lim_{z \to -\infty} U(z) = 1 \quad \text{and} \quad \lim_{z \to +\infty} U(z) = 0. \tag{2.174}$$

Letting $U' = V$, the equation (2.173) is equivalent to the system

$$\frac{dU}{dz} = V, \quad \frac{dV}{dz} = -cV - U + U^2 \tag{2.175}$$

in the phase plane (U, V). This system has two equilibrium points $(0, 0)$ and $(1, 0)$ corresponding to two steady states. Our travelling wave solution corresponds to an orbit connecting $(1, 0)$ to $(0, 0)$, with $0 < U < 1$.

We first examine the local behavior of the orbits near the equilibrium points. The coefficients matrices of the linearized systems at $(0, 0)$ and $(1, 0)$ are, respectively,

$$J(0, 0) = \begin{pmatrix} 0 & 1 \\ -1 & -c \end{pmatrix} \quad \text{and} \quad J(1, 0) = \begin{pmatrix} 0 & 1 \\ 1 & -c \end{pmatrix}.$$

The eigenvalues of $J(0, 0)$ are

$$\lambda_\pm = \frac{1}{2} \left[-c \pm \sqrt{c^2 - 4} \right],$$

with corresponding eigenvectors

$$\mathbf{h}_\pm = \begin{pmatrix} -c \mp \sqrt{c^2 - 4} \\ 2 \end{pmatrix}.$$

If $c \geq 2$ the eigenvalues are both negative while if $c < 2$ they are complex. Therefore

$$(0, 0) \text{ is a } \begin{cases} \text{stable node if } c \geq 2 \\ \text{stable focus if } c < 2. \end{cases}$$

The eigenvalues of $J(1, 0)$ are

$$\mu_\pm = \frac{1}{2} \left[-c \pm \sqrt{c^2 + 4} \right],$$

of opposite sign, hence $(1, 0)$ is a saddle point. The unstable and stable separatrices leave $(1, 0)$ along the directions of the two eigenvectors

$$\mathbf{k}_+ = \begin{pmatrix} c + \sqrt{c^2 + 4} \\ 2 \end{pmatrix} \quad \text{and} \quad \mathbf{k}_- = \begin{pmatrix} c - \sqrt{c^2 + 4} \\ 2 \end{pmatrix},$$

respectively.

Now, the constraint $0 < U < 1$ rules out the case $c < 2$, since in this case U changes sign along the orbit approaching $(0, 0)$. For $c \geq 2$, all orbits[56] in a neighborhood of the origin approach $(0, 0)$ for $z \to +\infty$ asymptotically with slope

---

[56] Except for two orbits on the stable manifold tangent to $\mathbf{h}_-$ at $(0, 0)$, in the case $c > 2$.
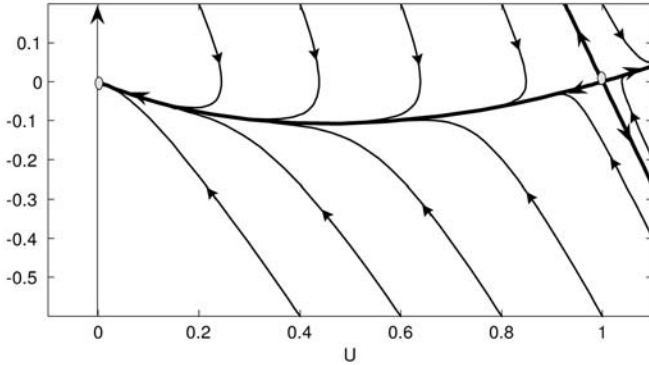
Fig. 2.18. Orbits of the system (2.175)

λ_+. On the other hand, the only orbit going to $(1,0)$ as $z \to -\infty$ and remaining in the region $0 < U < 1$ is the unstable separatrix γ of the saddle point.

Figure 2.18 shows the orbits configuration in the region of interest (see Problem 2.23). The conclusion is that *for each $c \geq 2$ there exists a unique travelling wave solution of equation (2.169) with speed c. Moreover U is strictly decreasing*.

In terms of original variables, there is a unique travelling wave solution for every speed c satisfying the inequality $c \geq c_{\min} = 2\sqrt{rD}$.

Thus, we have a continuous "spectrum" of possible speeds of propagation. It turns out that the minimum speed $c = c_{\min}$ is particularly important.

Indeed, having found a travelling solution is only the beginning of the story. There is a number of questions that arise naturally. Among them, the study of the *stability* of the travelling waves or of the asymptotic behavior (as $t \to +\infty$) of a solution with an initial data u_0 of *transitional* type, that is

$$u_0(x) = \begin{cases} 1 & x \leq a \\ 0 < u_0 < 1 & a < x < b \\ 0 & x \geq b. \end{cases} \qquad (2.176)$$

Should we expect that the travelling wave is insensitive to small perturbations? Does the solution with initial condition (2.176) evolve towards one of the travelling waves we have just found?

The interested reader can find the answers in the many specialized texts or papers on the subject[57]. Here we only mention that among the travelling wave solutions we have found, *only the minimum speed one* can be the asymptotic representation of solutions with transitional type initial condition. The biological implication of this result is that $c_{\min}$ *determines the required speed of propagation of an advantageous gene*.

---

[57] See for instance, the books by *Murray*, vol I, 2001, or *Grindrod*, 1991.

Problems

2.1. Use the method of separation of variables to solve the following initial-Neumann problem:

$$\begin{cases} u_t - u_{xx} = 0 & 0 < x < L, t > 0 \\ u(x,0) = x & 0 < x < L \\ u_x(0,t) = u_x(L,t) = 0 & t > 0. \end{cases}$$

2.2. Use the method of separation of variables to solve the following non homogeneous initial-Neumann problem:

$$\begin{cases} u_t - u_{xx} = tx & 0 < x < \pi, t > 0 \\ u(x,0) = 1 & 0 \le x \le \pi \\ u_x(0,t) = u_x(L,t) = 0 & t > 0. \end{cases}$$

[*Hint*: Write the candidate solution as $u(x,t) = \sum_{k \ge 0} c_k(t) v_k(x)$ where v_k are the eigenfunctions of the eigenvalue problem associated with the homogeneous equation].

2.3. Use the method of separation of variables to solve (at least formally) the following mixed problem:

$$\begin{cases} u_t - D u_{xx} = 0 & 0 < x < \pi, t > 0 \\ u(x,0) = g(x) & 0 \le x \le \pi \\ u_x(0,t) = 0 & \\ u_x(L,t) + u(L,t) = U & t > 0. \end{cases}$$

[*Answer*: $u(x,t) = \sum_{k \ge 0} c_k e^{-D\mu_k^2 t} \cos \mu_k x$, where the numbers μ_k are the positive solutions of the equation $\mu \tan \mu = 1$].

2.4. Prove that, if $w_t - D\Delta w = 0$ in Q_T and $w \in C\left(\overline{Q}_T\right)$, then

$$\min_{\partial_p Q_T} w \le w(\mathbf{x},t) \le \max_{\partial_p Q_T} w \qquad \text{for every } (\mathbf{x},t) \in Q_T.$$

2.5. Prove Corollary 2.1.

[*Hint*: b). Let $u = v - w$, $M = \max_{\overline{Q}_T} |f_1 - f_2|$ and apply Theorem 2.2 to $$z_\pm = \pm u - Mt].$$

2.6. Let $g(t) = M$ for $0 \le t \le 1$ and $g(t) = M - (1-t)^4$ for $1 < t \le 2$. Let u be the solution of $u_t - u_{xx} = 0$ in $Q_2 = (0,2) \times (0,2)$, $u = g$ on $\partial_p Q_2$. Compute $u(1,1)$ and check that it is the maximum of u. Is this in contrast with the strong maximum principle of Remark 2.4?

2.7. Suppose $u = u(x,t)$ is a solution of the heat equation in a plane domain $D_T = Q_T \setminus \left(\bar{Q}_1 \cup \bar{Q}_2\right)$ where Q_1 and Q_2 are the rectangles in figure 2.19. Assume that u attains its maximum M at the interior point (x_1, t_1). Where else $u = M$?
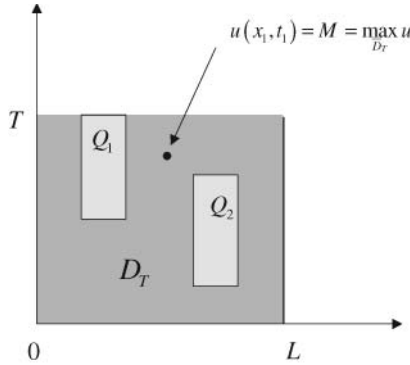
Fig. 2.19. At which points (x, t), $u(x, t) = M$?

2.8. Find the similarity solutions of the equation $u_t - u_{xx} = 0$ of the form $u(x, t) = U(x/\sqrt{t})$ and express the result in term of the *error function*

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz.$$

Find the solution of the problem $u_t - u_{xx} = 0$ in $x > 0, t > 0$ satisfying the conditions $u(0, t) = 1$ and $u(x, 0) = 0$, $x > 0$.

2.9. Determine for which α and β there exist similarity solutions to $u_t - u_{xx} = f(x)$ of the form $t^\alpha U(x/t^\beta)$ in each one of the following cases:

(a) $f(x) = 0$,   (b) $f(x) = 1$,   (c) $f(x) = x$.

[*Answer*: (a) α arbitrary, $\beta = 1/2$.   (b) $\alpha = 1$, $\beta = 1/2$.   (c) $\alpha = 3/2$, $\beta = 1/2$].

2.10. (*Reflecting barriers and Neumann condition*). Consider the symmetric random walk of Section 2.4. Suppose that a perfectly *reflecting* barrier is located at the point $L = \overline{m}h + \frac{h}{2} > 0$. By this we mean that if the particle hits the point $L - \frac{h}{2}$ at time t and moves to the right, then it is reflected and it comes back to $L - \frac{h}{2}$ at time $t + \tau$. Show that when $h, \tau \to 0$ and $h^2/\tau = 2D$, $p = p(x, t)$ is a solution of the problem

$$\begin{cases} p_t - Dp_{xx} = 0 & x < L, t > 0 \\ p(x, 0) = \delta & x < L \\ p_x(L, t) = 0 & t > 0 \end{cases}$$

and moreover $\int_{-\infty}^{L} p(x, t)\, dx = 1$. Compute explicitly the solution.

[*Answer*: $p(x, t) = \Gamma_D(x, t) + \Gamma_D(x - 2L, t)$].

2.11. (*Absorbing barrires and Dirichlet condition*). Consider the symmetric random walk of Section 2.4. Suppose that a perfectly *absorbing* barrier is located at the point $L = \overline{m}h > 0$. By this we mean that if the particle hits the point $L - h$

at time t and moves to the right then it is absorbed and stops at L. Show that when $h, \tau \to 0$ and $h^2/\tau = 2D$, $p = p(x,t)$ is a solution of the problem

$$\begin{cases} p_t - Dp_{xx} = 0 & x < L, t > 0 \\ p(x,0) = \delta & x < L \\ p(L,t) = 0 & t > 0 \end{cases}$$

Compute explicitly the solution.

[*Answer*: $p(x,t) = \Gamma_D(x,t) - \Gamma_D(x - 2L, t)$.]

2.12. Use the partial Fourier transform $\hat{u}(\xi, t) = \int_{\mathbb{R}} e^{-ix\xi} u(x,t) \, dx$ to solve the global Cauchy problem (2.129) and rediscover formula (2.130).

2.13. Prove Theorem 2.5 under the condition

$$z(x,t) \leq C, \quad x \in \mathbb{R}, \ 0 \leq t \leq T,$$

using the following steps.

a) Let $\sup_{\mathbb{R}} z(x,0) = M_0$ and define

$$w(x,t) = \frac{2C}{L^2}\left(\frac{x^2}{2} + Dt\right) + M_0.$$

Check that $w_t - Dw_{xx} = 0$ and use the maximum principle to show that $w \geq z$ in the rectangle $R_L = [-L, L] \times [0, T]$.

b) Fix an arbitrary point (x_0, t_0) and choose L large enough to have $(x_0, t_0) \in R_L$. Using a) deduce that $z(x_0, t_0) \leq M_0$.

2.14. Find an explicit formula for the solution of the global Cauchy problem

$$\begin{cases} u_t = Du_{xx} + bu_x + cu & x \in \mathbb{R}, t > 0 \\ u(x,0) = g(x) & x \in \mathbb{R}. \end{cases}$$

where D, b, c are constant coefficients. Show that, if $c < 0$ and g is bounded, $u(x,t) \to 0$ as $t \to +\infty$

[*Hint*: Choose h, k such that $v(x,t) = u(x,t) e^{hx+kt}$ is a solution of $v_t = Dv_{xx}$].

2.15. Find an explicit formula for the solution of the Cauchy problem

$$\begin{cases} u_t = u_{xx} & x > 0, t > 0 \\ u(x,0) = g(x) & x \geq 0 \\ u(0,t) = 0 & t > 0. \end{cases}$$

with g continuous and $g(0) = 0$.

[*Hint*: Extend g to $x < 0$ by odd reflection: $g(-x) = -g(x)$. Solve the corresponding global Cauchy problem and write the result as an integral on $(0, +\infty)$].

2.16. Let $Q_T = \Omega \times (0, T)$, with Ω bounded domain in $\mathbb{R}^n$. Let $u \in C^{2,1}(Q_T) \cap C(\overline{Q}_T)$ satisfy the equation

$$u_t = D\Delta u + \mathbf{b}(\mathbf{x},t) \cdot \nabla u + c(\mathbf{x},t) u \qquad \text{in } Q_T$$

where $\mathbf{b}$ and c are continuous in $\overline{Q}_T$. Show that if $u \geq 0$ (resp. $u \leq 0$) on $\partial_p Q_T$ then $u \geq 0$ (resp. $u \leq 0$) in Q_T.

[*Hint*: Assume first that $c(\mathbf{x},t) \leq a < 0$. Then reduce to this case by setting $u = v e^{kt}$ with a suitable $k > 0$].

2.17. Fill in the details in the arguments of Section 6.2, leading to formulas (2.108) and (2.109).

2.18. Solve the following initial-Dirichlet problem in $B_1 = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| < 1\}$:

$$\begin{cases} u_t = \Delta u & \mathbf{x} \in B_1, t > 0 \\ u(\mathbf{x}, 0) = 0 & \mathbf{x} \in B_1 \\ u(\boldsymbol{\sigma}, t) = 1 & \boldsymbol{\sigma} \in \partial B_1, t > 0. \end{cases}$$

Compute $\lim_{t \to +\infty} u$.

[*Hint*: The solution is radial so that $u = u(r, t)$, $r = |\mathbf{x}|$. Observe that $\Delta u = u_{rr} + \dfrac{2}{r} u_r = \dfrac{1}{r}(ru)_{rr}$. Let $v = ru$, reduce to homogeneous Dirichlet condition and use separation of variables].

2.19. Solve the following initial-Dirichlet problem

$$\begin{cases} u_t = \Delta u & \mathbf{x} \in K, t > 0 \\ u(\mathbf{x}, 0) = 0 & \mathbf{x} \in K \\ u(\boldsymbol{\sigma}, t) = 1 & \boldsymbol{\sigma} \in \partial K, t > 0. \end{cases}$$

where K is the rectangular box

$$K = \{(x, y, z) \in \mathbb{R}^3 : 0 < x < a, 0 < y < b, 0 < z < c\}.$$

Compute $\lim_{t \to +\infty} u$.

2.20. Solve the following initial-Neumann problem in $B_1 = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| < 1\}$:

$$\begin{cases} u_t = \Delta u & \mathbf{x} \in B_1, t > 0 \\ u(\mathbf{x}, 0) = |\mathbf{x}| & \mathbf{x} \in B_1 \\ u_\nu(\boldsymbol{\sigma}, t) = 1 & \boldsymbol{\sigma} \in \partial B_1, t > 0. \end{cases}$$

2.21. Solve the following non homogeneous initial-Dirichlet problem in the unit sphere B_1 $(u = u(r, t), r = |\mathbf{x}|)$:

$$\begin{cases} u_t - (u_{rr} + \dfrac{2}{r} u_r) = q e^{-t} & 0 < r < 1, t > 0 \\ u(r, 0) = U & 0 \leq r \leq 1 \\ u(1, t) = 0 & t > 0. \end{cases}$$

[*Answer*: The solution is

$$u(r, t) = \frac{2}{r} \sum_{n=1}^{\infty} \frac{(-1)^n}{\lambda_n} \sin(\lambda_n r) \left\{ \frac{q}{1 - \lambda_n^2} \left(e^{-t} - e^{-\lambda_n^2 t} \right) - U e^{-\lambda_n^2 t} \right\}$$

where $\lambda_n = n\pi$].

2.22. Using the maximum principle, compare the values of two call options $C^{(1)}$ and $C^{(2)}$ in the following cases:

(*a*) Same exercise price and $T_1 > T_2$. (*b*) Same expiry time and $E_1 > E_2$.

2.23. Justify carefully the orbit configuration of figure 2.18 and in particular that the unstable orbit γ connects the two equilibrium points of system (2.175), by filling in the details in the following steps:

1. Let $\mathbf{F} = V\mathbf{i} + \left(-cV + U^2 - U\right)\mathbf{j}$ and $\mathbf{n}$ be the interior normal to the boundary of the triangle Ω in figure 2.20. Show that, if β is large enough, $\mathbf{F} \cdot \mathbf{n} > 0$ along $\partial\Omega$.

2. Deduce that all the orbits of system (2.175) starting at a point in Ω cannot leave Ω (i.e. Ω is a *positively invariant region*) and converge to the origin as $z \to +\infty$.

3. Finally, deduce that the unstable separatrix γ of the saddle point $(1,0)$ approaches $(0,0)$ as $z \to +\infty$.
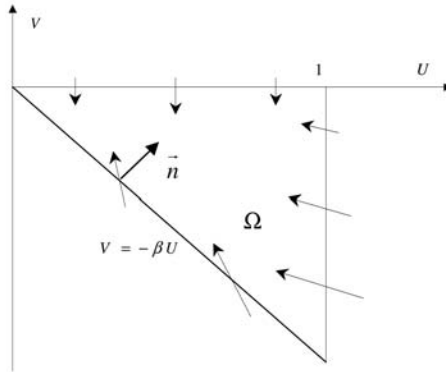
Fig. 2.20. Trapping region for the orbits of the vector field $\mathbf{F} = V\mathbf{i} + \left(-cV + U^2 - U\right)\mathbf{j}$

3

The Laplace Equation

Introduction – Well Posed Problems. Uniqueness – Harmonic Functions – Fundamental Solution and Newtonian Potential – The Green Function – Uniqueness in Unbounded Domains – Surface Potentials

3.1 Introduction

The Laplace equation $\Delta u = 0$ occurs frequently in applied sciences, in particular in the study of the *steady state phenomena*. Its solutions are called *harmonic* functions. For instance, the equilibrium position of a perfectly elastic membrane is a harmonic function as it is the velocity potential of a homogeneous fluid. Also, the steady state temperature of a homogeneous and isotropic body is a harmonic function and in this case Laplace equation constitutes the stationary counterpart (time independent) of the diffusion equation.

Slightly more generally, Poisson's equation $\Delta u = f$ plays an important role in the theory of *conservative fields* (electrical, magnetic, gravitational,...) where the vector field is derived from the gradient of a potential.

For example, let $\mathbf{E}$ be a force field due to a distribution of electric charges in a domain $\Omega \subset \mathbb{R}^3$. Then, in standard units, div $\mathbf{E} = 4\pi\rho$, where ρ represents the density of the charge distribution. When a *potential* u exists such that $\nabla u = -\mathbf{E}$, then $\Delta u = \text{div}\nabla u = -4\pi\rho$, which is Poisson's equation. If the electric field is created by charges located outside Ω, then $\rho = 0$ in Ω and u is harmonic therein. Analogously, the potential of a gravitational field due to a mass distribution is a harmonic function in a region free from mass.

In dimension two, the theories of harmonic and holomorphic functions are strictly connected[1]. Indeed, the real and the imaginary part of a holomorphic

---

[1] A complex function $f = f(z)$ is *holomorphic* in an open subset Ω of the complex plane if for every $z_0 \in \Omega$, the limit

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

function are harmonic. For instance, since the functions

$$z^m = r^m \left(\cos m\theta + i \sin m\theta\right), \qquad m \in \mathbb{N},$$

$(r, \theta$ polar coordinates) are holomorphic in the whole plane $\mathbb{C}$, the functions

$$u\left(r, \theta\right) = r^m \cos m\theta \quad \text{and} \quad v\left(r, \theta\right) = r^m \sin m\theta \qquad m \in \mathbb{N},$$

are harmonic in $\mathbb{R}^2$ (called *elementary harmonics*). In Cartesian coordinates, they are harmonic polynomials; for $m = 1, 2, 3$ we find

$$x, \ y, \ xy, \ x^2 - y^2, \ x^3 - 3xy^2, \ 3x^2y - y^3.$$

Other examples are

$$u\left(x, y\right) = e^{\alpha x} \cos \alpha y, \quad v\left(x, y\right) = e^{\alpha x} \sin \alpha y \qquad \left(\alpha \in \mathbb{R}\right),$$

the real and imaginary parts of $f\left(z\right) = e^{i\alpha z}$, both harmonic in $\mathbb{R}^2$, and

$$u\left(r, \theta\right) = \log r, \quad v\left(r, \theta\right) = \theta,$$

the real and imaginary parts of $f\left(z\right) = \log_0 z \equiv \log r + i\theta$, harmonic in $\mathbb{R}^2 \backslash \left(0, 0\right)$ and $\mathbb{R}^2 \backslash \left\{\theta = 0\right\},$ respectively.

In this chapter we present the formulation of the most important well posed problems and the classical properties of harmonic functions, focusing mainly on dimensions two and three. As in Chapter 2, we emphasize some probabilistic aspects, exploiting the connection among random walks, Brownian motion and the Laplace operator. A central notion is the concept of *fundamental solution*, that we develop in conjunction with the very basic elements of the so called *potential theory*.

3.2 Well Posed Problems. Uniqueness

Consider the Poisson equation

$$\Delta u = f \qquad \text{in } \Omega \tag{3.1}$$

where $\Omega \subset \mathbb{R}^n$ is a **bounded domain**. The well posed problems associated with equation (3.1) are the stationary counterparts of the corresponding problems for the diffusion equation. Clearly here there is no initial condition. On the boundary $\partial \Omega$ we may assign:

- *Dirichlet data*

$$u = g, \tag{3.2}$$

---

exists and it is finite.

- *Neumann data*

$$\partial_{\boldsymbol{\nu}} u = h, \tag{3.3}$$

where $\boldsymbol{\nu}$ is the outward normal unit vector to $\partial\Omega$,

- a *Robin (radiation) condition*

$$\partial_{\boldsymbol{\nu}} u + \alpha u = h \qquad (\alpha > 0), \tag{3.4}$$

- a *mixed condition*; for instance,

$$
\begin{aligned}
u &= g &&\text{on } \Gamma_D \\
\partial_{\boldsymbol{\nu}} u &= h &&\text{on } \Gamma_N,
\end{aligned}
\tag{3.5}
$$

where $\Gamma_D \cup \Gamma_N = \partial\Omega$, $\Gamma_D \cap \Gamma_N = \varnothing$, and Γ_N is a relatively open subset of $\partial\Omega$.

When $g = h = 0$ we say that the above boundary conditions are *homogeneous*.

We give some interpretations. If u is the position of a perfectly flexible membrane and f is an external distributed load (vertical force per unit surface), then (3.1) models a steady state.

The Dirichlet condition corresponds to fixing the position of the membrane at its boundary. Robin condition describes an elastic attachment at the boundary while a homogeneous Neumann condition corresponds to a free vertical motion of the boundary.

If u is the steady state concentration of a substance, the Dirichlet condition prescribes the level of u at the boundary, while the Neumann condition assigns the flux of u through the boundary.

Using Green's identity (1.13) we can prove the following uniqueness result.

Theorem 3.1. *Let $\Omega \subset \mathbb{R}^n$ be a smooth, bounded domain. Then there exists at most one solution $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ of (3.1), satisfying on $\partial\Omega$ one of the conditions (3.2), (3.4) or (3.5).*

In the case of the Neumann condition, that is when

$$\partial_{\boldsymbol{\nu}} u = h \qquad \text{on } \partial\Omega,$$

two solutions differ by a constant.

Proof. Let u and v be solutions of the same problem, sharing the same boundary data, and let $w = u - v$. Then w is harmonic and satisfies homogeneous boundary conditions (one among (3.2)-(3.5)). Substituting $u = v = w$ into (1.13) we find

$$\int_{\Omega} |\nabla w|^2 \, d\mathbf{x} = \int_{\partial\Omega} w \partial_{\boldsymbol{\nu}} w \, d\sigma.$$

If Dirichlet or mixed conditions hold, we have

$$\int_{\partial\Omega} w \partial_{\boldsymbol{\nu}} w \, d\sigma = 0.$$

When a Robin condition holds

$$\int_{\partial\Omega} w\partial_\nu w \, d\sigma = -\int_{\partial\Omega} \alpha w^2 d\sigma \le 0.$$

In any case we obtain that

$$\int_\Omega |\nabla w|^2 \, d\mathbf{x} \le 0. \tag{3.6}$$

From (3.6) we infer $\nabla w = \mathbf{0}$ and therefore $w = u - v = $ constant. This concludes the proof in the case of Neumann condition. In the other cases, the constant must be zero (why?), hence $u = v$. $\square$

Remark 3.1. Consider the Neumann problem $\Delta u = f$ in Ω, $\partial_\nu u = h$ on $\partial\Omega$. Integrating the equation on Ω and using Gauss' formula we find

$$\int_\Omega f \, d\mathbf{x} = \int_{\partial\Omega} h \, d\sigma. \tag{3.7}$$

The relation (3.7) appears as a *compatibility* condition on the data f and h, that has *necessarily* to be satisfied in order for the Neumann problem to admit a solution. Thus, when having to solve a Neumann problem, the first thing to do is to check the validity of (3.7). If it does not hold, the problem does not have any solution. We will examine later the physical meaning of (3.7).

3.3 Harmonic Functions

3.3.1 Discrete harmonic functions

In Chapter 2 we have examined the connection between Brownian motion and diffusion equation. We go back now to the multidimensional symmetric random walk considered in Section 2.6, analyzing its relation with the Laplace operator Δ. For simplicity we will work in dimension $n = 2$ but both arguments and conclusions may be easily extended to any dimension $n > 2$. We fix a time step $\tau > 0$, a space step $h > 0$ and denote by $h\mathbb{Z}^2$ the *lattice* of points $\mathbf{x} = (x_1, x_2)$ whose coordinates are integer multiples of h. Let $p(\mathbf{x}, t) = p(x_1, x_2, t)$ be the transition probability function, giving the probability to find our random particle at $\mathbf{x}$ at time t. From the total probability formula we found a difference equation for p, that we rewrite in dimension two:

$$p(\mathbf{x}, t + \tau) = \frac{1}{4}\{p(\mathbf{x}+h\mathbf{e}_1, t) + p(\mathbf{x}-h\mathbf{e}_1, t) + p(\mathbf{x}+h\mathbf{e}_2, t) + p(\mathbf{x}-h\mathbf{e}_2, t)\}. \tag{3.8}$$

We can write this formula in a more significant way by introducing the *mean value operator* M_h, whose action on a function $u = u(\mathbf{x})$ is defined by the following

formula:

$$M_h f\left(\mathbf{x}\right) = \frac{1}{4}\left\{u\left(\mathbf{x}+h\mathbf{e}_1\right) + u\left(\mathbf{x}-h\mathbf{e}_1\right) + u\left(\mathbf{x}+h\mathbf{e}_2\right) + u\left(\mathbf{x}-h\mathbf{e}_2\right)\right\}$$
$$= \frac{1}{4}\sum_{|\mathbf{x}-\mathbf{y}|=h} u\left(\mathbf{y}\right).$$

Note that $M_h u\left(\mathbf{x}\right)$ gives the average of u over the points of the lattice $h\mathbb{Z}^2$ at distance h from $\mathbf{x}$. We say that these points constitute the *discrete neighborhood of* $\mathbf{x}$ *of radius* h.

It is clear that (3.8) can be written in the form

$$p\left(\mathbf{x}, t + \tau\right) = M_h p\left(\mathbf{x}, t\right). \tag{3.9}$$

In (3.9) the probability p at time $t + \tau$ is determined by the action of M_h at the previous time, and then it is natural to interpret the mean value operator as *the generator of the random walk*.

Now we come to the Laplacian. If u is twice continuously differentiable, it is not difficult to show that[2]

$$\lim_{h \to 0} \frac{M_h u\left(\mathbf{x}\right) - u\left(\mathbf{x}\right)}{h^2} \to \frac{1}{4}\Delta u\left(\mathbf{x}\right). \tag{3.10}$$

The formula (3.10) induces to define, for any fixed $h > 0$, a *discrete Laplace operator* through the formula

$$\Delta_h^* = M_h - I$$

where I denotes the *identity* operator (i.e. $Iu = u$). The operator Δ_h^* acts on functions u defined in the whole lattice $h\mathbb{Z}^2$ and, coherently, we say that u is *d-harmonic* (*d* for *discrete*) if $\Delta_h^* u = 0$.

Thus, the value of a *d-harmonic* function at any point $\mathbf{x}$ is given by the average of the values at the points in the discrete neighborhood of $\mathbf{x}$ of radius h.

We can proceed further and define a discrete Dirichlet problem. Let A be a subset of $h\mathbb{Z}^2$.

We say that A is *connected* if, given any couple of points $\mathbf{x}_0$, $\mathbf{x}_1$ in A, it is possible to connect them by a walk[3] on $h\mathbb{Z}^2$ entirely contained in A.

Moreover, we say that $\mathbf{x} \in A$ is an *interior* point of A if its $h-$neighborhood is contained in A. The points of A that are not interior are called *boundary points* (Fig. 3.1). The set of the boundary points of A, the *boundary* of A, is denoted by ∂A.

---

[2] Using a second order Taylor's polynomial, after some simplifications, we get:

$$M_h u\left(\mathbf{x}\right) = u\left(\mathbf{x}\right) + \frac{h^2}{4}\left\{u_{x_1 x_1}\left(\mathbf{x}\right) + u_{x_2 x_2}\left(\mathbf{x}\right)\right\} + o\left(h^2\right)$$

from which formula (3.10) comes easily.

[3] Recall that consecutive points in a walk have distance h.

○ [Interior points]   ● Boundary points
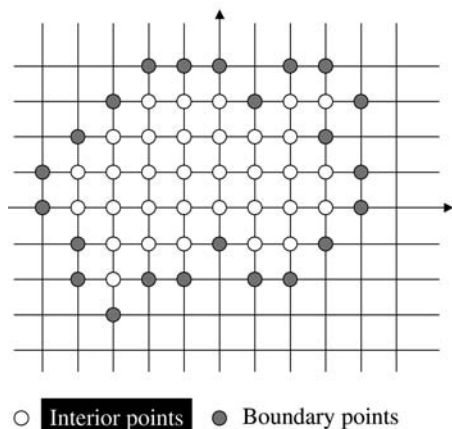
Fig. 3.1. A domain for the discrete Dirichlet problem

Discrete Dirichlet problem. Let A be a *bounded connected* subset of $h\mathbb{Z}^2$ and g be a function defined on the boundary ∂A of A. We want to determine u, defined on A, such that

$$\begin{cases} \Delta_h^* u = 0 \text{ at the interior points of } A \\ u = g \quad \text{ on } \partial A. \end{cases} \tag{3.11}$$

We deduce immediately three important properties of a solution u:

1. Maximum principle: *If u attains its maximum or its minimum at an interior point then u is constant.* Indeed, suppose $\mathbf{x} \in A$ is an interior point and $u(\mathbf{x}) = M \geq u(\mathbf{y})$ for every $\mathbf{y} \in A$. Since $u(\mathbf{x})$ is the average of the four values of u at the points at distance h from $\mathbf{x}$, at all these points u must be equal to M. Let $\mathbf{x}_1 \neq \mathbf{x}$ be one of these neighboring points. By the same argument, $u(\mathbf{y}) = M$ for every $\mathbf{y}$ in the h−neighborhood of $\mathbf{x}_1$. Since A is connected, proceeding in this way we prove that $u(\mathbf{y}) = M$ at every point of A.
2. *u attains its maximum and its minimum on ∂A.* This is an immediate consequence of 1.
3. *The solution of the discrete Dirichlet problem is unique* (exercise).

The discrete Dirichlet problem (3.11) has a remarkable probabilistic interpretation that can be used to construct its solution. Let us go back to our random particle. First of all, we want to show that whatever its starting point $\mathbf{x} \in A$ is, the particle hits the boundary ∂A *with probability one.*

For every $\Gamma \subseteq \partial A$, we denote by

$$P(\mathbf{x}, \Gamma)$$

the *probability that the particle starting from $\mathbf{x} \in A$ hits ∂A for the first time at a point $\mathbf{y} \in \Gamma$.* We have to prove that $P(\mathbf{x}, \partial A) = 1$ for every $\mathbf{x} \in A$.

Clearly, if $\mathbf{x} \in \Gamma$ we have $P(\mathbf{x}, \Gamma) = 1$, while if $\mathbf{x} \in \partial A \backslash \Gamma$, $P(\mathbf{x}, \Gamma) = 0$. It turns out that, for fixed Γ, the function

$$w_\Gamma(\mathbf{x}) = P(\mathbf{x}, \Gamma)$$

is *d-harmonic* in the interior of A, that is $\Delta^* w_\Gamma = 0$. To see this, denote by

$$p(1, \mathbf{x}, \mathbf{y})$$

the *one step transition probability,* i.e. the probability to go from $\mathbf{x}$ to $\mathbf{y}$ in one step. Given the symmetry of the walk, we have $p(1, \mathbf{x}, \mathbf{y}) = 1/4$ if $|\mathbf{x} - \mathbf{y}| = 1$ and $p(1, \mathbf{x}, \mathbf{y}) = 0$ otherwise.

Now, to hit Γ starting from $\mathbf{x}$, the particle first hits a point $\mathbf{y}$ in its $h-$neighborhood and from there it reaches Γ, independently of the first step. Then, by the total probability formula we can write

$$w_\Gamma(\mathbf{x}) = P(\mathbf{x}, \Gamma) = \sum_{\mathbf{y} \in h\mathbb{Z}^2} p(1, \mathbf{x}, \mathbf{y}) P(\mathbf{y}, \Gamma) = M_h P(\mathbf{x}, \Gamma) = M_h w_\Gamma(\mathbf{x}),$$

which entails

$$(I - M_h) w_\Gamma = \Delta_h^* w_\Gamma = 0.$$

Thus, w_Γ is $d - harmonic$ in A. In particular $w_{\partial A}(\mathbf{x}) = P(\mathbf{x}, \partial A)$ is $d - harmonic$ in A and $w_{\partial A} = 1$ on ∂A. On the other hand, the function $z(\mathbf{x}) \equiv 1$ satisfies the same discrete Dirichlet problem, so that, by the uniqueness property 3 above,

$$w_{\partial A}(\mathbf{x}) = P(\mathbf{x}, \partial A) \equiv 1 \text{ in } A. \tag{3.12}$$

This means that the particle hits the boundary ∂A *with probability one.* As a consequence, observe that the set function

$$\Gamma \mapsto P(\mathbf{x}, \Gamma)$$

defines a probability measure on ∂A, for any fixed $\mathbf{x} \in A$.

We construct now the solution u to (3.11). Interpret the boundary data g as a *payoff*: if the particle starts from $\mathbf{x}$ and hits the boundary for the first time at $\mathbf{y}$, it wins $g(\mathbf{y})$. We have:

Theorem 3.2. *The value $u(\mathbf{x})$ is given by the expected value of the winnings $g(\cdot)$ with respect to the probability $P(\mathbf{x}, \cdot)$. That is*

$$u(\mathbf{x}) = \sum_{\mathbf{y} \in \partial A} g(\mathbf{y}) P(\mathbf{x}, \{\mathbf{y}\}). \tag{3.13}$$

Proof. Each term

$$g(\mathbf{y}) P(\mathbf{x}, \{\mathbf{y}\}) = g(\mathbf{y}) w_{\{\mathbf{y}\}}(\mathbf{x})$$

is $d-$harmonic in A and therefore u is $d-$harmonic in A as well. Moreover, if $\mathbf{x} \in \partial A$ then $u(\mathbf{x}) = g(\mathbf{x})$ since each term in the sum is equal to $g(\mathbf{x})$ if $\mathbf{y} = \mathbf{x}$ or to zero if $\mathbf{y} \neq \mathbf{x}$. $\square$

As $h \to 0$, formula (3.10) shows that, formally, $d-$harmonic functions "become" harmonic. Thus, it seems reasonable that appropriate versions of the above properties and results should hold in the continuous case. We start with the mean value properties.

3.3.2 Mean value properties

Guided by their discrete characterization, we want to establish some fundamental properties of harmonic functions. To be precise, we say that a function u is *harmonic* in a domain $\Omega \subseteq \mathbb{R}^n$ if $u \in C^2(\Omega)$ and $\Delta u = 0$ in Ω.

Since $d-$harmonic functions are defined through a mean value property, we expect that harmonic functions inherit a mean value property of the following kind: the value at the center of any ball $B \subset\subset \Omega$, i.e. compactly contained in Ω, equals the average of the values on the boundary ∂B. Actually, something more is true.

Theorem 3.3. *Let u be harmonic in $\Omega \subseteq \mathbb{R}^n$. Then, for any ball $B_R(\mathbf{x}) \subset\subset \Omega$ the following mean value formulas hold:*

$$u(\mathbf{x}) = \frac{n}{\omega_n R^n} \int_{B_R(\mathbf{x})} u(\mathbf{y}) \, d\mathbf{y} \tag{3.14}$$

$$u(\mathbf{x}) = \frac{1}{\omega_n R^{n-1}} \int_{\partial B_R(\mathbf{x})} u(\boldsymbol{\sigma}) \, d\sigma \tag{3.15}$$

where ω_n is the measure of ∂B_1.

Proof (for $n = 2$). Let us start from the second formula. For $r < R$ define

$$g(r) = \frac{1}{2\pi r} \int_{\partial B_r(\mathbf{x})} u(\boldsymbol{\sigma}) \, d\sigma.$$

Perform the change of variables $\boldsymbol{\sigma} = \mathbf{x} + r\boldsymbol{\sigma}'$. Then $\boldsymbol{\sigma}' \in \partial B_1(\mathbf{0})$, $d\sigma = r d\sigma'$ and

$$g(r) = \frac{1}{2\pi} \int_{\partial B_1(\mathbf{0})} u(\mathbf{x} + r\boldsymbol{\sigma}') \, d\sigma'.$$

Let $v(\mathbf{y}) = u(\mathbf{x} + r\mathbf{y})$ and observe that

$$\nabla v(\mathbf{y}) = r\nabla u(\mathbf{x} + r\mathbf{y})$$
$$\Delta v(\mathbf{y}) = r^2 \Delta u(\mathbf{x} + r\mathbf{y}).$$

Then we have

$$g'(r) = \frac{1}{2\pi} \int_{\partial B_1(\mathbf{0})} \frac{d}{dr} u(\mathbf{x}+r\boldsymbol{\sigma}') \, d\sigma' = \frac{1}{2\pi} \int_{\partial B_1(\mathbf{0})} \nabla u(\mathbf{x}+r\boldsymbol{\sigma}') \cdot \boldsymbol{\sigma}' d\sigma'$$

$$= \frac{1}{2\pi r} \int_{\partial B_1(\mathbf{0})} \nabla v(\boldsymbol{\sigma}') \cdot \boldsymbol{\sigma}' d\sigma' = \text{(divergence theorem)}$$

$$= \frac{1}{2\pi r} \int_{B_1(\mathbf{0})} \Delta v(\mathbf{y}) \, d\mathbf{y} = \frac{r}{2\pi} \int_{B_1(\mathbf{0})} \Delta u(\mathbf{x}+r\mathbf{y}) \, d\mathbf{y} = 0.$$

Thus, g is constant and since $g(r) \to u(\mathbf{x})$ for $r \to 0$, we get (3.15).

To obtain (3.14), let $R = r$ in (3.15), multiply by r and integrate both sides between 0 and R. We find

$$\frac{R^2}{2} u(\mathbf{x}) = \frac{1}{2\pi} \int_0^R dr \int_{\partial B_r(\mathbf{x})} u(\boldsymbol{\sigma}) \, d\sigma = \frac{1}{2\pi} \int_{B_R(\mathbf{x})} u(\mathbf{y}) \, d\mathbf{y}$$

from which (3.14) follows. □

Even more significant is a converse of Theorem 3.2. We say that a **continuous** function u satisfies the mean value property in Ω, if (3.14) or (3.15) holds for any ball $B_R(\mathbf{x}) \subset\subset \Omega$. It turns out that if u is continuous and possesses the mean value property in a domain Ω, then u is harmonic in Ω. Thus we obtain a characterization of harmonic functions through a mean value property, as in the discrete case. As a by product we deduce that every harmonic function in a domain Ω is continuously differentiable of any order in Ω, that is, it belongs to $C^\infty(\Omega)$. Notice that this is not a trivial fact since it involves derivatives not appearing in the expression of the Laplace operator. For instance, $u(x,y) = x + y|y|$ is a solution of $u_{xx} + u_{xy} = 0$ in all $\mathbb{R}^2$ but it is not twice differentiable with respect to y at $(0,0)$.

Theorem 3.4. *Let* $u \in C(\Omega)$. *If* u *satisfies the mean value property, then* $u \in C^\infty(\Omega)$ *and it is harmonic in* Ω.

We postpone the proof to the end of the Section 3.4.

3.3.3 Maximum principles

As in the discrete case, a function satisfying the mean value property in a domain[4] Ω cannot attain its maximum or minimum at an *interior point of* Ω, unless it is constant. In case Ω is bounded and u (non constant) is continuous up to the boundary of Ω, it follows that u attains both its maximum and minimum **only on** $\partial\Omega$. This result expresses a maximum principle that we state precisely in the following theorem.

---

[4] Recall that a *domain* is an open *connected* set.

Theorem 3.5. *Let $u \in C(\Omega)$, $\Omega \subseteq \mathbb{R}^n$. If u has the mean value property and attains its maximum or minimum at $\mathbf{p} \in \Omega$, then u is constant. In particular, if Ω is bounded and $u \in C(\overline{\Omega})$ is not constant, then, for every $\mathbf{x} \in \Omega$,*

$$u(\mathbf{x}) < \max_{\partial\Omega} u \quad and \quad u(\mathbf{x}) > \min_{\partial\Omega} u \qquad \text{(from strong maximum principle)}.$$

Proof. $(n = 2)$. Let $\mathbf{p}$ be a minimum point[5] for u:

$$m = u(\mathbf{p}) \leq u(\mathbf{y}), \qquad \forall \mathbf{y} \in \Omega.$$

We want to show that $u \equiv m$ in Ω. Let $\mathbf{q}$ be another arbitrary point in Ω. Since Ω is connected, it is possible to find a finite sequence of circles $B(\mathbf{x}_j) \subset\subset \Omega$, $j = 0, ..., N$, such that (Fig. 3.2):

- $\mathbf{x}_j \in B(\mathbf{x}_{j-1})$, $j = 1, ..., N$
- $x_0 = \mathbf{p}$, $x_N = \mathbf{q}$.
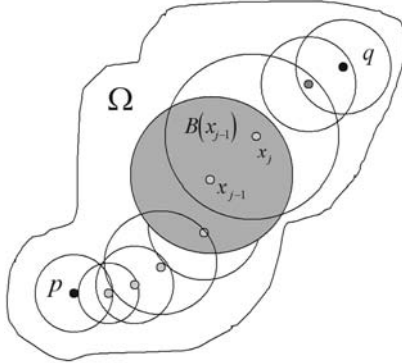
Fig. 3.2. A sequence of overlapping circles connecting the points $\mathbf{p}$ and $\mathbf{q}$

The mean value property gives

$$m = u(\mathbf{p}) = \frac{1}{|B(\mathbf{p})|} \int_{B(\mathbf{p})} u(\mathbf{y}) \, d\mathbf{y}.$$

Suppose there exists $\mathbf{z} \in B(\mathbf{p})$ such that $u(\mathbf{z}) > m$. Then, given a circle $B_r(\mathbf{z}) \subset B(\mathbf{p})$, we can write:

$$m = \frac{1}{|B(\mathbf{p})|} \int_{B(\mathbf{p})} u(\mathbf{y}) \, d\mathbf{y} \tag{3.16}$$

$$= \frac{1}{|B(\mathbf{p})|} \left\{ \int_{B(\mathbf{p}) \setminus B_r(\mathbf{z})} u(\mathbf{y}) \, d\mathbf{y} + \int_{B_r(\mathbf{z})} u(\mathbf{y}) \, d\mathbf{y} \right\}.$$

---

[5] The argument for the maximum is the same.

Since $u(\mathbf{y}) \geq m$ for every $\mathbf{y}$ and, by the mean value again,

$$\int_{B_r(\mathbf{z})} u(\mathbf{y}) \, d\mathbf{y} = u(\mathbf{z}) \left| B_r(\mathbf{z}) \right| > m \left| B_r(\mathbf{z}) \right|,$$

continuing from (3.16) we obtain

$$> \frac{1}{|B(\mathbf{p})|} \left\{ m \left| B(\mathbf{p}) \setminus B_r(\mathbf{z}) \right| + m \left| B_r(\mathbf{z}) \right| \right\} = m$$

and therefore the contradiction $m > m$.

Thus it must be that $u \equiv m$ in $B(\mathbf{p})$ and in particular $u(\mathbf{x}_1) = m$. We repeat now the same argument with $\mathbf{x}_1$ in place of $\mathbf{p}$ to show that $u \equiv m$ in $B(\mathbf{x}_1)$ and in particular $u(\mathbf{x}_2) = m$. Iterating the procedure we eventually deduce that $u(\mathbf{x}_N) = u(\mathbf{q}) = m$. Since $\mathbf{q}$ is an arbitrary point of Ω, we conclude that $u \equiv m$ in Ω. $\square$

An important consequence of the maximum principle is the following corollary.

Corollary 3.1. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and $g \in C(\partial\Omega)$. The problem*

$$\begin{cases} \Delta u = 0 & \text{in } \Omega \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{3.17}$$

has at most a solution $u_g \in C^2(\Omega) \cap C(\overline{\Omega})$. Moreover, let u_{g_1} and u_{g_2} be the solutions corresponding to the data $g_1, g_2 \in C(\partial\Omega)$. Then:

(a) (Comparison). If $g_1 \geq g_2$ on $\partial\Omega$ and $g_1 \neq g_2$, then

$$u_{g_1} > u_{g_2} \qquad \text{in } \Omega. \tag{3.18}$$

(b) (Stability).

$$\left| u_{g_1}(\mathbf{x}) - u_{g_2}(\mathbf{x}) \right| \leq \max_{\partial\Omega} \left| g_1 - g_2 \right| \qquad \text{for every } \mathbf{x} \in \Omega. \tag{3.19}$$

Proof. We first show (a) and (b). Let $w = u_{g_1} - u_{g_2}$. Then w is harmonic and $w = g_1 - g_2 \geq 0$ on $\partial\Omega$. Since $g_1 \neq g_2$, w is not constant and from Theorem 3.5

$$w(\mathbf{x}) > \min_{\partial\Omega}(g_1 - g_2) \geq 0 \quad \text{for every } \mathbf{x} \in \Omega.$$

This is (3.18). To prove (b), apply Theorem 3.5 to w and $-w$ to find

$$\pm w(\mathbf{x}) \leq \max_{\partial\Omega} \left| g_1 - g_2 \right| \quad \text{for every } \mathbf{x} \in \Omega$$

which is equivalent to (3.19).

Now if $g_1 = g_2$, (3.19) implies $w = u_{g_1} - u_{g_2} \equiv 0$, so that the Dirichlet problem (3.17) has at most one solution. $\square$

Remark 3.2. Inequality (3.19) is a *stability estimate*. Indeed, suppose g is known within an absolute error less than ε, or, in other words, suppose g_1 is an approximation of g and $\max_{\partial\Omega} |g - g_1| < \varepsilon$; then (3.19) gives

$$\max_{\overline{\Omega}} \left| u_{g_1} - u_g \right| < \varepsilon$$

so that the approximate solution is known within the same absolute error.

3.3.4 The Dirichlet problem in a circle. Poisson's formula

To prove the existence of a solution to one of the boundary value problems we considered in Section 3.2 is not an elementary task. In Chapter 8, we solve this question in a general context, using the more advanced tools of Functional Analysis. However, in special cases, elementary methods, like separation of variables, work. We use it to compute the solution of the Dirichlet problem in a circle. Precisely, let $B_R = B_R(\mathbf{p})$ be the circle of radius R centered at $\mathbf{p} = (p_1, p_2)$ and $g \in C(\partial B_R)$. We want to prove the following theorem.

Theorem 3.6. *The unique solution $u \in C^2(B_R) \cap C(\overline{B}_R)$ of the problem*

$$\begin{cases} \Delta u = 0 \ in \ B_R \\ u = g \quad on \ \partial B_R. \end{cases} \tag{3.20}$$

*is given by **Poisson's formula***

$$u(\mathbf{x}) = \frac{R^2 - |\mathbf{x} - \mathbf{p}|^2}{2\pi R} \int_{\partial B_R(\mathbf{p})} \frac{g(\boldsymbol{\sigma})}{|\mathbf{x} - \boldsymbol{\sigma}|^2} d\sigma. \tag{3.21}$$

In particular, $u \in C^\infty(B_R)$.

    Proof. The symmetry of the domain suggests the use of polar coordinates

$$x_1 = p_1 + r \cos\theta \qquad x_2 = p_2 + r \sin\theta.$$

Accordingly, let

$$U(r, \theta) = u(p_1 + r \cos\theta, p_2 + r \sin\theta), \ G(\theta) = g(p_1 + R \cos\theta, p_2 + R \sin\theta).$$

The Laplace equation becomes[6]

$$U_{rr} + \frac{1}{r} U_r + \frac{1}{r^2} U_{\theta\theta} = 0, \qquad 0 < r < R, \ 0 \le \theta \le 2\pi, \tag{3.22}$$

with the Dirichlet condition

$$U(R, \theta) = G(\theta), \qquad 0 \le \theta \le 2\pi.$$

Since we ask that u be continuous in $\overline{B}_R$, then U and G have to be continuous in $[0, R] \times [0, 2\pi]$ and $[0, 2\pi]$, respectively; moreover both have to be $2\pi-$periodic with respect to θ.

    We use now the method of separation of variables, by looking first for solutions of the form

$$U(r, \theta) = v(r) w(\theta)$$

---

[6] Appendix B.

with v, w bounded and w $2\pi-$periodic. Substitution in (3.22) gives

$$v''(r) w(\theta) + \frac{1}{r} v'(r) w(\theta) + \frac{1}{r^2} v(r) w''(\theta) = 0$$

or, separating the variables,

$$-\frac{r^2 v''(r) + r v'(r)}{v(r)} = \frac{w''(\theta)}{w(\theta)}.$$

This identity is possible only when the two quotients have a common constant value λ. Thus we are lead to the ordinary differential equation

$$r^2 v''(r) + r v'(r) - \lambda v(r) = 0 \qquad (3.23)$$

and to the eigenvalue problem

$$\begin{cases} w''(\theta) - \lambda w(\theta) = 0 \\ w(0) = w(2\pi). \end{cases} \qquad (3.24)$$

We leave to the reader to check that problem (3.24) has only the zero solution for $\lambda \geq 0$. If $\lambda = -\mu^2$, $\mu > 0$, the differential equation in (3.24) has the general integral

$$w(\theta) = a \cos \mu\theta + b \sin \mu\theta \qquad (a, b \in \mathbb{R}).$$

The $2\pi-$periodicity forces $\mu = m$, a nonnegative integer.

The equation (3.23), with $\lambda = m$, has the general solution[7]

$$v(r) = d_1 r^{-m} + d_2 r^m \qquad (d_1, d_2 \in \mathbb{R}).$$

Since v has to be bounded we exclude r^{-m}, $m > 0$ and hence $d_1 = 0$.

We have found a countable number of $2\pi-$periodic harmonic functions

$$r^m \{a_m \cos m\theta + b_m \sin m\theta\} \qquad m = 0, 1, 2, ... \qquad (3.25)$$

We superpose now the (3.25) by writing

$$U(r, \theta) = a_0 + \sum_{m=1}^{\infty} r^m \{a_m \cos m\theta + b_m \sin m\theta\} \qquad (3.26)$$

with the coefficients a_m and b_m still to be chosen in order to satisfy the boundary condition

$$\lim_{(r,\theta)\to(R,\xi)} U(r, \theta) = G(\xi) \qquad \forall \xi \in [0, 2\pi]. \qquad (3.27)$$

---

[7] It is an Euler equation. The change of variables $s = \log r$ reduces it to the equation

$$v''(s) - m^2 v(s) = 0.$$

Case $G \in C^1([0, 2\pi])$. In this case G can be expanded in a uniformly convergent Fourier series

$$G(\xi) = \frac{\alpha_0}{2} + \sum_{m=1}^{\infty} \{\alpha_m \cos m\xi + \beta_m \sin m\xi\}$$

where

$$\alpha_m = \frac{1}{\pi} \int_0^{2\pi} G(\varphi) \cos m\varphi \, d\varphi, \qquad \beta_m = \frac{1}{\pi} \int_0^{2\pi} G(\varphi) \sin m\varphi \, d\varphi.$$

Then, the boundary condition (3.27) is satisfied if we choose

$$a_0 = \frac{\alpha_0}{2}, \quad a_m = R^{-m} \alpha_m, \quad b_m = R^{-m} \beta_m.$$

Substitution of these values of a_0, a_m, b_m into (3.26) gives, for $r \leq R$,

$$U(r, \theta) = \frac{\alpha_0}{2} + \frac{1}{\pi} \sum_{m=1}^{\infty} \left(\frac{r}{R}\right)^m \int_0^{2\pi} G(\varphi) \{\cos m\varphi \cos m\theta + \sin m\varphi \sin m\theta\} \, d\varphi$$

$$= \frac{1}{\pi} \int_0^{2\pi} G(\varphi) \left[\frac{1}{2} + \sum_{m=1}^{\infty} \left(\frac{r}{R}\right)^m \{\cos m\varphi \cos m\theta + \sin m\varphi \sin m\theta\}\right] d\varphi$$

$$= \frac{1}{\pi} \int_0^{2\pi} G(\varphi) \left[\frac{1}{2} + \sum_{m=1}^{\infty} \left(\frac{r}{R}\right)^m \cos m(\varphi - \theta)\right] d\varphi.$$

Note that in the second equality above, the exchange of sum and integration is possible because of the uniform convergence of the series. Moreover, for $r < R$, we can differentiate under the integral sign and then term by term as many times as we want (why?). Therefore, since for every $m \geq 1$ the functions

$$\left(\frac{r}{R}\right)^m \cos m(\varphi - \theta)$$

are smooth and harmonic, also $U \in C^{\infty}(B_R)$ and is harmonic for $r < R$.

To obtain a better formula, observe that

$$\sum_{m=1}^{\infty} \left(\frac{r}{R}\right)^m \cos m(\varphi - \theta) = \text{Re} \left[\sum_{m=1}^{\infty} \left(e^{i(\varphi - \theta)} \frac{r}{R}\right)^m\right].$$

Since

$$\text{Re} \sum_{m=1}^{\infty} \left(e^{i(\varphi - \theta)} \frac{r}{R}\right)^m = \text{Re} \frac{1}{1 - e^{i(\varphi - \theta)} \frac{r}{R}} - 1 = \frac{R^2 - rR \cos(\varphi - \theta)}{R^2 + r^2 - 2rR \cos(\varphi - \theta)} - 1$$

$$= \frac{rR \cos(\varphi - \theta) - r^2}{R^2 + r^2 - 2rR \cos(\varphi - \theta)}$$

we find

$$\frac{1}{2} + \sum_{m=1}^{\infty} \left(\frac{r}{R}\right)^m \cos m(\varphi - \theta) = \frac{1}{2}\frac{R^2 - r^2}{R^2 + r^2 - 2rR\cos(\varphi - \theta)}. \tag{3.28}$$

Inserting (3.28) into the formula for U, we get **Poisson's formula** in polar coordinates:

$$U(r, \theta) = \frac{R^2 - r^2}{2\pi} \int_0^{2\pi} \frac{G(\varphi)}{R^2 + r^2 - 2Rr\cos(\theta - \varphi)} d\varphi. \tag{3.29}$$

Going back to Cartesian coordinates[8] we obtain Poisson's formula (3.21). Corollary 3.1 assures that (3.29) is indeed the unique solution of the Dirichlet problem (3.20).

Case $G \in C([0, 2\pi])$. We drop now the additional hypothesis that $G(\theta)$ is continuously differentiable. Even with G only continuous, formula (3.29) makes perfect sense and defines a harmonic function in B_R and it can be shown that[9]

$$\lim_{(r,\theta)\to(R,\xi)} U(r, \theta) = G(\xi), \qquad \forall\xi \in [0, 2\pi].$$

Therefore (3.29) is the unique (by Corollary 3.1) solution to (3.20). $\square$

• *Poisson's formula in dimension $n > 2$.* Theorem 3.6 has an appropriate extension in any number of dimensions. When $B_R = B_R(\mathbf{p})$ is an $n-$ dimensional ball, the solution of the Dirichlet problem (3.20) is given by (see subsection 3.5.3, for the case $n = 3$)

$$u(\mathbf{x}) = \frac{R^2 - |\mathbf{x} - \mathbf{p}|^2}{\omega_n R} \int_{\partial B_R(\mathbf{p})} \frac{g(\boldsymbol{\sigma})}{|\mathbf{x} - \boldsymbol{\sigma}|^n} d\sigma. \tag{3.30}$$

An important consequence of Poisson's formula is the possibility to control the derivatives of any order of a harmonic function u at a point $\mathbf{p}$ by the maximum of u in a small ball centered at $\mathbf{p}$. We show it for first and second derivatives in the following corollary.

Corollary 3.2. *Let u be a harmonic function in a domain Ω and $B_R(\mathbf{p}) \subset\subset \Omega$. Then*

$$\left|u_{x_j}(\mathbf{p})\right| \le \frac{n}{R} \max_{\partial B_R(\mathbf{p})} |u|, \qquad \left|u_{x_j x_k}(\mathbf{p})\right| \le \frac{c(n)}{R^2} \max_{\partial B_R(\mathbf{p})} |u|. \tag{3.31}$$

---

[8] With $\boldsymbol{\sigma} = R(\cos\varphi, \sin\varphi)$, $d\sigma = Rd\varphi$ and

$$\begin{aligned}|\mathbf{x} - \boldsymbol{\sigma}|^2 &= (r\cos\theta - R\cos\varphi)^2 + (r\sin\theta - R\sin\varphi)^2 \\ &= R^2 + r^2 - 2Rr(\cos\varphi\cos\theta + \sin\varphi\sin\theta) \\ &= R^2 + r^2 - 2Rr\cos(\theta - \varphi).\end{aligned}$$

[9] See Problem 3.20.

Proof. From (3.30) we have

$$u\left(\mathbf{x}\right) = \frac{R^2 - |\mathbf{x} - \mathbf{p}|^2}{\omega_n R} \int_{\partial B_R(\mathbf{p})} \frac{u\left(\boldsymbol{\sigma}\right)}{|\mathbf{x} - \boldsymbol{\sigma}|^n} d\sigma.$$

Since we want to compute the derivatives at $\mathbf{p}$, we can differentiate under the integral obtaining:

$$u_{x_j}\left(\mathbf{x}\right) = \frac{-2\left(x_j - p_j\right)}{\omega_n R} \int_{\partial B_R(\mathbf{p})} \frac{u\left(\boldsymbol{\sigma}\right)}{|\mathbf{x} - \boldsymbol{\sigma}|^n} d\sigma$$
$$-n\frac{R^2 - 2|\mathbf{x} - \mathbf{p}|^2}{\omega_n R} \int_{\partial B_R(\mathbf{p})} \frac{x_j - \sigma_j}{|\mathbf{x} - \boldsymbol{\sigma}|^{n+2}} u\left(\boldsymbol{\sigma}\right) d\sigma.$$

Now, at $\mathbf{x} = \mathbf{p}$ we have

$$\frac{|p_j - \sigma_j|}{|\mathbf{p} - \boldsymbol{\sigma}|^{n+2}} \le R^{-n-1}.$$

Therefore, since $|\partial B_R\left(\mathbf{p}\right)| = \omega_n R^{n-1}$,

$$\left|u_{x_j}\left(\mathbf{p}\right)\right| \le \frac{nR}{\omega_n} \int_{\partial B_R(\mathbf{p})} \frac{|p_j - \sigma_j|}{|\mathbf{p} - \boldsymbol{\sigma}|^{n+2}} |u\left(\boldsymbol{\sigma}\right)| d\sigma \le \frac{n}{R} \max_{\partial B_R(\mathbf{p})} |u|.$$

Similarly we get the estimates for the second derivatives; we leave the details to the reader. $\square$

We are now in position to prove Theorem 3.4, the converse of the mean value property $(m.v.p.)$.

• *Proof of Theorem* 3.4. First observe that if two functions satisfy the $m.v.p.$ in a domain Ω, their difference satisfies this property as well. Let $u \in C\left(\Omega\right)$ satisfying the $m.v.p.$ and consider a circle $B \subset\subset \Omega$. We want to show that u is harmonic and infinitely differentiable in Ω. Denote by v the solution of the Dirichlet problem

$$\begin{cases} \Delta v = 0 \text{ in } B \\ v = u \quad \text{on } \partial B. \end{cases}$$

From Theorem 3.6 we know that $v \in C^\infty\left(B\right) \cap C\left(\overline{B}\right)$ and, being harmonic, it satisfies the $m.v.p.$ in B. Then, also $w = v - u$ satisfies the $m.v.p.$ in B and therefore (Theorem 3.4) it attains its maximum and minimum on ∂B. Since $w = 0$ on ∂B, we conclude that $u = v$ in B. Since B is arbitrary, $u \in C^\infty\left(\Omega\right)$ and is harmonic in Ω. $\square$

3.3.5 Harnack's inequality and Liouville's theorem

From the mean value and Poisson's formulas we deduce another maximum principle, known as *Harnack's inequality*:

Theorem 3.7. *Let u be harmonic and nonnegative in the ball $B_R = B_R(\mathbf{0}) \subset \mathbb{R}^n$. Then for any $\mathbf{x} \in B_R$,*

$$\frac{R^{n-2}(R - |\mathbf{x}|)}{(R + |\mathbf{x}|)^{n-1}} u(\mathbf{0}) \leq u(\mathbf{x}) \leq \frac{R^{n-2}(R + |\mathbf{x}|)}{(R - |\mathbf{x}|)^{n-1}} u(\mathbf{0}). \tag{3.32}$$

Proof ($n = 3$). From Poisson's formula:

$$u(\mathbf{x}) = \frac{R^2 - |\mathbf{x}|^2}{4\pi R} \int_{\partial B_R} \frac{u(\boldsymbol{\sigma})}{|\boldsymbol{\sigma} - \mathbf{x}|^3} d\sigma.$$

Observe that $R - |\mathbf{x}| \leq |\boldsymbol{\sigma} - \mathbf{x}| \leq R + |\mathbf{x}|$ and $R^2 - |\mathbf{x}|^2 = (R - |\mathbf{x}|)(R + |\mathbf{x}|)$. Then, by the mean value property,

$$u(\mathbf{x}) \leq \frac{R + |\mathbf{x}|}{(R - |\mathbf{x}|)^2} \frac{1}{4\pi R} \int_{\partial B_R} u(\boldsymbol{\sigma}) \, d\sigma = \frac{R(R + |\mathbf{x}|)}{(R - |\mathbf{x}|)^2} u(\mathbf{0}).$$

Analogously,

$$u(\mathbf{x}) \geq \frac{R(R - |\mathbf{x}|)}{(R + |\mathbf{x}|)^2} \frac{1}{4\pi R^2} \int_{\partial B_R} u(\boldsymbol{\sigma}) \, d\sigma = \frac{R(R - |\mathbf{x}|)}{(R + |\mathbf{x}|)^2} u(\mathbf{0}).$$

$\square$

Harnack's inequality has an important consequence: the only harmonic functions in $\mathbb{R}^n$ bounded from below or above are the constant functions.

Corollary 3.3. (Liouville′s Theorem). *If u is harmonic in $\mathbb{R}^n$ and $u(\mathbf{x}) \geq M$, then u is constant.*

Proof ($n = 3$). The function $w = u - M$ is harmonic in $\mathbb{R}^3$ and nonnegative. Fix $\mathbf{x} \in \mathbb{R}^3$ and choose $R > |\mathbf{x}|$; Harnack's inequality gives

$$\frac{R(R - |\mathbf{x}|)}{(R + |\mathbf{x}|)^2} w(\mathbf{0}) \leq w(\mathbf{x}) \leq \frac{R(R + |\mathbf{x}|)}{(R - |\mathbf{x}|)^2} w(\mathbf{0}). \tag{3.33}$$

Letting $R \to \infty$ in (3.33) we get

$$w(\mathbf{0}) \leq w(\mathbf{x}) \leq w(\mathbf{0})$$

whence $w(\mathbf{0}) = w(\mathbf{x})$. Since $\mathbf{x}$ is arbitrary we conclude that w, and therefore also u, is constant. $\square$

3.3.6 A probabilistic solution of the Dirichlet problem

In Section 3.1 we solved the discrete Dirichlet problem via a probabilistic method. The key ingredients in the construction of the solution, leading to formula (3.13), were the mean value property and the absence of memory of the random walk (each

step is independent of the preceding ones). In the continuous case the appropriate version of those tools are available, with the Markov property encoding the absence of memory of Brownian motion[10]. Thus it is reasonable that a suitable continuous version of formula (3.13) should give the solution of the Dirichlet problem for the Laplace operator. As before, we work in dimension $n = 2$, but methods and conclusions can be extended without much effort to any number of dimensions.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain and $g \in C(\partial\Omega)$. We want to derive a representation formula for the unique solution $u \in C^2(\Omega) \cap C(\overline{\Omega})$ of the problem

$$\begin{cases} \Delta u = 0 \text{ in } \Omega \\ u = g \quad \text{on } \partial\Omega. \end{cases} \tag{3.34}$$

Let $\mathbf{X}(t)$ be the position of a Brownian particle started at $\mathbf{x} \in \Omega$ and define *the first exit time from Ω, $\tau = \tau(\mathbf{x})$,* as follows (Fig. 3.3):

$$\tau(\mathbf{x}) = \left\{ \inf_{t \geq 0} t : \mathbf{X}(t) \in \mathbb{R}^2 \backslash \Omega \right\}.$$
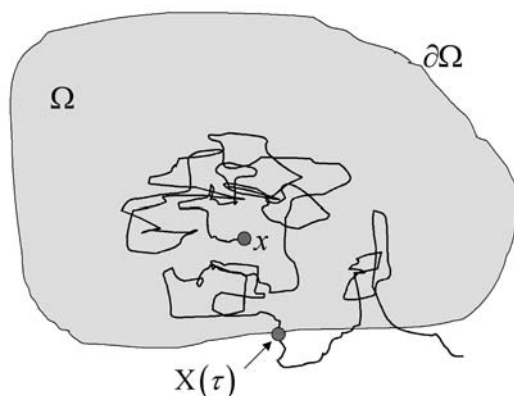
Fig. 3.3. First exit point from Ω

The time τ is a *stopping time*: to decide whether the event $\{\tau \leq t\}$ occurs or not, *it suffices to observe the process until time t.* In fact, for fixed $t \geq 0$, to decide whether or not $\tau \leq t$ is true, it is enough to consider the event

$$E = \{\mathbf{X}(s) \in \Omega, \text{ for all times } s \text{ from } 0 \text{ until } t, t \text{ included}\}.$$

If this event occurs, then it must be that $\tau > t$. If E does not occur, it means that there are points $\mathbf{X}(s)$ outside Ω for some $s \leq t$ and therefore it must be that $\tau \leq t$.

The first thing we have to check is that the particle leaves Ω in finite time, almost surely. Precisely:

---

[10] Section 2.6.

Lemma 3.1. *For every* $\mathbf{x} \in \Omega$, $\tau(\mathbf{x})$ *is finite with probability 1, that is:*

$$P\{\tau(\mathbf{x}) < \infty\} = 1.$$

Proof. It is enough to show that our particle remains inside any circle $B_r = B_r(\mathbf{x}) \subset \Omega$ with zero probability. If we denote by τ_r the first exit time from B_r, we have to prove that $P\{\tau_r = \infty\} = 0$.

Suppose $\mathbf{X}(t) \in B_r$ until $t = k$ (k integer). Then, for $j = 1, 2, ..., k$, it must be that

$$|\mathbf{X}(j) - \mathbf{X}(j-1)| < 2r.$$

Thus (the occurrence of) the event $\{\tau_r > k\}$ implies (the occurrence of) all the events

$$E_j = \{|\mathbf{X}(j) - \mathbf{X}(j-1)| < 2r\} \qquad j = 1, 2, ..., k$$

and therefore also of their intersection. As a consequence

$$P\{\tau_r > k\} \leq P\{\cap_{j=1}^k E_j\}. \tag{3.35}$$

On the other hand, the increments $\mathbf{X}(j) - \mathbf{X}(j-1)$ are mutually independent and equidistributed according to a standard normal law, hence we can write

$$P\{E_j\} = \frac{1}{2\pi} \int_{\{|\mathbf{z}|<2r\}} \exp\left(-\frac{|\mathbf{z}|^2}{2}\right) d\mathbf{z} \equiv \gamma < 1$$

and

$$P\{\cap_{j=1}^k E_j\} = \prod_{j=1}^k P\{E_j\} = \gamma^k. \tag{3.36}$$

Since $\{\tau_r = \infty\}$ implies $\{\tau_r > k\}$, from (3.35) and (3.36) we have

$$P\{\tau_r = \infty\} \leq P\{\tau_r > k\} \leq \gamma^k.$$

Letting $k \to +\infty$ we get $P\{\tau_r = \infty\} = 0$. $\square$

Lemma 3.1 implies that $\mathbf{X}(\tau)$ hits the boundary $\partial\Omega$ in finite time, with probability 1. We can therefore introduce on $\partial\Omega$ a probability distribution associated with the random variable $\mathbf{X}(\tau)$ by setting

$$P(\mathbf{x}, \tau, F) = P\{\mathbf{X}(\tau) \in F\} \qquad\qquad (\tau = \tau(\mathbf{x})),$$

for every "reasonable" subset $F \subset \partial\Omega$[11]. $P(\mathbf{x}, \tau, F)$ is called the *escape probability from Ω through F*. For fixed $\mathbf{x}$ in Ω, the set function

$$F \longmapsto P(\mathbf{x}, \tau(\mathbf{x}), F)$$

defines a probability measure on $\partial\Omega$, since $P(\mathbf{x}, \tau(\mathbf{x}), \partial\Omega) = 1$, according to Lemma 3.1.

---

[11] Precisely, for every *Borel set* (Appendix B).

By analogy with formula (3.13), we can now guess the type of formula we expect for the solution u of problem (3.34). To get the value $u(\mathbf{x})$, let a Brownian particle start from $\mathbf{x}$, and let $\mathbf{X}(\tau) \in \partial\Omega$ its first exit point from Ω. Then, compute the random "gain" $g(\mathbf{X}(\tau))$ and take its expected value with respect to the distribution $P(\mathbf{x}, \tau, \cdot)$. This is $u(\mathbf{x})$. Everything works if $\partial\Omega$ is not too bad. Precisely, we have:

Theorem 3.8. *Let Ω be a bounded Lipschitz domain and $g \in C(\partial\Omega)$. The unique solution $u \in C^2(\Omega) \cap C(\overline{\Omega})$ of problem (3.34) is given by*

$$u(\mathbf{x}) = E^{\mathbf{x}}[g(\mathbf{X}(\tau))] = \int_{\partial\Omega} g(\boldsymbol{\sigma}) P(\mathbf{x}, \tau(\mathbf{x}), d\sigma). \qquad (3.37)$$

Proof (sketch). For fixed $F \subseteq \partial\Omega$, consider the function

$$u_F \colon \mathbf{x} \longmapsto P(\mathbf{x}, \tau(\mathbf{x}), F)).$$

We claim that u_F is *harmonic* in Ω. Assuming that u_F is continuous[12] in Ω, from Theorem 3.4, it is enough to show that u_F satisfies the mean value property. Let $B_R = B_R(\mathbf{x}) \subset\subset \Omega$. If $\tau_R = \tau_R(\mathbf{x})$ is the first exit time from B_R, then $\mathbf{X}(\tau_R)$ has a uniform distribution on ∂B_R, due to the invariance by rotation of the Brownian motion.

This means that, starting from the center, the escape probability from B_R through any arc $K \subset \partial B_R$ is given by

$$\frac{\text{length of } K}{2\pi R}.$$

Now, before reaching F, the particle must hit ∂B_R. Since τ_R is a stopping time, we may use the strong Markov property. Thus, after τ_R, $\mathbf{X}(t)$ can be considered as a Brownian motion with uniform initial distribution on ∂B_R, expressed by the formula (Fig. 3.4)

$$\mu(ds) = \frac{ds}{2\pi R},$$

where ds is the length element on ∂B_R. Therefore, the particle escapes ∂B_R through some arc of length ds centered at a point $\mathbf{s}$ and from there it reaches F with probability $P(\mathbf{s}, \tau(\mathbf{s}), F)\mu(ds)$. By integrating this probability on ∂B_R we obtain $P(\mathbf{x}, \tau(\mathbf{x}), F)$, namely:

$$P(\mathbf{x}, \tau(\mathbf{x}), F) = \int_{\partial B_R(\mathbf{x})} P(\mathbf{s}, \tau(\mathbf{s}), F)\mu(ds) = \frac{1}{2\pi R} \int_{\partial B_R(\mathbf{x})} P(\mathbf{s}, \tau(\mathbf{s}), F)ds.$$

which is the mean value property for u_F.

Observe now that if $\boldsymbol{\sigma} \in \partial\Omega$ then $\tau(\boldsymbol{\sigma}) = 0$ and hence

$$P(\boldsymbol{\sigma}, \tau(\boldsymbol{\sigma}), F) = \begin{cases} 1 \text{ if } \boldsymbol{\sigma} \in F \\ 0 \text{ if } \boldsymbol{\sigma} \in \partial\Omega\backslash F. \end{cases}$$

---

[12] Which should be at least intuitively clear.
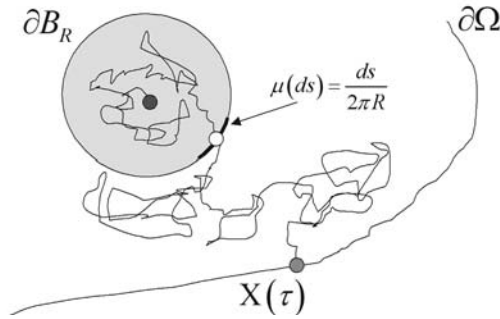
Fig. 3.4. Strong Markov property of a Brownian particle

Therefore, on $\partial\Omega$, u_F coincides with the characteristic function of F. Thus if $d\sigma$ is an arc element on $\partial\Omega$ centered in $\boldsymbol{\sigma}$, intuitively, the function

$$\mathbf{x} \longmapsto g\left(\boldsymbol{\sigma}\right) P\left(\mathbf{x}, \tau\left(\mathbf{x}\right), d\sigma\right) \tag{3.38}$$

is harmonic in Ω, it attains the value $g\left(\boldsymbol{\sigma}\right)$ on $d\sigma$ and it is zero on $\partial\Omega\backslash d\sigma$. To obtain the harmonic function equal to g on $\partial\Omega$, we integrate over $\partial\Omega$ all the contributions from (3.38). This gives the representation (3.37).

Rigorously, to assert that (3.37) is indeed the required solution, we should check that $u\left(\mathbf{x}\right) \to g\left(\boldsymbol{\sigma}\right)$ when $\mathbf{x} \to \boldsymbol{\sigma}$. It can be proved[13] that this is true if Ω is, for instance, a Lipschitz domain. $\square$

Remark 3.3. The measure

$$F \longmapsto P(\mathbf{x}, \tau\left(\mathbf{x}\right), F)$$

is called the *harmonic measure at* $\mathbf{x}$ *of the domain* Ω and in general it can not be expressed by an explicit formula. In the particular case $\Omega = B_R\left(\mathbf{p}\right)$, Poisson's formula (3.21) indicates that the harmonic measure for the circle $B_R\left(\mathbf{p}\right)$ is given by

$$P(\mathbf{x}, \tau\left(\mathbf{x}\right), d\sigma) = \frac{1}{2\pi R} \frac{R^2 - \left|\mathbf{x} - \mathbf{p}\right|^2}{\left|\boldsymbol{\sigma} - \mathbf{x}\right|^2} d\sigma.$$

Remark 3.4. Formula (3.37) shows that the value of the solution at a point $\mathbf{x}$ *depends on the boundary data on all* $\partial\Omega$ (except for sets of length zero). In the case of figure 3.5, a change of the data g on the arc AB affects the value of the solution at $\mathbf{x}$, even if this point is far from AB and near $\partial\Omega$.

3.3.7 Recurrence and Brownian motion

We have seen that the solution of a Dirichlet problem can be constructed by using the general properties of a Brownian motion. On the other hand, the deterministic

---

[13] The proof is rather delicate (see *Øksendal*, 1995).
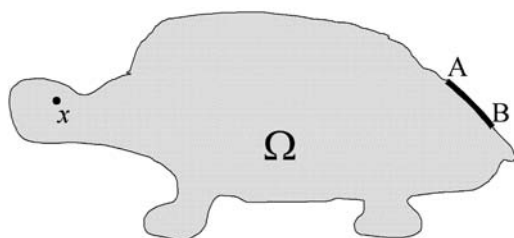
Fig. 3.5. A modification of the Dirichlet data on the arc AB, affects the value of the solution at $\mathbf{x}$

solution of some Dirichlet problems can be used to deduce interesting properties of Brownian motion. We examine two simple examples. Recall from Section 3.1 that $\ln |\mathbf{x}|$ is harmonic in the plane except $\mathbf{x} = \mathbf{0}$.

Let a, R be real numbers, $R > a > 0$. It is easy to check that the function

$$u_R(\mathbf{x}) = \frac{\ln |R| - \ln |\mathbf{x}|}{\ln R - \ln a}$$

is harmonic in the ring $B_{a,R} = \left\{ \mathbf{x} \in \mathbb{R}^2; a < |\mathbf{x}| < R \right\}$ and moreover

$$u_R(\mathbf{x}) = 1 \text{ on } \partial B_a(\mathbf{0}), \qquad u_R(\mathbf{x}) = 0 \text{ on } \partial B_R(\mathbf{0}).$$

Thus $u(\mathbf{x})$ represents the escape probability from the ring through $\partial B_a(\mathbf{0})$, starting at $\mathbf{x}$:

$$u_R(\mathbf{x}) = P_R(\mathbf{x}, \tau(\mathbf{x}), \partial B_a(\mathbf{0})).$$

Letting $R \to +\infty$, we get

$$P_R(\mathbf{x}, \tau(\mathbf{x}), \partial B_a(\mathbf{0})) = \frac{\ln |R| - \ln |\mathbf{x}|}{\ln R - \ln a} \to 1 = P_\infty(\mathbf{x}, \tau(\mathbf{x}), \partial B_a(\mathbf{0})).$$

This means that, starting at $\mathbf{x}$, the probability we (sooner or later) *enter* the circle $B_a(\mathbf{0})$, is 1. Due to the invariance by translations of the Brownian motion, the origin can be replaced by any other point without changing the conclusions. Moreover, since we have proved in Lemma 3.1 that the exit probability from any circle is also 1, we can state the following result: *given any point $\mathbf{x}$ and any circle in the plane, a Brownian particle started at $\mathbf{x}$ enters the circle and exit from it an infinite number of times, with probability* 1. We say that a bidimensional Brownian motion is *recurrent*.

In three dimensions a Brownian motion is *not* recurrent. In fact (see the next section), the function

$$u(\mathbf{x}) = \frac{\frac{1}{|\mathbf{x}|} - \frac{1}{R}}{\frac{1}{a} - \frac{1}{R}}$$

is harmonic in the spherical shell $B_{a,R} = \left\{ \mathbf{x} \in \mathbb{R}^3; a < |\mathbf{x}| < R \right\}$ and

$$u(\mathbf{x}) = 1 \text{ on } \partial B_a(\mathbf{0}), \qquad u(\mathbf{x}) = 0 \text{ on } \partial B_R(\mathbf{0}).$$

Then $u\left(\mathbf{x}\right)$ represents the escape probability from the shell through $\partial B_a\left(\mathbf{0}\right)$, starting at $\mathbf{x}$:

$$u_R\left(\mathbf{x}\right) = P_R\left(\mathbf{x},\tau\left(\mathbf{x}\right),\partial B_a\left(\mathbf{0}\right)\right).$$

This time, letting $R\rightarrow +\infty$, we find

$$P_R\left(\mathbf{x},\tau\left(\mathbf{x}\right),\partial B_a\left(\mathbf{0}\right)\right) = \frac{\frac{1}{|\mathbf{x}|} - \frac{1}{R}}{\frac{1}{a} - \frac{1}{R}} \rightarrow \frac{a}{|\mathbf{x}|} = P_\infty\left(\mathbf{x},\tau\left(\mathbf{x}\right),\partial B_a\left(\mathbf{0}\right)\right).$$

Thus, the probability to enter, sooner or later, the sphere $B_a\left(\mathbf{0}\right)$ *is not* 1 and it becomes smaller and smaller as the distance of $\mathbf{x}$ from the origin increases.

3.4 Fundamental Solution and Newtonian Potential

3.4.1 The fundamental solution

The (3.37) is not the only representation formula for the solution of the Dirichlet problem. We shall derive deterministic formulas involving various types of *potentials*, constructed using a special function, called the *fundamental solution* of the Laplace operator.

As we did for the diffusion equation, let us look at the invariance properties characterizing the operator Δ: the invariances by *translations* and by *rotations*.

Let $u = u\left(\mathbf{x}\right)$ be harmonic in $\mathbb{R}^n$. Invariance by translations means that the function $v\left(\mathbf{x}\right) = u\left(\mathbf{x} - \mathbf{y}\right)$, for each fixed $\mathbf{y}$, is also harmonic, as it is immediate to check.

Invariance by rotations means that, given a rotation in $\mathbb{R}^n$, represented by an orthogonal matrix $\mathbf{M}$ (i.e. $\mathbf{M}^T = \mathbf{M}^{-1}$), also $v\left(\mathbf{x}\right) = u\left(\mathbf{Mx}\right)$ is harmonic in $\mathbb{R}^n$. To check it, observe that, if we denote by D^2u the Hessian of u, we have

$$\Delta u = \mathrm{Tr}D^2u = \text{ trace of the Hessian of } u.$$

Since

$$D^2v\left(\mathbf{x}\right) = \mathbf{M}^T D^2u\left(\mathbf{Mx}\right)\mathbf{M}$$

and $\mathbf{M}$ is orthogonal, we have

$$\Delta v\left(\mathbf{x}\right) = \mathrm{Tr}[\mathbf{M}^T D^2u\left(\mathbf{Mx}\right)\mathbf{M}] = \mathrm{Tr}D^2u\left(\mathbf{Mx}\right) = \Delta u\left(\mathbf{Mx}\right) = 0$$

and therefore v is harmonic.

Now, a typical rotation invariant quantity is *the distance function from a point*, for instance from the origin, that is $r = |\mathbf{x}|$. Thus, let us look for *radially symmetric* harmonic functions $u = u\left(r\right)$.

Consider first $n = 2$; using polar coordinates and recalling (3.22), we find

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} = 0$$

so that

$$u(r) = C \log r + C_1.$$

In dimension $n = 3$, using spherical coordinates (r, ψ, θ), $r > 0$, $0 < \psi < \pi$, $0 < \theta < 2\pi$, the operator Δ has the following expression[14]:

$$\Delta = \underbrace{\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}}_{\text{radial part}} + \frac{1}{r^2}\underbrace{\left\{\frac{1}{(\sin\psi)^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial\psi^2} + \cot\psi\frac{\partial}{\partial\psi}\right\}}_{\text{spherical part (Laplace-Beltrami operator)}}.$$

The Laplace equation for $u = u(r)$ becomes

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} = 0$$

whose general integral is

$$u(r) = \frac{C}{r} + C_1 \qquad C, C_1 \text{ arbitrary constants.}$$

Choose $C_1 = 0$ and $C = \frac{1}{4\pi}$ if $n = 3$, $C = -\frac{1}{2\pi}$ if $n = 2$. The function

$$\Phi(\mathbf{x}) = \begin{cases} -\frac{1}{2\pi}\log|\mathbf{x}| & n = 2 \\ \dfrac{1}{4\pi|\mathbf{x}|} & n = 3 \end{cases} \tag{3.39}$$

is called the **fundamental solution** for the Laplace operator Δ. As we shall prove in Chapter 7, the above choice of the constant C is made in order to have

$$\Delta\Phi(\mathbf{x}) = -\delta(\mathbf{x})$$

where $\delta(\mathbf{x})$ denotes *the Dirac measure at* $\mathbf{x} = \mathbf{0}$.

The physical meaning of Φ is remarkable: if $n = 3$, in standard units, $4\pi\Phi$ represents the electrostatic potential due to a unitary charge located at the origin and vanishing at infinity[15].

Clearly, if the origin is replaced by a point $\mathbf{y}$, the corresponding potential is $\Phi(\mathbf{x} - \mathbf{y})$ and

$$\Delta_{\mathbf{x}}\Phi(\mathbf{x} - \mathbf{y}) = -\delta(\mathbf{x} - \mathbf{y}).$$

By symmetry, we also have $\Delta_{\mathbf{y}}\Phi(\mathbf{x} - \mathbf{y}) = -\delta(\mathbf{x} - \mathbf{y})$.

Remark 3.5. In dimension $n > 3$, the fundamental solution of the Laplace operator is $\Phi(\mathbf{x}) = \omega_n^{-1}|\mathbf{x}|^{2-n}$.

---

[14] Appendix D.

[15] In dimension 2,

$$2\pi\Phi(x_1, x_2) = -\log\sqrt{x_1^2 + x_2^2}$$

represents the potential due to a charge of density 1, distributed along the x_3 axis.

3.4.2 The Newtonian potential

Suppose that $f(\mathbf{x})$ is the density of a charge located inside a compact set in $\mathbb{R}^3$. Then $\Phi(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) d\mathbf{y}$ represents the potential at $\mathbf{x}$ due to the charge $f(\mathbf{y}) d\mathbf{y}$ inside a small region of volume $d\mathbf{y}$ around $\mathbf{y}$. The full potential is given by the sum of all the contributions; we get

$$u(\mathbf{x}) = \int_{\mathbb{R}^3} \Phi(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{f(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|} d\mathbf{y} \tag{3.40}$$

which is the *convolution between* f *and* Φ and it is called the **Newtonian potential** of f. Formally, we have

$$\Delta u(\mathbf{x}) = \int_{\mathbb{R}^3} \Delta_{\mathbf{x}} \Phi(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} = -\int_{\mathbb{R}^3} \delta(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} = -f(\mathbf{x}). \tag{3.41}$$

Under suitable hypotheses on f, (3.41) is indeed true (see Theorem 3.9 below). Clearly, u is not the only solution of $\Delta v = -f$, since $u + c$, c constant, is a solution as well. However, the Newtonian potential is the only solution vanishing at infinity. All this is stated precisely in the theorem below, where, for simplicity, we assume $f \in C^2(\mathbb{R}^3)$ with compact support[16]. We have:

Theorem 3.9. *Let* $f \in C^2(\mathbb{R}^3)$ *with* **compact** *support. Let* u *be the Newtonian potential of* f, *defined by (3.40). Then,* u *is the only solution in* $\mathbb{R}^3$ *of*

$$\Delta u = -f \tag{3.42}$$

belonging to $C^2(\mathbb{R}^3)$ *and vanishing at infinity.*

Proof. The uniqueness part follows from Liouville's Theorem. Let $v \in C^2(\mathbb{R}^3)$ another solution to (3.42), vanishing at infinity. Then $u - v$ is a *bounded* harmonic function in all $\mathbb{R}^3$ and therefore is constant. Since it vanishes at infinity it must be zero; thus $u = v$.

To show that (3.40) belongs to $C^2(\mathbb{R}^3)$ and satisfies (3.42), observe that we can write (3.40) in the alternative form

$$u(\mathbf{x}) = \int_{\mathbb{R}^3} \Phi(\mathbf{y}) f(\mathbf{x} - \mathbf{y}) \, d\mathbf{y} = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{f(\mathbf{x} - \mathbf{y})}{|\mathbf{y}|} d\mathbf{y}.$$

Since $1/|\mathbf{y}|$ is integrable near the origin and f is zero outside a compact set, we can take first and second order derivatives under the integral sign to get

$$u_{x_j x_k}(\mathbf{x}) = \int_{\mathbb{R}^3} \Phi(\mathbf{y}) f_{x_j x_k}(\mathbf{x} - \mathbf{y}) \, d\mathbf{y}. \tag{3.43}$$

---

[16] Recall that the *support* of a continuous function f is the *closure of the set where* f *is not zero.*

Since $f_{x_j x_k} \in C\left(\mathbb{R}^3\right)$, formula (3.43) shows that also $u_{x_j x_k}$ is continuous and therefore $u \in C^2\left(\mathbb{R}^3\right)$.

It remains to prove (3.42). Since $\Delta_{\mathbf{x}} f\left(\mathbf{x} - \mathbf{y}\right) = \Delta_{\mathbf{y}} f\left(\mathbf{x} - \mathbf{y}\right)$, from (3.43), we have

$$\Delta u(\mathbf{x}) = \int_{\mathbb{R}^3} \Phi\left(\mathbf{y}\right) \Delta_{\mathbf{x}} f\left(\mathbf{x} - \mathbf{y}\right) d\mathbf{y} = \int_{\mathbb{R}^3} \Phi\left(\mathbf{y}\right) \Delta_{\mathbf{y}} f\left(\mathbf{x} - \mathbf{y}\right) d\mathbf{y}.$$

We want to integrate by parts using formula (1.13) but since Φ has a singularity at $\mathbf{y} = \mathbf{0}$, we have first to isolate the origin, by choosing a small ball $B_r = B_r\left(\mathbf{0}\right)$ and writing

$$\Delta u(\mathbf{x}) = \int_{B_r(\mathbf{0})} \cdots\, d\mathbf{y} + \int_{\mathbb{R}^3 \setminus B_r(\mathbf{0})} \cdots\, d\mathbf{y} \equiv \mathbf{I}_r + \mathbf{J}_r. \tag{3.44}$$

We have, using spherical coordinates,

$$|\mathbf{I}_r| \le \frac{\max|\Delta f|}{4\pi} \int_{B_r(\mathbf{0})} \frac{1}{|\mathbf{y}|}\, d\mathbf{y} = \max|\Delta f| \int_0^r \rho\, d\rho = \frac{\max|\Delta f|}{2} r^2$$

so that

$$\mathbf{I}_r \to 0 \qquad \text{if } r \to 0.$$

Keeping in mind that f vanishes outside a compact set, we can integrate $\mathbf{J}_r$ by parts (twice), obtaining

$$\mathbf{J}_r = \frac{1}{4\pi r} \int_{\partial B_r} \nabla_{\boldsymbol{\sigma}} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}\, d\sigma - \int_{\mathbb{R}^3 \setminus B_r(\mathbf{0})} \nabla \Phi\left(\mathbf{y}\right) \cdot \nabla_{\mathbf{y}} f\left(\mathbf{x} - \mathbf{y}\right)\, d\mathbf{y}$$

$$= \frac{1}{4\pi r} \int_{\partial B_r} \nabla_{\boldsymbol{\sigma}} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}\, d\sigma - \int_{\partial B_r} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) \nabla \Phi\left(\boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}\, d\sigma$$

since $\Delta \Phi = 0$ in $\mathbb{R}^3 \setminus B_r\left(\mathbf{0}\right)$. We have:

$$\frac{1}{4\pi r} \left|\int_{\partial B_r} \nabla_{\boldsymbol{\sigma}} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}\, d\sigma\right| \le r \max|\nabla f| \to 0 \qquad \text{as } r \to 0.$$

On the other hand, $\nabla \Phi\left(\mathbf{y}\right) = -\mathbf{y}\,|\mathbf{y}|^{-3}$ and the outward pointing[17] unit normal on ∂B_r is $\boldsymbol{\nu}_{\boldsymbol{\sigma}} = -\boldsymbol{\sigma}/r$, so that

$$\int_{\partial B_r} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) \nabla \Phi\left(\boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}\, d\sigma = \frac{1}{4\pi r^2} \int_{\partial B_r} f\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma \to f\left(\mathbf{x}\right) \qquad \text{as } r \to 0.$$

Thus, $\mathbf{J}_r \to -f\left(\mathbf{x}\right)$ as $r \to 0$. Passing to the limit as $r \to 0$ in (3.44) we get $\Delta u\left(\mathbf{x}\right) = -f\left(\mathbf{x}\right)$. $\square$

Remark 3.6. Theorem 3.9 actually holds under much less restrictive hypotheses on f. For instance, it is enough that $f \in C^1\left(\mathbb{R}^3\right)$ and $|f\left(\mathbf{x}\right)| \le C\,|\mathbf{x}|^{-3-\varepsilon}$, $\varepsilon > 0$.

---

[17] With respect to $\mathbb{R}^3 \setminus B_r$.

Remark 3.7. An appropriate version of Theorem 3.9 holds in dimension $n = 2$, with the Newtonian potential replaced by the *logarithmic potential*

$$u(\mathbf{x}) = \int_{\mathbb{R}^2} \Phi(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} = -\frac{1}{2\pi} \int_{\mathbb{R}^2} \log|\mathbf{x} - \mathbf{y}| \ f(\mathbf{y}) \, d\mathbf{y}. \qquad (3.45)$$

The logarithmic potential does not vanish at infinity; its asymptotic behavior is (see Problem 3.11)

$$u(\mathbf{x}) = -\frac{M}{2\pi} \log|\mathbf{x}| + O\left(\frac{1}{|\mathbf{x}|}\right) \qquad \text{as } |\mathbf{x}| \to +\infty \qquad (3.46)$$

where

$$M = \int_{\mathbb{R}^2} f(\mathbf{y}) \, d\mathbf{y}.$$

Indeed, the logarithmic potential is the only solution of $\Delta u = -f$ in $\mathbb{R}^2$ satisfying (3.46).

3.4.3 A divergence-curl system. Helmholtz decomposition formula

Using the properties of the Newtonian potential we can solve the following two problems, that appear in several applications e.g. to linear elasticity, fluid dynamics or electrostatics.

(**1**) *Reconstruction of a vector field in $\mathbb{R}^3$ from the knowledge of its **curl** and **divergence***. Precisely, given a scalar f and a vector field $\boldsymbol{\omega}$, we want to find a vector field $\mathbf{u}$ such that

$$\begin{cases} \text{div } \mathbf{u} = f \\ \text{curl } \mathbf{u} = \boldsymbol{\omega} \end{cases} \qquad \text{in } \mathbb{R}^3.$$

We assume that $\mathbf{u}$ has continuous second derivatives and vanishes at infinity, as it is required in most applications.

(**2**) *Decomposition of a vector field $\mathbf{u} \in \mathbb{R}^3$ into the sum of a **divergence free** vector field and a **curl free** vector field*. Precisely, given $\mathbf{u}$, we want to find φ and a vector field $\mathbf{w}$ such that the following *Helmholtz decomposition formula* holds

$$\mathbf{u} = \nabla\varphi + \text{curl } \mathbf{w}. \qquad (3.47)$$

Consider problem (**1**). First of all observe that, since div curl $\mathbf{u} = 0$, a necessary condition for the existence of a solution is div $\boldsymbol{\omega} = 0$.

Let us check uniqueness. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are solutions sharing the same data f and $\boldsymbol{\omega}$, their difference $\mathbf{w} = \mathbf{u}_1 - \mathbf{u}_2$ vanishes at infinity and satisfies

$$\text{div } \mathbf{w} = 0 \qquad \text{and} \qquad \text{curl } \mathbf{w} = \mathbf{0} \qquad \text{in } \mathbb{R}^3.$$

From curl $\mathbf{w} = \mathbf{0}$ we infer the existence of a scalar function U such that

$$\nabla U = \mathbf{w}.$$

From div $\mathbf{w} = 0$ we deduce

$$\text{div}\nabla U = \Delta U = 0.$$

Thus U is harmonic. Hence its derivatives, i.e. the components w_j of $\mathbf{w}$, are bounded harmonic functions in $\mathbb{R}^3$. Liouville's theorem implies that each w_j is constant and therefore identically zero since it vanishes at infinity. We conclude that, under the stated assumptions, the *solution of problem* (**1**) *is unique*.

To find $\mathbf{u}$, split it into $\mathbf{u} = \mathbf{v} + \mathbf{z}$ and look for $\mathbf{v}$ and $\mathbf{z}$ such that

$$\begin{aligned}
\text{div } \mathbf{z} &= 0 & \text{curl } \mathbf{z} &= \boldsymbol{\omega} \\
\text{div } \mathbf{v} &= f & \text{curl } \mathbf{v} &= \mathbf{0}.
\end{aligned}$$

As before, from curl $\mathbf{v} = \mathbf{0}$ we infer the existence of a scalar function φ such that $\nabla \varphi = \mathbf{v}$, while div $\mathbf{v} = f$ implies $\Delta \varphi = f$. We have seen that, under suitable hypotheses on f, φ is given by the Newtonian potential of f, that is:

$$\varphi(\mathbf{x}) = -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} f(\mathbf{y}) \, d\mathbf{y}$$

and $\mathbf{v} = \nabla\varphi$. To find $\mathbf{z}$, recall the identity

$$\text{curl curl } \mathbf{z} = \nabla(\text{div } \mathbf{z}) - \Delta\mathbf{z}.$$

Since div $\mathbf{z} = 0$, we get

$$\Delta\mathbf{z} = -\text{curl curl } \mathbf{z} = \text{curl } \boldsymbol{\omega}$$

so that

$$\mathbf{z}(\mathbf{x}) = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} \text{curl } \boldsymbol{\omega}(\mathbf{y}) \, d\mathbf{y}.$$

Let us summarize the conclusions in the next theorem, also specifying the hypotheses[18] on f and $\boldsymbol{\omega}$.

Theorem 3.10. *Let $f \in C^1(\mathbb{R}^3)$, $\boldsymbol{\omega} \in C^2(\mathbb{R}^3)$ such that div $\boldsymbol{\omega} = 0$ and, for $|\mathbf{x}|$ large,*

$$|f(\mathbf{x})| \le \frac{M}{|\mathbf{x}|^{3+\varepsilon}}, \qquad |\text{curl}\boldsymbol{\omega}(\mathbf{x})| \le \frac{M}{|\mathbf{x}|^{3+\varepsilon}} \qquad (\varepsilon > 0).$$

Then, the unique solution vanishing at infinity of the system

$$\begin{cases} \text{div } \mathbf{u} = f \\ \text{curl } \mathbf{u} = \boldsymbol{\omega} \end{cases} \qquad in \ \mathbb{R}^3$$

is given by vector field

$$\mathbf{u}(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} \text{curl } \boldsymbol{\omega}(\mathbf{y}) \, d\mathbf{y} - \nabla \int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} f(\mathbf{y}) \, d\mathbf{y}. \qquad (3.48)$$

---

[18] See Remark 3.6.

Consider now problem (**2**). If **u**, div **u** and curl **u** satisfy the hypotheses of theorem 3.10, we can write

$$\mathbf{u}(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} \text{curl curl } \mathbf{u}\,(\mathbf{y})\ d\mathbf{y} - \nabla \int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|}\ \text{div } \mathbf{u}\,(\mathbf{y})\,d\mathbf{y}.$$

Since **u** is rapidly vanishing at infinity, we have[19]

$$\int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} \text{curl curl } \mathbf{u}\,(\mathbf{y})\,d\mathbf{y} = \text{curl} \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} \text{curl } \mathbf{u}\,(\mathbf{y})\,d\mathbf{y}. \tag{3.49}$$

We conclude that

$$\mathbf{u} = \nabla \varphi + \text{curl } \mathbf{w}$$

where

$$\varphi\,(\mathbf{x}) = -\int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} \text{div } \mathbf{u}\,(\mathbf{y})\,d\mathbf{y}$$

and

$$\mathbf{w}\,(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} \text{curl } \mathbf{u}\,(\mathbf{y})\,d\mathbf{y}.$$

• *An application to fluid dynamics.* Consider the three dimensional flow of an incompressible fluid of constant density ρ and viscosity μ, subject to a conservative external force[20] $\mathbf{F} = \nabla f$. If $\mathbf{u} = \mathbf{u}\,(\mathbf{x},t)$ denotes the velocity field and $p = p\,(\mathbf{x},t)$ is the hydrostatic pressure, the laws of conservation of mass and linear momentum give for **u** and p the celebrated *Navier-Stokes equations*:

$$\text{div } \mathbf{u} = 0 \tag{3.50}$$

and

$$\frac{D\mathbf{u}}{Dt} = \mathbf{u}_t + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu \Delta \mathbf{u} + \frac{1}{\rho}\nabla f \qquad (\nu = \mu/\rho). \tag{3.51}$$

We look for solution of (3.50), (3.51) subject to a given initial condition

$$\mathbf{u}\,(\mathbf{x},0) = \mathbf{g}\,(\mathbf{x}) \qquad \mathbf{x} \in \mathbb{R}^3, \tag{3.52}$$

where **g** is also divergence free:

$$\text{div } \mathbf{g} = 0.$$

---

[19] In fact, if $|g\,(\mathbf{x})| \leq \frac{M}{|\mathbf{x}|^{3+\varepsilon}}$, one can show that

$$\frac{\partial}{\partial x_j} \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} g\,(\mathbf{y})\,d\mathbf{y} = \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} \frac{\partial g}{\partial y_j}\,d\mathbf{y}.$$

[20] Gravity, for instance.

The quantity $\frac{D\mathbf{u}}{Dt}$ is called the *material derivative of* $\mathbf{u}$, given by the sum of $\mathbf{u}_t$, the fluid acceleration due to the non-stationary character of the motion, and of $(\mathbf{u}\cdot\nabla)\,\mathbf{u}$, the inertial acceleration due to fluid transport[21].

In general, the system (3.50), (3.51) is extremely difficult to solve. In the case of slow flow, for instance due to high viscosity, the inertial term becomes negligible, compared for instance to $\nu\Delta\mathbf{u}$, and (3.51) simplifies to the linearized equation

$$\mathbf{u}_t = -\frac{1}{\rho}\nabla p + \nu\Delta\mathbf{u} + \nabla f. \tag{3.53}$$

It is possible to find an explicit formula for the solution of (3.50), (3.52), (3.53) by writing everything in terms of $\boldsymbol{\omega} = $curl $\mathbf{u}$. In fact, taking the curl of (3.53) and (3.52), we obtain, since curl$(\nabla p + \nu\Delta\mathbf{u} + \nabla f) = \nu\Delta\boldsymbol{\omega}$,

$$\begin{cases} \boldsymbol{\omega}_t = \nu\Delta\boldsymbol{\omega} & \mathbf{x}\in\mathbb{R}^3,\ t>0 \\ \boldsymbol{\omega}\,(\mathbf{x},0) = \text{curl } \mathbf{g}\,(\mathbf{x}) & \mathbf{x}\in\mathbb{R}^3. \end{cases}$$

This is a global Cauchy problem for the heat equation. If $\mathbf{g}\in C^2\left(\mathbb{R}^3\right)$ and curl $\mathbf{g}$ is bounded, we have

$$\boldsymbol{\omega}\,(\mathbf{x},t) = \frac{1}{(4\pi\nu t)^{3/2}}\int_{\mathbb{R}^3}\exp\left(-\frac{|\mathbf{y}|^2}{4\nu t}\right)\text{curl } \mathbf{g}\,(\mathbf{x}-\mathbf{y})\ d\mathbf{y}. \tag{3.54}$$

Moreover, for $t>0$, we can take the divergence operator under the integral in (3.54) and deduce that div $\boldsymbol{\omega} =0$. Therefore, if curl $\mathbf{g}\,(\mathbf{x})$ vanishes rapidly at infinity[22], we can recover $\mathbf{u}$ by solving the system

$$\text{curl } \mathbf{u} = \boldsymbol{\omega}, \quad \text{div } \mathbf{u} =0,$$

according to formula (3.48) with $f=0$.

Finally to find the pressure, from (3.53) we have

$$\nabla p = -\rho\mathbf{u}_t + \mu\Delta\mathbf{u} - \nabla f. \tag{3.55}$$

Since $\boldsymbol{\omega}_t = \nu\Delta\boldsymbol{\omega}$, the right hand side has zero curl; hence (3.55) can be solved and determines p up to an additive constant (as it should be).

In conclusion: *Let, $f\in C^1\left(\mathbb{R}^3\right)$, $\mathbf{g}\in C^2\left(\mathbb{R}^3\right)$, with div $\mathbf{g}=0$ and curl $\mathbf{g}$ rapidly vanishing at infinity. There exist a unique $\mathbf{u}\in C^2\left(\mathbb{R}^3\right)$, with curl $\mathbf{u}$ vanishing at infinity, and $p\in C^1\left(\mathbb{R}^3\right)$ unique up to an additive constant, satisfying the system* (3.50), (3.52), (3.53).

---

[21] The $i-$component of $(\mathbf{v}\cdot\nabla)\,\mathbf{v}$ is given by $\sum_{j=1}^3 v_j\frac{\partial v_i}{\partial x_j}$. Let us compute $\frac{D\mathbf{v}}{Dt}$, for example, for a plane fluid uniformly rotating with angular speed ω. Then $\mathbf{v}\,(x,y) = -\omega y\mathbf{i}+\omega x\mathbf{j}$. Since $\mathbf{v}_t = \mathbf{0}$, the motion is stationary and

$$\frac{D\mathbf{v}}{Dt} = (\mathbf{v}\cdot\nabla)\,\mathbf{v} = \left(-\omega y\frac{\partial}{\partial x}+\omega x\frac{\partial}{\partial y}\right)(-\omega y\mathbf{i}+\omega x\mathbf{j}) = -\omega^2\,(-x\mathbf{i}+y\mathbf{j})$$

which is centrifugal acceleration.

[22] $|\text{curl } \mathbf{g}\,(\mathbf{x})| \le M/\left|\mathbf{x}\right|^{3+\varepsilon}$, $\varepsilon>0$, it is enough.

3.5 The Green Function

3.5.1 An integral identity

Formula (3.40) gives a representation of the solution to Poisson's equation in all $\mathbb{R}^3$. In bounded domains, any representation formula has to take into account the boundary values, as indicated in the following theorem.

Theorem 3.11. *Let $\Omega \subset \mathbb{R}^n$ be a smooth, bounded domain and $u \in C^2\left(\overline{\Omega}\right)$. Then, for every $\mathbf{x} \in \Omega$,*

$$u(\mathbf{x}) = -\int_{\Omega} \Phi(\mathbf{x} - \mathbf{y}) \, \Delta u(\mathbf{y}) \, d\mathbf{y} +$$
$$+ \int_{\partial\Omega} \Phi(\mathbf{x} - \boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_\sigma} u(\boldsymbol{\sigma}) \, d\sigma - \int_{\partial\Omega} u(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma}) \, d\sigma \tag{3.56}$$

The last two terms in the right hand side of (3.56) are called *single* and *double layer potentials*, respectively. We are going to examine these surface potentials later. The first one is the Newtonian potential of $-\Delta u$ in Ω.

Proof. We give it for $n = 3$. Fix $\mathbf{x} \in \Omega$, and consider the fundamental solution

$$\Phi(\mathbf{x} - \mathbf{y}) = \frac{1}{4\pi r_{\mathbf{xy}}} \qquad r_{\mathbf{xy}} = |\mathbf{x} - \mathbf{y}|$$

as a function of $\mathbf{y}$: we write $\Phi(\mathbf{x} - \cdot)$.

We would like to apply *Green's identity* (1.15)

$$\int_{\Omega}(v\Delta u - u\Delta v)d\mathbf{x} = \int_{\partial\Omega}(v\partial_{\boldsymbol{\nu}} u - u\partial_{\boldsymbol{\nu}} v)d\sigma \tag{3.57}$$

to u and $\Phi(\mathbf{x} - \cdot)$. However, $\Phi(\mathbf{x} - \cdot)$ has a singularity in $\mathbf{x}$, so that it cannot be inserted directly into (3.57). Let us isolate the singularity inside a ball $B_\varepsilon(\mathbf{x})$, with ε small. In the domain $\Omega_\varepsilon = \Omega \backslash \overline{B}_\varepsilon(\mathbf{x})$, $\Phi(\mathbf{x} - \cdot)$ is smooth and harmonic.

Thus, replacing Ω with Ω_ε, we can apply (3.57) to u and $\Phi(\mathbf{x} - \cdot)$. Since

$$\partial\Omega_\varepsilon = \partial\Omega \cup \partial B_\varepsilon(\mathbf{x}),$$

and $\Delta_{\mathbf{y}}\Phi(\mathbf{x} - \mathbf{y}) = 0$, we find:

$$\int_{\Omega_\varepsilon} \frac{1}{r_{\mathbf{xy}}}\Delta u \, d\mathbf{y} = \int_{\partial\Omega_\varepsilon} \left(\frac{1}{r_{\mathbf{x\sigma}}} \frac{\partial u}{\partial \boldsymbol{\nu}_\sigma} - u \frac{\partial}{\partial \boldsymbol{\nu}_\sigma} \frac{1}{r_{\mathbf{x\sigma}}} \right) d\sigma$$
$$= \int_{\partial\Omega}(\cdots) \, d\sigma + \int_{\partial B_\varepsilon(\mathbf{x})} \frac{1}{r_{\mathbf{x\sigma}}} \frac{\partial u}{\partial \boldsymbol{\nu}_\sigma} \, d\sigma - \int_{\partial B_\varepsilon(\mathbf{x})} u \frac{\partial}{\partial \boldsymbol{\nu}_\sigma} \frac{1}{r_{\mathbf{x\sigma}}} \, d\sigma. \tag{3.58}$$

We let now $\varepsilon \to 0$ in (3.58). We have:

$$\int_{\Omega_\varepsilon} \frac{1}{r_{\mathbf{x\sigma}}}\Delta u \, d\mathbf{y} \to \int_{\Omega} \frac{1}{r_{\mathbf{x\sigma}}}\Delta u \, d\mathbf{y} \qquad \text{as } \varepsilon \to 0 \tag{3.59}$$

since $\Delta u \in C\left(\overline{\Omega}\right)$ and $r_{\mathbf{x\sigma}}^{-1}$ is positive and integrable in Ω.

On $\partial B_\varepsilon(\mathbf{x})$, we have $r_{\mathbf{x}\boldsymbol{\sigma}} = \varepsilon$ and $|\partial_{\boldsymbol{\nu}} u| \leq M$, since $|\nabla u|$ is bounded; then

$$\left| \int_{\partial B_\varepsilon(\mathbf{x})} \frac{1}{r_{\mathbf{x}\boldsymbol{\sigma}}} \frac{\partial u}{\partial \boldsymbol{\nu_\sigma}} \, d\sigma \right| \leq 4\pi\varepsilon M \to 0 \qquad \text{as } \varepsilon \to 0. \tag{3.60}$$

The most delicate term is

$$\int_{\partial B_\varepsilon(\mathbf{x})} u \, \frac{\partial}{\partial \boldsymbol{\nu_\sigma}} \frac{1}{r_{\mathbf{x}\boldsymbol{\sigma}}} \, d\sigma.$$

On $\partial B_\varepsilon(\mathbf{x})$, the outward pointing (with respect to Ω_ε) unit normal at $\boldsymbol{\sigma}$ is $\boldsymbol{\nu_\sigma} = \frac{\mathbf{x} - \boldsymbol{\sigma}}{\varepsilon}$, so that

$$\frac{\partial}{\partial \boldsymbol{\nu_\sigma}} \frac{1}{r_{\mathbf{x}\boldsymbol{\sigma}}} = \nabla_{\mathbf{y}} \frac{1}{r_{\mathbf{x}\boldsymbol{\sigma}}} \cdot \boldsymbol{\nu_\sigma} = \frac{\mathbf{x} - \boldsymbol{\sigma}}{\varepsilon^3} \frac{\mathbf{x} - \boldsymbol{\sigma}}{\varepsilon} = \frac{1}{\varepsilon^2}.$$

As a consequence,

$$\int_{\partial B_\varepsilon(\mathbf{x})} u \, \frac{\partial}{\partial \boldsymbol{\nu_\sigma}} \frac{1}{r_{\mathbf{x}\boldsymbol{\sigma}}} \, d\sigma = \frac{1}{\varepsilon^2} \int_{\partial B_\varepsilon(\mathbf{x})} u \, d\sigma \to 4\pi \, u(\mathbf{x}) \tag{3.61}$$

as $\varepsilon \to 0$, by the continuity of u.

Letting $\varepsilon \to 0$ in (3.58), from (3.59), (3.60), (3.61) we obtain (3.56). $\square$

3.5.2 The Green function

The function Φ defined in (3.39) is the fundamental solution for the Laplace operator Δ in all $\mathbb{R}^n$ ($n = 2, 3$). We can also define a fundamental solution for the Laplace operator in any open set and in particular in any *bounded* domain $\Omega \subset \mathbb{R}^n$, representing the potential due to a unit charge placed at a point $\mathbf{x} \in \Omega$ and equal to zero on $\partial\Omega$.

This function, that we denote by $G(\mathbf{x}, \mathbf{y})$, is called the *Green function in Ω*, for the operator Δ; for fixed $\mathbf{x} \in \Omega$, G satisfies

$$\Delta_{\mathbf{y}} G(\mathbf{x}, \mathbf{y}) = -\delta_{\mathbf{x}} \qquad \text{in } \Omega$$

and

$$G(\mathbf{x}, \boldsymbol{\sigma}) = 0, \qquad \boldsymbol{\sigma} \in \partial\Omega.$$

More explicitly, the Green's function can be written in the form

$$G(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y}) - \varphi(\mathbf{x}, \mathbf{y})$$

where φ, for fixed $\mathbf{x} \in \Omega$, solves the Dirichlet problem

$$\begin{cases} \Delta_{\mathbf{y}} \varphi = 0 & \text{in } \Omega \\ \varphi(\mathbf{x}, \boldsymbol{\sigma}) = \Phi(\mathbf{x} - \boldsymbol{\sigma}) & \text{on } \partial\Omega. \end{cases} \tag{3.62}$$

Two important properties of the Green function are the following (see Problem 3.14):

(a) *Positivity:* $G(\mathbf{x}, \mathbf{y}) > 0$ for every $\mathbf{x}, \mathbf{y} \in \Omega$, with $G(\mathbf{x}, \mathbf{y}) \to +\infty$ when $\mathbf{x} - \mathbf{y} \to \mathbf{0}$;

(b) *Symmetry:* $G(\mathbf{x}, \mathbf{y}) = G(\mathbf{y}, \mathbf{x})$.

The existence of the Green function for a particular domain depends on the solvability of the Dirichlet problem (3.62). From Theorem 3.8, we know that this is the case if Ω is a Lipschitz domain, for instance.

Even if we know that the Green function exists, explicit formulas are available only for special domains. Sometimes a technique known as *method of electrostatic images* works. In this method $\varphi(\mathbf{x}, \cdot)$ is considered as the potential due to an imaginary charge q placed at a suitable point $\mathbf{x}^*$, the *image of* $\mathbf{x}$, in the complement of Ω. The charge q and the point $\mathbf{x}^*$ have to be chosen so that $\varphi(\mathbf{x}, \cdot)$ on $\partial \Omega$ is equal to the potential created by the unit charge in $\mathbf{x}$.

The simplest way to illustrate the method is to find the Green function for the upper half-space, although this is an unbounded domain. Clearly, we require that G vanishes at infinity.

• *Green's function for the upper half space in* $\mathbb{R}^3$. Let $\mathbb{R}^3_+$ be the upper half space :

$$\mathbb{R}^3_+ = \{(x_1, x_2, x_3) : x_3 > 0\}.$$

Fix $\mathbf{x} = (x_1, x_2, x_3)$ and observe that if we choose $\mathbf{x}^* = (x_1, x_2, -x_3)$ then, on $y_3 = 0$ we have

$$|\mathbf{x}^* - \mathbf{y}| = |\mathbf{x} - \mathbf{y}|.$$

Thus, if $\mathbf{x} \in \mathbb{R}^3_+$, $\mathbf{x}^*$ belongs to the complement of $\mathbb{R}^3_+$, the function

$$\varphi(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x}^* - \mathbf{y}) = \frac{1}{4\pi |\mathbf{x}^* - \mathbf{y}|}$$

is harmonic in $\mathbb{R}^3_+$ and $\varphi(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y})$ on the plane $y_3 = 0$. In conclusion,

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} - \frac{1}{4\pi |\mathbf{x}^* - \mathbf{y}|} \tag{3.63}$$

is the Green function for the upper half space.

• *Green function for sphere.* Let $\Omega = B_R = B_R(\mathbf{0}) \subset \mathbb{R}^3$. To find the Green function for B_R, set

$$\varphi(\mathbf{x}, \mathbf{y}) = \frac{q}{4\pi |\mathbf{x}^* - \mathbf{y}|},$$

$\mathbf{x}$ fixed in B_R, and try to determine $\mathbf{x}^*$, outside B_R, and q, so that

$$\frac{q}{4\pi |\mathbf{x}^* - \mathbf{y}|} = \frac{1}{4\pi |\mathbf{x} - \mathbf{y}|} \tag{3.64}$$

when $|\mathbf{y}| = R$. The (3.64) gives

$$|\mathbf{x}^* - \mathbf{y}|^2 = q^2 |\mathbf{x} - \mathbf{y}|^2 \tag{3.65}$$

or

$$|\mathbf{x}^*|^2 - 2\mathbf{x}^* \cdot \mathbf{y} + R^2 = q^2(|\mathbf{x}|^2 - 2\mathbf{x} \cdot \mathbf{y} + R^2).$$

Rearranging the terms we have

$$|\mathbf{x}^*|^2 + R^2 - q^2(R^2 + |\mathbf{x}|^2) = 2\mathbf{y} \cdot (\mathbf{x}^* - q^2\mathbf{x}). \tag{3.66}$$

Since the left hand side does not depend on $\mathbf{y}$, it must be that $\mathbf{x}^* = q^2\mathbf{x}$ and

$$q^4 |\mathbf{x}|^2 - q^2(R^2 + |\mathbf{x}|^2) + R^2 = 0$$

from which $q = R/|\mathbf{x}|$. This works for $\mathbf{x} \neq \mathbf{0}$ and gives

$$G(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi}\left[\frac{1}{|\mathbf{x} - \mathbf{y}|} - \frac{R}{|\mathbf{x}|\,|\mathbf{x}^* - \mathbf{y}|}\right], \qquad \mathbf{x}^* = \frac{R^2}{|\mathbf{x}|^2}\mathbf{x},\; \mathbf{x} \neq \mathbf{0}. \tag{3.67}$$

Since

$$|\mathbf{x}^* - \mathbf{y}| = |\mathbf{x}|^{-1}\left(R^4 - 2R^2\mathbf{x} \cdot \mathbf{y} + \mathbf{y}\,|\mathbf{x}|^2\right)^{1/2},$$

when $\mathbf{x} \to \mathbf{0}$ we have

$$\varphi(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi}\frac{R}{|\mathbf{x}|\,|\mathbf{x}^* - \mathbf{y}|} \to \frac{1}{4\pi R}$$

and therefore we can define

$$G(\mathbf{0}, \mathbf{y}) = \frac{1}{4\pi}\left[\frac{1}{|\mathbf{y}|} - \frac{1}{R}\right].$$
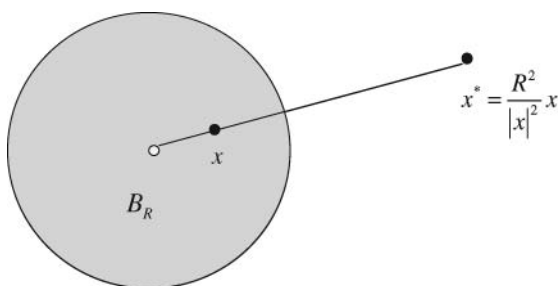
Fig. 3.6. The image $\mathbf{x}^*$ of $\mathbf{x}$ in the construction of the Green's function for the sphere

3.5.3 Green's representation formula

From Theorem 3.11 we know that every smooth function u can be written as the sum of a volume (Newtonian) potential with density $-\Delta u$, a single layer potential

of *density* $\partial_{\boldsymbol{\nu}} u$ and a double layer potential of *moment* u. Suppose u solves the Dirichlet problem

$$\begin{cases} \Delta u = f & \text{in } \Omega \\ \quad u = g & \text{on } \partial\Omega. \end{cases} \tag{3.68}$$

Then (3.56) gives, for $\mathbf{x} \in \Omega$,

$$u(\mathbf{x}) = -\int_{\Omega} \Phi(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} +$$
$$+ \int_{\partial\Omega} \Phi(\mathbf{x} - \boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} u(\boldsymbol{\sigma}) \, d\sigma - \int_{\partial\Omega} g(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} \Phi(\mathbf{x} - \boldsymbol{\sigma}) \, d\sigma. \tag{3.69}$$

This representation formula for u is not satisfactory, since it involves the data f and g but also the normal derivative $\partial_{\boldsymbol{\nu}_{\sigma}} u$, which is unknown. To get rid of $\partial_{\boldsymbol{\nu}_{\sigma}} u$, let $G(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y}) - \varphi(\mathbf{x}, \mathbf{y})$ be the Green function in Ω. Since $\varphi(\mathbf{x}, \cdot)$ is harmonic in Ω, we can apply (3.57) to u and $\varphi(\mathbf{x}, \cdot)$; we find

$$0 = \int_{\Omega} \varphi(\mathbf{x}, \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} +$$
$$- \int_{\partial\Omega} \varphi(\mathbf{x}, \boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} u(\boldsymbol{\sigma}) \, d\sigma + \int_{\partial\Omega} g(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} \varphi(\mathbf{x}, \boldsymbol{\sigma}) \, d\sigma. \tag{3.70}$$

Adding (3.69), (3.70) and recalling that $\varphi(\mathbf{x}, \boldsymbol{\sigma}) = \Phi(\mathbf{x} - \boldsymbol{\sigma})$ on $\partial\Omega$, we obtain:

Theorem 3.12. *Let Ω be a smooth domain and u be a smooth solution of (3.68). Then:*

$$u(\mathbf{x}) = -\int_{\Omega} f(\mathbf{y}) \, G(\mathbf{x}, \mathbf{y}) \, d\mathbf{y} - \int_{\partial\Omega} g(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} G(\mathbf{x}, \boldsymbol{\sigma}) \, d\sigma. \tag{3.71}$$

Thus the solution of the Dirichlet problem (3.68) can be written as the sum of the two Green potentials in the right hand side of (3.71) and it is known as soon as the Green function in Ω is known. In particular, if u is harmonic, then

$$u(\mathbf{x}) = -\int_{\partial\Omega} g(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_{\sigma}} G(\mathbf{x}, \boldsymbol{\sigma}) \, d\sigma. \tag{3.72}$$

Comparing with (3.37), we deduce that

$$-\partial_{\boldsymbol{\nu}_{\sigma}} G(\mathbf{x}, \boldsymbol{\sigma}) \, d\sigma$$

represents the *harmonic measure* in Ω. The function

$$P(\mathbf{x}, \boldsymbol{\sigma}) = -\partial_{\boldsymbol{\nu}_{\sigma}} G(\mathbf{x}, \boldsymbol{\sigma})$$

is called **Poisson's kernel**. Since $G(\cdot, \boldsymbol{\sigma}) > 0$ inside Ω and vanishes on Ω, P is *nonnegative* (actually positive).

On the other hand, the formula

$$u(\mathbf{x}) = -\int_{\Omega} f(\mathbf{y}) \, G(\mathbf{x}, \mathbf{y}) \, d\mathbf{y}$$

gives the solution of the Poisson equation $\Delta u = f$ in Ω, vanishing on $\partial\Omega$. From the positivity of G we have that:

$$f \geq 0 \quad \text{in } \Omega \text{ implies } u \leq 0 \text{ in } \Omega,$$

which is another form of the maximum principle.

• *Poisson's kernel and Poisson's formula.* From (3.67) we can compute Poisson's kernel for the sphere $B_R\left(\mathbf{0}\right)$. We have, recalling that $\mathbf{x}^* = R^2\left|\mathbf{x}\right|^{-2}\mathbf{x}$, if $\mathbf{x} \neq \mathbf{0}$,

$$\nabla_{\mathbf{y}}\left[\frac{1}{\left|\mathbf{x}-\mathbf{y}\right|} - \frac{R}{\left|\mathbf{x}\right|\left|\mathbf{x}^*-\mathbf{y}\right|}\right] = \frac{\mathbf{x}-\mathbf{y}}{\left|\mathbf{x}-\mathbf{y}\right|^3} - \frac{R}{\left|\mathbf{x}\right|}\frac{\mathbf{x}^*-\mathbf{y}}{\left|\mathbf{x}^*-\mathbf{y}\right|^3}.$$

If $\boldsymbol{\sigma} \in \partial B_R\left(\mathbf{0}\right)$, from (3.65) we have $\left|\mathbf{x}^* - \boldsymbol{\sigma}\right| = R\left|\mathbf{x}\right|^{-1}\left|\mathbf{x} - \boldsymbol{\sigma}\right|$, therefore

$$\nabla_{\mathbf{y}}G\left(\mathbf{x},\boldsymbol{\sigma}\right) = \frac{1}{4\pi}\left[\frac{\mathbf{x}-\boldsymbol{\sigma}}{\left|\mathbf{x}-\boldsymbol{\sigma}\right|^3} - \frac{\left|\mathbf{x}\right|^2}{R^2}\frac{\mathbf{x}^*-\boldsymbol{\sigma}}{\left|\mathbf{x}-\boldsymbol{\sigma}\right|^3}\right] = \frac{-\boldsymbol{\sigma}}{4\pi\left|\mathbf{x}-\boldsymbol{\sigma}\right|^3}\left[1 - \frac{\left|\mathbf{x}\right|^2}{R^2}\right].$$

Since on $\partial B_R\left(\mathbf{0}\right)$ the exterior unit normal is $\boldsymbol{\nu}_{\boldsymbol{\sigma}} = \boldsymbol{\sigma}/R$, we have

$$P\left(\mathbf{x},\boldsymbol{\sigma}\right) = -\partial_{\boldsymbol{\nu}_{\boldsymbol{\sigma}}}G\left(\mathbf{x},\boldsymbol{\sigma}\right) = -\nabla_{\mathbf{y}}G\left(\mathbf{x},\boldsymbol{\sigma}\right)\cdot\boldsymbol{\nu}_{\boldsymbol{\sigma}} = \frac{R^2 - \left|\mathbf{x}\right|^2}{4\pi R}\frac{1}{\left|\mathbf{x}-\boldsymbol{\sigma}\right|^3}.$$

As a consequence, we obtain Poisson's formula

$$u\left(\mathbf{x}\right) = \frac{R^2 - \left|\mathbf{x}\right|^2}{4\pi R}\int_{\partial B_R(\mathbf{0})}\frac{g\left(\boldsymbol{\sigma}\right)}{\left|\mathbf{x}-\boldsymbol{\sigma}\right|^3}d\sigma \tag{3.73}$$

for the unique solution of the Dirichlet problem $\Delta u = 0$ in $B_R\left(\mathbf{0}\right)$ and $u = g$ on $\partial B_R\left(\mathbf{0}\right)$.

3.5.4 The Neumann function

We can find a representation formula for the solution of a Neumann problem as well. Let u be a smooth solution of the problem

$$\begin{cases} \Delta u = f & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}}u = h & \text{on } \partial\Omega \end{cases} \tag{3.74}$$

where f and h have to satisfy the solvability condition

$$\int_{\partial\Omega}h\left(\boldsymbol{\sigma}\right)d\sigma = \int_{\Omega}f\left(\mathbf{y}\right)d\mathbf{y}, \tag{3.75}$$

keeping in mind that u is uniquely determined up to an additive constant. From Theorem 3.11 we can write

$$u\left(\mathbf{x}\right) = -\int_{\Omega}\Phi\left(\mathbf{x}-\mathbf{y}\right)f\left(\mathbf{y}\right)d\mathbf{y} +$$
$$+\int_{\partial\Omega}h\left(\boldsymbol{\sigma}\right)\Phi\left(\mathbf{x}-\boldsymbol{\sigma}\right)d\sigma - \int_{\partial\Omega}u\left(\boldsymbol{\sigma}\right)\partial_{\boldsymbol{\nu}_{\boldsymbol{\sigma}}}\Phi\left(\mathbf{x}-\boldsymbol{\sigma}\right)d\sigma. \tag{3.76}$$

and this time we should get rid of the second integral, containing the unknown data u on $\partial\Omega$. Mimicking what we have done for the Dirichlet problem, we try to find an analog of the Green function, that is a function $N = N(\mathbf{x}, \mathbf{y})$ given by

$$N(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x} - \mathbf{y}) - \psi(\mathbf{x}, \mathbf{y})$$

where, for $\mathbf{x}$ fixed, ψ is a solution of

$$\begin{cases} \Delta_{\mathbf{y}} \psi = 0 & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}_\sigma} \psi(\mathbf{x}, \boldsymbol{\sigma}) = \partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma}) & \text{on } \partial\Omega, \end{cases}$$

in order to have $\partial_{\boldsymbol{\nu}_\sigma} N(\mathbf{x}, \boldsymbol{\sigma}) = 0$ on $\partial\Omega$. But this Neumann problem has no solution because the compatibility condition

$$\int_{\partial\Omega} \partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma})\, d\sigma = 0$$

is not satisfied. In fact, letting $u \equiv 1$ in (3.56), we get

$$\int_{\partial\Omega} \partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma})\, d\sigma = -1. \tag{3.77}$$

Thus, taking into account (3.77), we require ψ to satisfy

$$\begin{cases} \Delta_{\mathbf{y}} \psi = 0 & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}_\sigma} \psi(\mathbf{x}, \boldsymbol{\sigma}) = \partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma}) + \frac{1}{|\partial\Omega|} & \text{on } \partial\Omega. \end{cases} \tag{3.78}$$

In this way,

$$\int_{\partial\Omega} \left(\partial_{\boldsymbol{\nu}_\sigma} \Phi(\mathbf{x} - \boldsymbol{\sigma}) + \frac{1}{|\partial\Omega|} \right) d\sigma = 0$$

and (3.78) is solvable. Note that, with this choice of ψ, we have

$$\partial_{\boldsymbol{\nu}_\sigma} N(\mathbf{x}, \boldsymbol{\sigma}) = -\frac{1}{|\partial\Omega|} \quad \text{on } \partial\Omega. \tag{3.79}$$

Apply now (3.57) to u and $\psi(\mathbf{x}, \cdot)$; we find:

$$0 = -\int_{\partial\Omega} \psi(\mathbf{x}, \boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_\sigma} u(\boldsymbol{\sigma})\, d\sigma + \int_{\partial\Omega} h(\boldsymbol{\sigma}) \, \partial_{\boldsymbol{\nu}_\sigma} \psi(\boldsymbol{\sigma})\, d\sigma + \int_\Omega \psi(\mathbf{y}) f(\mathbf{y})\, d\mathbf{y}. \tag{3.80}$$

Adding (3.80) to (3.76) and using (3.79) we obtain:

Theorem 3.13. *Let Ω be a smooth domain and u be a smooth solution of (3.74). Then:*

$$u(\mathbf{x}) - \frac{1}{|\partial\Omega|} \int_{\partial\Omega} u(\boldsymbol{\sigma})\, d\sigma = \int_{\partial\Omega} h(\boldsymbol{\sigma}) N(\mathbf{x}, \boldsymbol{\sigma})\, d\sigma - \int_\Omega f(\mathbf{y}) N(\mathbf{x}, \mathbf{y})\, d\mathbf{y}.$$

Thus, the solution of the Neumann problem (3.74) can also be written as the sum of two potentials, up to the additive constant $c = \frac{1}{|\partial\Omega|} \int_{\partial\Omega} u(\boldsymbol{\sigma})\, d\sigma$, the mean value of u.

The function N is called *Neumann's function* (also Green's function for the Neumann problem) and it is defined up to an additive constant.

3.6 Uniqueness in Unbounded Domains

3.6.1 Exterior problems

Boundary value problems in unbounded domains occur in important applications, for instance in the motion of fluids past an obstacle, capacity problems or scattering of acoustic or electromagnetic waves.

As in the case of Poisson's equation in all $\mathbb{R}^n$, a problem in an unbounded domain requires suitable conditions at infinity to be well posed.

Consider for example the Dirichlet problem

$$\begin{cases} \Delta u = 0 & \text{in } |\mathbf{x}| > 1 \\ u = 0 & \text{on } |\mathbf{x}| = 1. \end{cases} \tag{3.81}$$

For every real number a,

$$u(\mathbf{x}) = a \log |\mathbf{x}| \quad \text{and} \quad u(\mathbf{x}) = a(1 - 1/|\mathbf{x}|)$$

are solutions to (3.81) in dimension two and three, respectively. Thus there is no uniqueness.

To restore uniqueness, a typical requirement in two dimensions is that u be bounded, while in three dimensions that $u(\mathbf{x})$ has a limit, say u_∞, as $|\mathbf{x}| \to \infty$: under these conditions, in both cases we select a unique solution.

Problem (3.81) is an *exterior Dirichlet problem*. Given a bounded domain Ω, we call *exterior of* Ω the set

$$\Omega_e = \mathbb{R}^n \backslash \overline{\Omega}.$$

Without loss of generality we will assume that $\mathbf{0} \in \Omega$ and for simplicity, we will consider only *connected* exterior sets, i.e. **exterior domains**. Note that $\partial \Omega_e = \partial \Omega$.

As we have seen in several occasions, maximum principles are very useful to prove uniqueness. In exterior three dimensional domains we have (for $n = 2$ see Problem 3.16):

Theorem 3.14. *Let $\Omega_e \subset \mathbb{R}^3$ be an exterior domain and $u \in C^2(\Omega_e) \cap C(\overline{\Omega}_e)$, be harmonic in Ω_e and vanishing as $|\mathbf{x}| \to \infty$. If $u \geq 0$ (resp. $u \leq 0$) on $\partial \Omega_e$ then $u \geq 0$ (resp. $u \leq 0$) in Ω_e.*

Proof. Let $u \geq 0$ on $\partial \Omega_e$. Fix $\varepsilon > 0$ and choose r_0 so large that $\Omega \subset \{|\mathbf{x}| < r\}$ and $u \geq -\varepsilon$ on $\{|\mathbf{x}| = r\}$, for every $r \geq r_0$. In the bounded set

$$\Omega_{e,r} = \Omega_e \cap \{|\mathbf{x}| < r\}$$

we can apply Theorem 3.5 and we get $u \geq -\varepsilon$ in this set. Since ε is arbitrary and r may be taken as large as we like, we deduce that $u \geq 0$ in Ω_e.

The argument for the case $u \leq 0$ on $\partial \Omega_e$ is similar and we leave the details to the reader. $\square$

An immediate consequence is the following uniqueness result in dimension $n = 3$ (for $n = 2$ see Problem 3.16):

Theorem 3.15. *(Exterior Dirichlet problem). Let $\Omega_e \subset \mathbb{R}^3$ be an exterior domain. Then there exists at most one solution $u \in C^2(\Omega_e) \cap C(\overline{\Omega}_e)$ of the Dirichlet problem*

$$\begin{cases} \Delta u = f & in\ \Omega_e \\ u = g & on\ \partial\Omega_e \\ u(\mathbf{x}) \to u_\infty & |\mathbf{x}| \to \infty. \end{cases} \tag{3.82}$$

Proof. Apply Theorem 3.14 to the difference of two solutions. $\square$

We point out another interesting consequence of Theorem 3.14 and Corollary 3.2: a harmonic function vanishing at infinity, for $|\mathbf{x}|$ large is controlled by the fundamental solution.

Actually, more is true:

Corollary 3.4. *Let u be harmonic in $\Omega_e \subset \mathbb{R}^3$ and $u(\mathbf{x}) \to 0$ as $|\mathbf{x}| \to \infty$. There exists r_0 such that, if $|\mathbf{x}| \geq r_0$,*

$$|u(\mathbf{x})| \leq \frac{M}{|\mathbf{x}|}, \qquad \left|u_{x_j}(\mathbf{x})\right| \leq \frac{M}{|\mathbf{x}|^2}, \qquad \left|u_{x_j x_k}(\mathbf{x})\right| \leq \frac{M}{|\mathbf{x}|^3} \tag{3.83}$$

where M depends on r_0.

Proof. Choose r_0 such that $|u(\mathbf{x})| \leq 1$ if $|\mathbf{x}| \geq r_0$. Let $w(\mathbf{x}) = u(\mathbf{x}) - r_0/|\mathbf{x}|$. Then w is harmonic for $|\mathbf{x}| \geq r_0$, $w(\mathbf{x}) \leq 0$ on $|\mathbf{x}| = r_0$ and vanishes at infinity. Then, by Theorem 3.14,

$$w(\mathbf{x}) \leq 0 \text{ in } \Omega_e \cap \{|\mathbf{x}| \geq r_0\}. \tag{3.84}$$

Setting $v(\mathbf{x}) = r_0/|\mathbf{x}| - u(\mathbf{x})$, a similar argument gives $v(\mathbf{x}) \geq 0$ in $\Omega_e \cap \{|\mathbf{x}| \geq r_0\}$. This and (3.84) imply $|u(\mathbf{x})| \leq r_0/|\mathbf{x}|$ in $\Omega_e \cap \{|\mathbf{x}| \geq r_0\}$.

The gradient bound follows from (3.31) and (3.83). In fact, choose $m \geq 2$ such that $(m-1)r_0 \leq |\mathbf{x}| \leq (m+1)r_0$. Then $\partial B_{(m-1)r_0}(\mathbf{x}) \subset \{|\mathbf{x}| \geq r_0\}$ and from (3.31)

$$\left|u_{x_j}(\mathbf{x})\right| \leq \frac{3}{(m-1)r_0} \max_{\partial B_{mr_0}(\mathbf{x})} |u|.$$

But we know that $\max_{\partial B_{mr_0}(\mathbf{x})} |u| \leq r_0/|\mathbf{x}|$ and $r_0 \geq |\mathbf{x}|/(m+1)$ so that we get

$$\left|u_{x_j}(\mathbf{x})\right| \leq \frac{m+1}{(m-1)} \frac{3\,r_0}{|\mathbf{x}|^2} \leq \frac{6\,r_0}{|\mathbf{x}|^2}$$

since $(m+1) < 2(m-1)$.

Similarly we can prove

$$\left|u_{x_j x_k}(\mathbf{x})\right| \leq \frac{c\,r_0}{|\mathbf{x}|^3}$$

$\square$

The estimates (3.83) assure the validity of the Green identity

$$\int_{\Omega_e} \nabla u \cdot \nabla v \, d\mathbf{x} = \int_{\partial \Omega_e} v \partial_{\boldsymbol{\nu}} u \, d\sigma \tag{3.85}$$

for any pair $u, v \in C^2(\Omega_e) \cap C^1(\overline{\Omega}_e)$, harmonic in Ω_e and vanishing at infinity. To see this, apply the identity (3.85) in the bounded domain $\Omega_{e,r} = \Omega_e \cap \{|\mathbf{x}| < r\}$. Then let $r \to \infty$ to get (3.85) in Ω_e.

In turn, via the Green identity (3.85), we can prove an appropriate version of Theorem 3.15 for the exterior problem

$$\begin{cases} \Delta u = f & \text{in } \Omega_e \\ \partial_{\boldsymbol{\nu}} u + ku = g & \text{on } \partial \Omega_e, \ (k \geq 0) \\ u \to u_\infty & \text{as } |\mathbf{x}| \to \infty. \end{cases} \tag{3.86}$$

Observe that $k \neq 0$ corresponds to the Robin problem while $k = 0$ corresponds to the Neumann problem.

Theorem 3.16. *(Exterior Neumann and Robin problems). Let $\Omega_e \subset \mathbb{R}^3$ be an exterior domain. Then there exists at most one solution $u \in C^2(\Omega_e) \cap C^1(\overline{\Omega}_e)$ of problem (3.86).*

Proof. Suppose u, v are solutions of (3.86) and let $w = u - v$. Then w is harmonic Ω^e, $\partial_{\boldsymbol{\nu}} w + kw = 0$ on $\partial \Omega_e$ and $w \to 0$ as $|\mathbf{x}| \to \infty$.

Apply the identity (3.85) with $u = v = w$. Since $\partial_{\boldsymbol{\nu}} w = -kw$ on $\partial \Omega_e$ we have:

$$\int_{\Omega_e} |\nabla w|^2 \, d\mathbf{x} = \int_{\partial \Omega_e} w \partial_{\boldsymbol{\nu}} w \, d\sigma = - \int_{\partial \Omega_e} kw^2 d\sigma \leq 0.$$

Thus $\nabla w = 0$ and w is constant, because Ω_e is connected. But w vanishes at infinity so that $w = 0$. $\square$

3.7 Surface Potentials

In this section we go back to examine the meaning and the main properties of the surface potentials appearing in the identity (3.56). A remarkable consequence is the possibility to convert a boundary value problem into a **boundary integral equation**. This kind of formulation can be obtained for more general operators and more general problems as soon as a fundamental solution is known. Thus it constitutes a flexible method with important implications. In particular, it constitutes the theoretical basis for the so called *boundary element method*, which may offer several advantages from the point of view of the computational cost in numerical approximations, due to a dimension reduction. Here we present the integral formulations of the main boundary value problems and state some basic results. The reader can find complete proofs and the integral formulation of more general or different problems in the literature at the end of the book (e.g. *Dautrait-Lions, vol 3, 1985*).

3.7.1 The double and single layer potentials

The last integral in (3.56) is of the form

$$\mathcal{D}\left(\mathbf{x};\mu\right) = \int_{\partial\Omega} \mu\left(\boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma \tag{3.87}$$

and it is called the **double layer potential of** μ. In three dimensions it represents the electrostatic potential generated by a *dipole* distribution[23] of moment μ on $\partial\Omega$.

To get a clue of the main features of $\mathcal{D}\left(\mathbf{x};\mu\right)$, it is useful to look at the particular case $\mu\left(\boldsymbol{\sigma}\right) \equiv 1$, that is

$$\mathcal{D}\left(\mathbf{x};1\right) = \int_{\partial\Omega} \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma. \tag{3.88}$$

Inserting $u \equiv 1$ into (3.56) we get

$$\mathcal{D}\left(\mathbf{x};1\right) = -1 \qquad \text{for every } \mathbf{x} \in \Omega. \tag{3.89}$$

On the other hand, if $\mathbf{x} \in \mathbb{R}^n \backslash \overline{\Omega}$ is fixed, $\Phi\left(\mathbf{x} - \cdot\right)$ is harmonic in Ω and can be inserted into (3.57) with $u \equiv 1$; the result is

$$\mathcal{D}\left(\mathbf{x};1\right) = 0 \qquad \text{for every } \mathbf{x} \in \mathbb{R}^n \backslash \overline{\Omega}. \tag{3.90}$$

What happens for $\mathbf{x} \in \partial\Omega$? First of all we have to check that $\mathcal{D}\left(\mathbf{x};1\right)$ is well defined (i.e. finite) on $\partial\Omega$. Indeed the singularity of $\partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right)$ becomes critical when $\mathbf{x} \in \partial\Omega$ since as $\boldsymbol{\sigma} \to \mathbf{x}$ the order of infinity equals the topological dimension of $\partial\Omega$. For instance, in the two dimensional case we have

$$\mathcal{D}\left(\mathbf{x};1\right) = -\frac{1}{2\pi} \int_{\partial\Omega} \partial_{\boldsymbol{\nu}_{\sigma}} \log|\mathbf{x} - \boldsymbol{\sigma}| \, d\sigma = -\frac{1}{2\pi} \int_{\partial\Omega} \frac{(\mathbf{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{\nu}_{\sigma}}{|\mathbf{x} - \boldsymbol{\sigma}|^2} d\sigma.$$

The order of infinity of the integrand is one and the boundary $\partial\Omega$ is a curve, a one dimensional object. In the three dimensional case we have

$$\mathcal{D}\left(\mathbf{x};1\right) = \frac{1}{4\pi} \int_{\partial\Omega} \frac{\partial}{\partial \boldsymbol{\nu}_{\sigma}} \frac{1}{|\mathbf{x} - \boldsymbol{\sigma}|} d\sigma = \frac{1}{4\pi} \int_{\partial\Omega} \frac{(\mathbf{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{\nu}_{\sigma}}{|\mathbf{x} - \boldsymbol{\sigma}|^3} d\sigma.$$

---

[23] For every $\boldsymbol{\sigma} \in \partial\Omega$, let $-q\left(\boldsymbol{\sigma}\right)$, $q\left(\boldsymbol{\sigma}\right)$ two charges placed at the points $\boldsymbol{\sigma}$, $\boldsymbol{\sigma}+h\boldsymbol{\nu}_{\sigma}$, respectively. If $h > 0$ is very small, the pair of charges constitutes a *dipole of axis* $\boldsymbol{\nu}_{\sigma}$. The induced potential at $\mathbf{x}$ is given by

$$u\left(\mathbf{x}, \boldsymbol{\sigma}\right) = q\left(\boldsymbol{\sigma}\right) \left[\Phi\left(\mathbf{x} - \left(\boldsymbol{\sigma}+h\boldsymbol{\nu}\right)\right) - \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right)\right] = q\left(\boldsymbol{\sigma}\right) h \left[\frac{\Phi\left(\mathbf{x} - \left(\boldsymbol{\sigma}+h\boldsymbol{\nu}\right)\right) - \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right)}{h}\right].$$

Since h is very small, setting $q\left(\boldsymbol{\sigma}\right) h = \mu\left(\boldsymbol{\sigma}\right)$, we can write, at first order of approximation,

$$u_h\left(\mathbf{x}\right) \simeq \mu\left(\boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right).$$

Integrating on $\partial\Omega$ we obtain $\mathcal{D}\left(\mathbf{x};\mu\right)$.

The order of infinity of the integrand is two and $\partial\Omega$ is a bidimensional surface.

However, if we assume that Ω is **a C^2 domain**, then it can be proved that $\mathcal{D}(\mathbf{x};1)$ is well defined on $\partial\Omega$.

To compute the value of $\mathcal{D}(\mathbf{x};1)$ on $\partial\Omega$ we first observe that formulas (3.89) and (3.90) follow immediately from the geometric interpretation of the integrand in $\mathcal{D}(\mathbf{x};1)$. Precisely, set $r_{\mathbf{x}\boldsymbol{\sigma}} = |\mathbf{x} - \boldsymbol{\sigma}|$ and, in dimension two, consider the quantity

$$d\sigma^* = -\frac{(\mathbf{x} - \boldsymbol{\sigma})\cdot\boldsymbol{\nu}_{\boldsymbol{\sigma}}}{r_{\mathbf{x}\boldsymbol{\sigma}}^2}\, d\sigma = \frac{(\boldsymbol{\sigma} - \mathbf{x})\cdot\boldsymbol{\nu}_{\boldsymbol{\sigma}}}{r_{\mathbf{x}\boldsymbol{\sigma}}^2}\, d\sigma.$$

We have (see Fig. 3.7),

$$\frac{(\boldsymbol{\sigma} - \mathbf{x})}{r_{\mathbf{x}\boldsymbol{\sigma}}}\cdot\boldsymbol{\nu}_{\boldsymbol{\sigma}} = \cos\varphi$$

and therefore

$$d\sigma' = \frac{(\boldsymbol{\sigma} - \mathbf{x})\cdot\boldsymbol{\nu}_{\boldsymbol{\sigma}}}{r_{\mathbf{x}\boldsymbol{\sigma}}}\, d\sigma = \cos\varphi\, d\sigma$$

is the projection of the length element $d\sigma$ on the circle $\partial B_{r_{\mathbf{x}\boldsymbol{\sigma}}}(\mathbf{x})$, up to an error of lower order. Then

$$d\sigma^* = \frac{d\sigma'}{r_{\mathbf{x}\boldsymbol{\sigma}}}$$

is the projection of $d\sigma$ on $\partial B_1(\mathbf{x})$.

Integrating on $\partial\Omega$, the contributions to $d\sigma^*$ sum up to 2π if $\mathbf{x}\in\Omega$ (case a) of figure 3.8) and to 0 if $\mathbf{x}\in\mathbb{R}^n\backslash\overline{\Omega}$, due to the sign compensations induced by the orientation of $\boldsymbol{\nu}_{\boldsymbol{\sigma}}$ (case c) of figure 3.8). Thus

$$\int_{\partial\Omega} d\sigma^* = \begin{cases} 2\pi \text{ if } \mathbf{x}\in\Omega \\ 0 \quad \text{ if } \mathbf{x}\in\mathbb{R}^2\backslash\overline{\Omega} \end{cases}$$
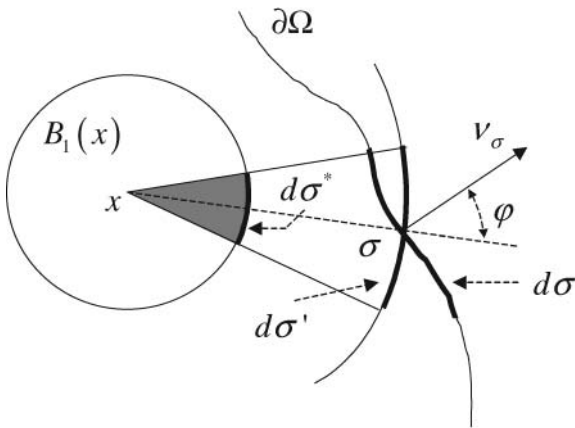
Fig. 3.7. Geometrical interpretation of the integrand in $\mathcal{D}(\mathbf{x},1)$ $(n = 2)$

which are equivalent to (3.89) and (3.90), since

$$\mathcal{D}\left(\mathbf{x};1\right) = -\frac{1}{2\pi}\int_{\partial\Omega} d\sigma^*.$$

The case b) in figure 3.8 corresponds to $\mathbf{x} \in \partial\Omega$. It should be now intuitively clear that the point $\mathbf{x}$ " sees" a total angle of only π radiants and $\mathcal{D}\left(\mathbf{x};1\right) = -1/2$.

The same kind of considerations hold in dimension three; this time, the quantity (see Fig. 3.9)

$$d\sigma^* = -\frac{(\mathbf{x} - \boldsymbol{\sigma}) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}}{r_{\mathbf{x}\boldsymbol{\sigma}}^3} d\sigma = \frac{(\boldsymbol{\sigma} - \mathbf{x}) \cdot \boldsymbol{\nu}_{\boldsymbol{\sigma}}}{r_{\mathbf{x}\boldsymbol{\sigma}}^3} d\sigma$$

is the projection on $\partial B_1\left(\mathbf{x}\right)$ (*solid angle*) of the surface element $d\sigma$. Integrating over $\partial\Omega$, the contributions to $d\sigma^*$ sum up to 4π if $\mathbf{x} \in \Omega$ and to 0 if $\mathbf{x} \in \mathbb{R}^3\backslash\overline{\Omega}$.

If $\mathbf{x} \in \partial\Omega$, it " sees" a total solid angle of measure 2π. Since

$$\mathcal{D}\left(\mathbf{x};1\right) = -\frac{1}{4\pi}\int_{\partial\Omega} d\sigma^*,$$

we find again the values $-1, 0, -1/2$ in the three cases, respectively.

a) 2π *b)* π *c)* 0
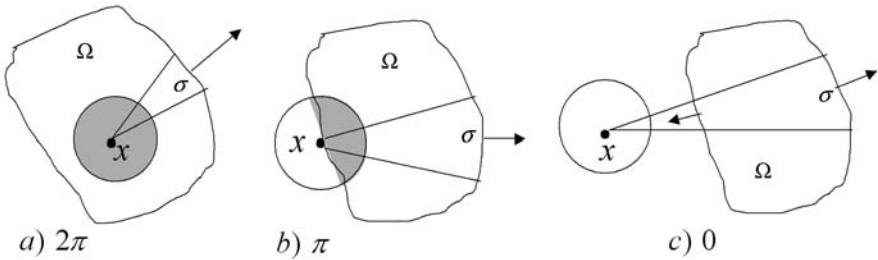
Fig. 3.8. Values of $\int_{\partial\Omega} d\sigma^*$ for $n = 2$

We gather the above results in the following Lemma.

Lemma 3.2 (Gauss). *Let* $\Omega \subset \mathbb{R}^n$ *be a bounded,* C^2*-domain. Then*

$$\mathcal{D}\left(\mathbf{x};1\right) = \int_{\partial\Omega} \partial_{\boldsymbol{\nu}_{\boldsymbol{\sigma}}}\Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma = \begin{cases} -1 & \mathbf{x} \in \Omega \\ -\dfrac{1}{2} & \mathbf{x} \in \partial\Omega \\ 0 & \mathbf{x} \in \mathbb{R}^n\backslash\overline{\Omega}. \end{cases}$$

Thus, when $\mu \equiv 1$, the double layer potential is constant outside $\partial\Omega$ and has *a jump discontinuity* across $\partial\Omega$. Observe that if $\mathbf{x} \in \partial\Omega$,

$$\lim_{\mathbf{z}\to\mathbf{x},\ \mathbf{z}\in\mathbb{R}^n\backslash\bar{\Omega}} \mathcal{D}\left(\mathbf{z};1\right) = \mathcal{D}\left(\mathbf{x};1\right) + \frac{1}{2}$$
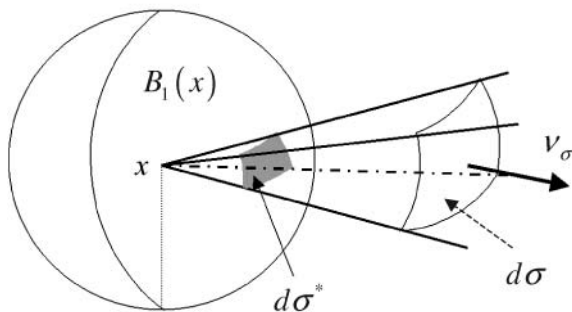
Fig. 3.9. The solid angle $d\sigma^*$, projected from $d\sigma$

and

$$\lim_{\mathbf{z}\to\mathbf{x},\ \mathbf{z}\in\Omega} \mathcal{D}\left(\mathbf{z};1\right) = \mathcal{D}\left(\mathbf{x};1\right) - \frac{1}{2}.$$

These formulas are the key for understanding the general properties of $\mathcal{D}\left(\mathbf{x};\mu\right)$, that we state in the following theorem.

Theorem 3.17. *Let $\Omega \subset \mathbb{R}^n$ be a bounded, C^2 domain and μ a continuous function on $\partial\Omega$. Then, $\mathcal{D}\left(\mathbf{x};\mu\right)$ is harmonic in $\mathbb{R}^n\backslash\partial\Omega$ and the following jump relations hold for every $\mathbf{x} \in \partial\Omega$:*

$$\lim_{\mathbf{z}\to\mathbf{x},\ \mathbf{z}\in\mathbb{R}^n\backslash\overline{\Omega}} \mathcal{D}\left(\mathbf{z};\mu\right) = \mathcal{D}\left(\mathbf{x};\mu\right) + \frac{1}{2}\mu\left(\mathbf{x}\right) \tag{3.91}$$

and

$$\lim_{\mathbf{z}\to\mathbf{x},\ \mathbf{z}\in\Omega} \mathcal{D}\left(\mathbf{z};\mu\right) = \mathcal{D}\left(\mathbf{x};\mu\right) - \frac{1}{2}\mu\left(\mathbf{x}\right). \tag{3.92}$$

Proof (Sketch). If $\mathbf{x} \notin \partial\Omega$ there is no problem in differentiating under the integral sign and, for $\boldsymbol{\sigma}$ fixed on $\partial\Omega$, the function

$$\partial_{\boldsymbol{\nu}_\sigma}\Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) = \nabla_{\boldsymbol{\sigma}}\Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right)\cdot\boldsymbol{\nu}_\sigma$$

is harmonic. Thus $\mathcal{D}\left(\mathbf{x};\mu\right)$ is harmonic in $\mathbb{R}^n\backslash\partial\Omega$.

Consider (3.91). This is not an elementary formula and we cannot take the limit under the integral sign, once more because of the critical singularity of $\partial_{\boldsymbol{\nu}_\sigma}\Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right)$ when $\mathbf{x} \in \partial\Omega$.

Let $\mathbf{z} \in\mathbb{R}^n\backslash\overline{\Omega}$. From Gauss Lemma 3.2

$$\mu\left(\mathbf{x}\right)\int_{\partial\Omega} \partial_{\boldsymbol{\nu}_\sigma}\Phi\left(\mathbf{z} - \boldsymbol{\sigma}\right) d\sigma = 0$$

and we can write

$$\mathcal{D}\left(\mathbf{z};\mu\right) = \int_{\partial\Omega} \partial_{\boldsymbol{\nu}_\sigma}\Phi\left(\mathbf{z} - \boldsymbol{\sigma}\right)\left[\mu\left(\boldsymbol{\sigma}\right) - \mu\left(\mathbf{x}\right)\right] d\sigma. \tag{3.93}$$

Now, when $\boldsymbol{\sigma}$ is near $\mathbf{x}$, the smoothness of $\partial\Omega$ and the continuity of μ mitigate the singularity of $\partial_{\nu_\sigma}\Phi(\mathbf{x}-\boldsymbol{\sigma})$ and allow to take the limit under the integral sign. Thus

$$\lim_{\mathbf{z}\to\mathbf{x}}\int_{\partial\Omega}\partial_{\nu_\sigma}\Phi(\mathbf{z}-\boldsymbol{\sigma})\left[\mu(\boldsymbol{\sigma})-\mu(\mathbf{x})\right]d\sigma=\int_{\partial\Omega}\partial_{\nu_\sigma}\Phi(\mathbf{x}-\boldsymbol{\sigma})\left[\mu(\boldsymbol{\sigma})-\mu(\mathbf{x})\right]d\sigma.$$

Exploiting once more Gauss lemma, we have

$$=\int_{\partial\Omega}\partial_{\boldsymbol{\nu}_\sigma}\Phi(\mathbf{x}-\boldsymbol{\sigma})\,\mu(\boldsymbol{\sigma})\,d\sigma-\mu(\mathbf{x})\int_{\partial\Omega}\partial_{\boldsymbol{\nu}_\sigma}\Phi(\mathbf{x}-\boldsymbol{\sigma})\,d\sigma$$

$$=\mathcal{D}(\mathbf{x};\mu)+\frac{1}{2}\mu(\mathbf{x})\,.$$

The proof of (3.92) is similar. $\square$

The second integral in (3.56) is of the form

$$\mathcal{S}(\mathbf{x},\psi)=\int_{\partial\Omega}\Phi(\mathbf{x}-\boldsymbol{\sigma})\,\psi(\boldsymbol{\sigma})\,d\sigma$$

and it is called the **single layer potential of** ψ.

In three dimensions it represents the electrostatic potential generated by a charge distribution of density ψ on $\partial\Omega$. If Ω is a C^2-domain and ψ is continuous on $\partial\Omega$, then $\mathcal{S}$ is *continuous across* $\partial\Omega$ and

$$\Delta\mathcal{S}=0\qquad\text{in }\mathbb{R}^n\backslash\partial\Omega,$$

because there is no problem in differentiating under the integral sign.

Since the flux of an electrostatic potential undergoes a jump discontinuity across a charged surface, we expect a jump discontinuity of the normal derivative of $\mathcal{S}$ across $\partial\Omega$. Precisely

Theorem 3.18. *Let* $\Omega\subset\mathbb{R}^n$ *be a bounded,* C^2*-domain and* ψ *a continuous function on* $\partial\Omega$. *Then,* $\mathcal{S}(\mathbf{x};\psi)$ *is harmonic in* $\mathbb{R}^n\backslash\partial\Omega$, *continuous across* $\partial\Omega$ *and the following jump relations hold for every* $\mathbf{x}\in\partial\Omega$:

$$\lim_{\mathbf{z}\to\mathbf{x},\,\mathbf{z}\in\mathbb{R}^n\backslash\overline{\Omega}}\partial_{\boldsymbol{\nu}_\mathbf{x}}\mathcal{S}(\mathbf{z};\psi)=\int_{\partial\Omega}\partial_{\boldsymbol{\nu}_\mathbf{x}}\Phi(\mathbf{x}-\boldsymbol{\sigma})\,\psi(\boldsymbol{\sigma})\,d\sigma-\frac{1}{2}\psi(\mathbf{x})\qquad(3.94)$$

and

$$\lim_{\mathbf{z}\to\mathbf{x},\,\mathbf{z}\in\Omega}\partial_{\boldsymbol{\nu}_\mathbf{x}}\mathcal{S}(\mathbf{z};\psi)=\int_{\partial\Omega}\partial_{\boldsymbol{\nu}_\mathbf{x}}\Phi(\mathbf{x}-\boldsymbol{\sigma})\,\psi(\boldsymbol{\sigma})\,d\sigma+\frac{1}{2}\psi(\mathbf{x})\,.\qquad(3.95)$$

3.7.2 The integral equations of potential theory

By means of the jump relations (3.92)-(3.95) of the double and single layer potentials we can reduce the main boundary value problems of potential theory into integral equations of a special form. Let $\Omega\subset\mathbb{R}^n$ be a smooth domain and $g\in C(\partial\Omega)$. We first show the reduction procedure for the *interior Dirichlet problem*

$$\begin{cases}\Delta u=0 & \text{in }\Omega\\ u=g & \text{on }\partial\Omega.\end{cases}\qquad(3.96)$$

The starting point is once more the identity (3.56), which gives for the solution u of (3.96) the representation

$$u\left(\mathbf{x}\right) = \int_{\partial\Omega} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} u\left(\boldsymbol{\sigma}\right) d\sigma - \int_{\partial\Omega} g\left(\boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma.$$

In subsection 3.5.3 we used the Green function to get rid of the single layer potential containing the unknown $\partial_{\boldsymbol{\nu}_{\sigma}} u$. Here we adopt a different strategy: we forget the single layer potential and try to represent u in the form of a double layer potential, by choosing an appropriate density. In other words, *we seek a continuous function μ on $\partial\Omega$, such that the solution u of* (3.96) *is given by*

$$u\left(\mathbf{x}\right) = \int_{\partial\Omega} \mu\left(\boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma = \mathcal{D}\left(\mathbf{x};\mu\right). \tag{3.97}$$

The function u given by (3.97) is harmonic in Ω so that we have only to check the boundary condition

$$\lim_{\mathbf{x}\to\mathbf{z}\in\partial\Omega} u\left(\mathbf{x}\right) = g\left(\mathbf{z}\right).$$

Letting $\mathbf{x}\to\mathbf{z}\in\partial\Omega$ and taking into account the jump relation (3.91), we obtain for μ the integral equation

$$\int_{\partial\Omega} \mu\left(\boldsymbol{\sigma}\right) \partial_{\boldsymbol{\nu}_{\sigma}} \Phi\left(\mathbf{z} - \boldsymbol{\sigma}\right) d\sigma - \frac{1}{2}\mu\left(\mathbf{z}\right) = g\left(\mathbf{z}\right) \qquad \mathbf{z}\in\partial\Omega. \tag{3.98}$$

If $\mu \in C\left(\partial\Omega\right)$ solves (3.98), then (3.97) is a solution of (3.96) in $C^2\left(\Omega\right)\cap C\left(\overline{\Omega}\right)$. The following theorem holds.

Theorem 3.19. *Let $\Omega \subset \mathbb{R}^n$ be a bounded, C^2 domain and g a continuous function on $\partial\Omega$. Then, the integral equation (3.98) has a unique solution $\mu \in C\left(\partial\Omega\right)$ and the solution $u \in C^2\left(\Omega\right)\cap C\left(\overline{\Omega}\right)$ of the Dirichlet problem (3.96) can be represented as the double layer potential of μ.*

We consider now the *interior Neumann problem*

$$\begin{cases} \Delta u = 0 & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u = g & \text{on } \partial\Omega \end{cases} \tag{3.99}$$

where $g \in C(\partial B_R)$ satisfies the solvability condition

$$\int_{\partial B_R} g\, d\sigma = 0. \tag{3.100}$$

This time *we seek a continuous function ψ on $\partial\Omega$, such that the solution u of* (3.99) *is given in the form*

$$u\left(\mathbf{x}\right) = \int_{\partial\Omega} \psi\left(\boldsymbol{\sigma}\right) \Phi\left(\mathbf{x} - \boldsymbol{\sigma}\right) d\sigma = \mathcal{S}\left(\mathbf{x},\psi\right). \tag{3.101}$$

The function u given by (3.101) is harmonic in Ω, so that we have only to check the boundary condition

$$\lim_{\substack{\mathbf{x} \to \mathbf{z} \in \partial\Omega}} \partial_{\boldsymbol{\nu}_\mathbf{z}} u(\mathbf{x}) = g(\mathbf{z}).$$

Letting $\mathbf{x} \to \mathbf{z} \in \partial\Omega$ and taking into account the jump relation (3.95), we obtain for ψ the integral equation

$$\int_{\partial\Omega} \psi(\boldsymbol{\sigma}) \partial_{\boldsymbol{\nu}_\mathbf{z}} \Phi(\mathbf{z} - \boldsymbol{\sigma}) \, d\sigma + \frac{1}{2} \psi(\mathbf{z}) = g(\mathbf{z}) \qquad\qquad \mathbf{z} \in \partial\Omega. \qquad\quad (3.102)$$

If $\psi \in C(\partial\Omega)$ solves (3.102), then (3.101) is a solution of (3.99) in $C^2(\Omega) \cap C^1(\overline{\Omega})$.

It turns out that the general solution of (3.102) has the form

$$\psi = \overline{\psi} + C_0 \psi_0 \qquad C_0 \in \mathbb{R},$$

where $\overline{\psi}$ is a particular solution of (3.102) and ψ_0 is a solution of the homogeneous equation

$$\int_{\partial\Omega} \psi_0(\boldsymbol{\sigma}) \partial_{\boldsymbol{\nu}_\mathbf{z}} \Phi(\mathbf{z} - \boldsymbol{\sigma}) \, d\sigma + \frac{1}{2} \psi_0(\mathbf{z}) = 0 \qquad\qquad \mathbf{z} \in \partial\Omega. \qquad\quad (3.103)$$

As expected, we have infinitely many solutions to the Neumann problem. Observe that

$$\mathcal{S}(\mathbf{x}, \psi_0) = \int_{\partial\Omega} \psi_0(\boldsymbol{\sigma}) \Phi(\mathbf{x} - \boldsymbol{\sigma}) \, d\sigma$$

is harmonic in Ω with vanishing normal derivative on $\partial\Omega$, because of (3.95) and (3.103). Consequently, $\mathcal{S}(\mathbf{x}, \psi_0)$ is constant and the following theorem holds.

Theorem 3.20. *Let $\Omega \subset \mathbb{R}^n$ be a bounded, C^2-domain and g a continuous function on $\partial\Omega$ satisfying (3.100). Then, the Neumann problem (3.99) has infinitely many solutions $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ of the form*

$$u(\mathbf{x}) = \mathcal{S}(\mathbf{x}, \overline{\psi}) + C,$$

where $\overline{\psi}$ is a particular solution of (3.102) and C is an arbitrary constant.

Another advantage of the method is that, in principle, exterior problems can be treated as the interior problems, with the same level of difficulty. It is enough to use the exterior jump conditions (3.91), (3.94) and proceed in the same way (see Problem 3.16).

As an example of an elementary application of the method we solve the interior Neumann problem for the circle.

• *The Neumann problem for the circle.* Let $B_R = B_R(\mathbf{0}) \subset \mathbb{R}^2$ and consider the Neumann problem

$$\begin{cases} \Delta u = 0 & \text{in } B_R \\ \partial_\nu u = g & \text{on } \partial B_R \end{cases}$$

where $g \in C(\partial B_R)$ satisfies the solvability condition (3.100). We know that u is unique up to an additive constant. We want to express the solution as a single layer potential:

$$u\left(\mathbf{x}\right) = -\frac{1}{2\pi} \int_{\partial B_R} \psi\left(\boldsymbol{\sigma}\right) \log\left|\mathbf{x} - \boldsymbol{\sigma}\right| d\sigma. \tag{3.104}$$

The Neumann condition $\partial_\nu u = g$ on ∂B_R translates into the following integral equation for the density ψ:

$$-\frac{1}{2\pi} \int_{\partial B_R} \frac{\left(\mathbf{z} - \boldsymbol{\sigma}\right) \cdot \boldsymbol{\nu}_z}{\left|\mathbf{z} - \boldsymbol{\sigma}\right|^2} \psi\left(\boldsymbol{\sigma}\right) d\sigma + \frac{1}{2}\psi\left(\mathbf{z}\right) = g\left(\mathbf{z}\right) \qquad \left(\mathbf{z} \in \partial B_R\right). \tag{3.105}$$

On ∂B_R we have

$$\boldsymbol{\nu}_z = \mathbf{z}/R \quad \text{and} \quad \left(\mathbf{z} - \boldsymbol{\sigma}\right) \cdot \mathbf{z} = R^2 - \mathbf{z} \cdot \boldsymbol{\sigma}$$

and

$$\left|\mathbf{z} - \boldsymbol{\sigma}\right|^2 = 2\left(R^2 - \mathbf{z} \cdot \boldsymbol{\sigma}\right)$$

so that (3.105) becomes

$$-\frac{1}{4\pi R} \int_{\partial B_R} \psi\left(\boldsymbol{\sigma}\right) d\sigma + \frac{1}{2}\psi\left(\mathbf{z}\right) = g\left(\mathbf{z}\right) \qquad \left(\mathbf{z} \in \partial B_R\right). \tag{3.106}$$

The solutions of the homogeneous equation $(g = 0)$ are constant functions $\psi_0\left(x\right) = C$ (why?). A particular solution $\bar{\psi}$ with

$$\int_{\partial B_R} \bar{\psi}\left(\boldsymbol{\sigma}\right) d\sigma = 0$$

is given by

$$\bar{\psi}\left(\mathbf{z}\right) = 2g\left(\mathbf{z}\right).$$

Thus, the general solution of (3.106) is

$$\psi\left(\mathbf{z}\right) = 2g\left(\mathbf{z}\right) + C \qquad C \in \mathbb{R}$$

and up to an additive constant, the solution of the Neumann problem is given by

$$u\left(\mathbf{x}\right) = -\frac{1}{\pi} \int_{\partial B_R} g\left(\boldsymbol{\sigma}\right) \log\left|\mathbf{x} - \boldsymbol{\sigma}\right| d\sigma.$$

Remark 3.8. The integral equations (3.98) and (4.114) are of the form

$$\int_{\partial \Omega} K\left(\mathbf{z}, \boldsymbol{\sigma}\right) \rho(\boldsymbol{\sigma}) d\sigma \pm \frac{1}{2}\rho\left(\mathbf{z}\right) = g\left(\mathbf{z}\right) \tag{3.107}$$

and are called *Fredholm integral equations of the first kind*. Their solution is based on the following so called **Fredholm alternative**: *either equation (3.107) has exactly one solution for every* $g \in C(\partial\Omega)$, *or the homogeneous equation*

$$\int_{\partial\Omega} K(\mathbf{z}, \boldsymbol{\sigma})\,\phi(\boldsymbol{\sigma})d\sigma \pm \frac{1}{2}\phi(\mathbf{z}) = 0$$

has a finite number $\phi_1, ..., \phi_N$ *of non trivial, linearly independent solutions.*

In this last case equation (3.107) is not always solvable and we have:

(a) the **adjoint** homogeneous equation

$$\int_{\partial\Omega} K(\boldsymbol{\sigma}, \mathbf{z})\,\phi^*(\boldsymbol{\sigma})d\sigma \pm \frac{1}{2}\phi^*(\mathbf{z}) = 0$$

has N non trivial linearly independent solutions $\phi_1^*, ..., \phi_N^*$;

(b) equation (3.107) has a solution if and only if g satisfies the following N compatibility conditions:

$$\int_{\partial\Omega} \phi_j^*(\boldsymbol{\sigma})g(\boldsymbol{\sigma})\,d\sigma = 0, \quad j = 1, ..., N \tag{3.108}$$

(c) if g satisfies (3.108), the general solution of (3.107) is given by

$$\rho = \overline{\rho} + C_1\phi_1 + ...C_N\phi_N$$

where $\overline{\rho}$ is a particular solution of equation (3.107) and $C_1, ..., C_N$ are arbitrary real constants.

The analogy with the solution of a system of linear algebraic equations should be evident. We will come back to Fredholm's alternative in Chapter 6.

Problems

3.1. Show that if u is harmonic in a domain Ω, also the derivatives of u of any order are harmonic in Ω.

3.2. We say that a function $u \in C^2(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ is *subharmonic* (resp. *superharmonic*) in Ω if $\Delta u \geq 0$ ($\Delta u \leq 0$) in Ω. Show that:

a) If u is subharmonic, then, for every $B_R(\mathbf{x}) \subset\subset \Omega$,

$$u(\mathbf{x}) \leq \frac{n}{\omega_n R^n} \int_{B_R(\mathbf{x})} u(\mathbf{y})\,d\mathbf{y}$$

and

$$u(\mathbf{x}) \leq \frac{1}{\omega_n R^{n-1}} \int_{\partial B_R(\mathbf{x})} u(\mathbf{y})\,d\mathbf{y}.$$

If u is superharmonic, the reverse inequalities hold.

b) If $u \in C\left(\overline{\Omega}\right)$ is subharmonic, (superharmonic), the maximum (minimum) of u is attained on $\partial\Omega$.

c) If u is harmonic in Ω then u^2 is subharmonic.

d) Let u be subharmonic in Ω and $F : \mathbb{R} \to \mathbb{R}$, smooth. Under which conditions on F is $F \circ u$ subharmonic?

3.3. Let $\Omega \subset \mathbb{R}^2$ be a bounded domain and $v \in C^2\left(\Omega\right) \cap C^1\left(\overline{\Omega}\right)$ be a solution of (*torsion problem*)

$$\begin{cases} v_{xx} + v_{yy} = -2 & \text{in } \Omega \\ v = 0 & \text{on } \partial\Omega. \end{cases}$$

Show that $u = |\nabla v|^2$ attains its maximum on $\partial\Omega$.

3.4. Let B_R be the unit circle centered at $(0,0)$. Use the method of separation of variables to solve the problem

$$\begin{cases} \Delta u = f & \text{in } B_R \\ u = 1 & \text{on } \partial B_R. \end{cases}$$

Find an explicit formula when $f(x, y) = y$.

[*Hint*: Use polar coordinates; expand $f = f(r, \cdot)$ in sine Fourier series in $[0, 2\pi]$ and derive a set of ordinary differential equations for the Fourier coefficients of $u(r, \cdot)$].

3.5. Let $B_{1,2} = \{(r, \theta) \in \mathbb{R}^2; 1 < r < 2\}$. Examine the solvability of the Neumann problem

$$\begin{cases} \Delta u = -1 & \text{in } B_{1,2} \\ u_\nu = \cos\theta & \text{on } r = 1 \qquad (\lambda \in \mathbb{R}) \\ u_\nu = \lambda(\cos\theta)^2 & \text{on } r = 2 \end{cases}$$

and write an explicit formula for the solution, when it exists.

3.6 (*Schwarz's reflection principle*). Let

$$B_1^+ = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, y > 0\right\}$$

and $u \in C^2\left(B_1^+\right) \cap C(\overline{B_1^+})$, harmonic B_1^+, $u(x, 0) = 0$. Show that the function

$$U(x, y) = \begin{cases} u(x, y) & y \geq 0 \\ -u(x, -y) & y < 0 \end{cases}$$

obtained from u by odd reflection with respect to y, is harmonic in all B_1.

[*Hint*: Let v be the solution of $\Delta v = 0$ in B_1, $v = U$ on ∂B_1. Define

$$w(x, y) = v(x, y) + v(x, -y)$$

and show that $w \equiv 0$ ...].
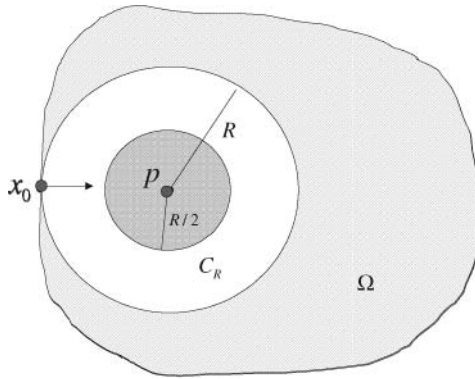
Fig. 3.10. Interior circle condition at $\mathbf{x}_0$

3.7. State and prove the *Schwarz reflection principle* in dimension three.

3.8. Let u be harmonic in $\mathbb{R}^3$ such that

$$\int_{\mathbb{R}^3} |u(\mathbf{x})|^2 \, d\mathbf{x} < \infty.$$

Show that $u \equiv 0$.

[*Hint.* Write the mean formula in a ball $B_R(\mathbf{0})$ for u. Use the Schwarz inequality and let $R \to +\infty$].

3.9. Let u be harmonic in $\mathbb{R}^n$ and $\mathbf{M}$ an orthogonal matrix of order n. Using the mean value property, show that $v(\mathbf{x}) = u(\mathbf{Mx})$ is harmonic in $\mathbb{R}^n$.

3.10. (*Hopf's maximum principle*). Let Ω be a domain in $\mathbb{R}^2$ and $u \in C^1(\overline{\Omega})$, harmonic and positive in Ω. Assume that $u(\mathbf{x}_0) = 0$ at a point $\mathbf{x}_0 \in \partial\Omega$ and that at $\mathbf{x}_0$ the following *interior circle condition* holds (see Fig. 3.10): *there exists a circle $C_R(\mathbf{p}) \subset \Omega$ such that*

$$C_R(\mathbf{p}) \cap \partial\Omega = \{\mathbf{x}_0\}.$$

(a) Show that the exterior normal derivative of u at $\mathbf{x}_0$ is (strictly) negative: $u_{\boldsymbol{\nu}}(\mathbf{x}_0) < 0$.

(b) Generalize to any number of dimensions.

[*Hint.* (a) Use the maximum principle to compare u with the function

$$w(\mathbf{x}) = \frac{\ln|R| - \ln|\mathbf{x} - \mathbf{p}|}{\ln R - \ln(R/2)} \min_{\partial C_{R/2}(\mathbf{p})} u$$

in the ring

$$A = C_R(\mathbf{p}) \setminus C_{R/2}(\mathbf{p}).$$

Then, compare the normal derivatives at $\mathbf{x}_0$].

3.11. Let $f \in C^2 \left(\mathbb{R}^2\right)$ with compact support K and

$$u \left(\mathbf{x}\right) = -\frac{1}{2\pi} \int_{\mathbb{R}^2} \log |\mathbf{x} - \mathbf{y}| \ f \left(\mathbf{y}\right) d\mathbf{y}.$$

Show that

$$u \left(\mathbf{x}\right) = -\frac{M}{2\pi} \log |\mathbf{x}| + O(|\mathbf{x}|^{-1}), \qquad \text{as } |\mathbf{x}| \to +\infty$$

where

$$M = \int_{\mathbb{R}^2} f \left(\mathbf{y}\right) d\mathbf{y}.$$

[*Hint:* Write

$$\log |\mathbf{x} - \mathbf{y}| = \log \left(|\mathbf{x} - \mathbf{y}| / |\mathbf{x}|\right) + \log |\mathbf{x}|$$

and show that, if $\mathbf{y} \in K$,

$$|\log \left(|\mathbf{x} - \mathbf{y}| / |\mathbf{x}|\right)| \leq C/ |\mathbf{x}|].$$

3.12. Prove the representation formula (3.56) in dimension two.

3.13. Compute the Green function for the circle of radius R.
[*Answer*:

$$G \left(\mathbf{x}, \mathbf{y}\right) = -\frac{1}{2\pi} [\log |\mathbf{x} - \mathbf{y}| - \log(\frac{|\mathbf{x}|}{R} \, |\mathbf{x}^* - \mathbf{y}|)],$$

where $\mathbf{x}^* = R^2 \mathbf{x} |\mathbf{x}|^{-2}$, $\mathbf{x} \neq \mathbf{0}]$.

3.14. Let $\Omega \subset \mathbb{R}^n$ be a bounded smooth domain and G be the Green function in Ω. Prove that, for every $\mathbf{x}, \mathbf{y} \in \Omega$, $\mathbf{x} \neq \mathbf{y}$:

(a) $G \left(\mathbf{x}, \mathbf{y}\right) > 0$;

(b) $G \left(\mathbf{x}, \mathbf{y}\right) = G \left(\mathbf{y}, \mathbf{x}\right)$.

[*Hint.* (a) Let $B_r \left(\mathbf{x}\right) \subset \Omega$ and let w be the harmonic function in $\Omega \backslash \overline{B}_r \left(\mathbf{x}\right)$ such that $w = 0$ on $\partial\Omega$ and $w = 1$ on $\partial B_r \left(\mathbf{x}\right)$. Show that, for every r small enough,

$$G \left(\mathbf{x}, \cdot\right) > w \left(\cdot\right)$$

in $\Omega \backslash \overline{B}_r \left(\mathbf{x}\right)$.

(b) For fixed $\mathbf{x} \in \Omega$, define $w_1 \left(\mathbf{y}\right) = G \left(\mathbf{x}, \mathbf{y}\right)$ and $w_2 \left(\mathbf{y}\right) = G \left(\mathbf{y}, \mathbf{x}\right)$. Apply Green's identity (3.57) in $\Omega \backslash B_r \left(\mathbf{x}\right)$ to w_1 and w_2. Let $r \to 0]$.

3.15. Compute the Green function for the half plane $\mathbb{R}^2_+ = \{(x, y) ; y > 0\}$ and (formally) derive the Poisson formula

$$u \left(x, y\right) = \frac{y}{\pi} \int_{\mathbb{R}} \frac{u \left(x, 0\right)}{\left(x - \xi\right)^2 + y^2} d\xi$$

for a bounded harmonic function in $\mathbb{R}^2_+$.

3.16. Prove that the exterior Dirichlet problem in the plane has a unique bounded solution $u \in C^2(\Omega_e) \cap C(\overline{\Omega}_e)$, through the following steps. Let w be the difference of two solutions. Then w is harmonic Ω_e, vanishes on $\partial \Omega_e$ and is bounded, say $|w| \leq M$.

1) Assume that the $\mathbf{0} \in \Omega$. Let $B_a(\mathbf{0})$ and $B_R(\mathbf{0})$ such that

$$B_a(\mathbf{0}) \subset \Omega \subset B_R(\mathbf{0})$$

and define

$$u_R(\mathbf{x}) = M \frac{\ln |\mathbf{x}| - \ln |a|}{\ln R - \ln a}.$$

Use the maximum principle to conclude that $w \leq u_R$, in the ring

$$B_{a,R} = \left\{ \mathbf{x} \in \mathbb{R}^2; a < |\mathbf{x}| < R \right\}.$$

2) Let $R \to \infty$ and deduce that $w \leq 0$ in Ω_e.

3) Proceed similarly to show that $w \geq 0$ in Ω_e.

3.17. Find the Poisson formula for the circle B_R, by representing the solution of $\Delta u = 0$ in B_R, $u = g$ on ∂B_R, as a double layer potential.

3.18. Consider the exterior Neumann-Robin problem in $\mathbb{R}^3$

$$\begin{cases} \Delta u = 0 & \text{in } \Omega_e \\ \partial_\nu u + ku = g & \text{on } \partial \Omega_e, \ (k \geq 0) \\ u \to 0 & \text{as } |\mathbf{x}| \to \infty. \end{cases} \tag{3.109}$$

(a) Show that the condition

$$\int_{\partial \Omega} g d\sigma = 0$$

is necessary for the solvability of (3.109) if $k = 0$.

(b) Represent the solution as a single layer potential and derive the integral equations for the unknown density.

[*Hint.* (a) Show that, for R large,

$$\int_{\partial \Omega} g \, d\sigma = \int_{\{|\mathbf{x}|=R\}} \partial_\nu u \, d\sigma.$$

Then let $R \to \infty$ and use Corollary 3.2].

3.19. Solve (formally) the Neumann problem in the half space $\mathbb{R}^3_+$, using a single layer potential.

3.20. Let $B = B_1(\mathbf{0}) \subset \mathbb{R}^2$. To complete the proof of Theorem 3.6 we must show that, if $g \in C(\partial B)$ and u is given by formula (3.21) with $R = 1$ and $\mathbf{p} = \mathbf{0}$, then

$$\lim_{\mathbf{x} \to \boldsymbol{\xi}} u(\mathbf{x}) = g(\boldsymbol{\xi}) \qquad \text{for every } \boldsymbol{\xi} \in \partial B.$$

Fill in the details in the following steps and conclude the proof.

1. First show that:

$$\frac{1 - |\mathbf{x}|^2}{2\pi} \int_{\partial B} \frac{1}{|\mathbf{x} - \boldsymbol{\sigma}|^2} d\sigma = 1$$

and that therefore

$$u(\mathbf{x}) - g(\boldsymbol{\xi}) = \frac{1 - |\mathbf{x}|^2}{2\pi} \int_{\partial B} \frac{g(\boldsymbol{\sigma}) - g(\boldsymbol{\xi})}{|\mathbf{x} - \boldsymbol{\sigma}|^2} d\sigma.$$

2. For $\delta > 0$, write

$$u(\mathbf{x}) - g(\boldsymbol{\xi}) = \frac{1 - |\mathbf{x}|^2}{2\pi} \int_{\partial B \cap \{|\boldsymbol{\sigma} - \boldsymbol{\xi}| < \delta\}} \cdots d\sigma + \frac{1 - |\mathbf{x}|^2}{2\pi} \int_{\partial B \cap \{|\boldsymbol{\sigma} - \boldsymbol{\xi}| > \delta\}} \cdots d\sigma$$
$$\equiv I + II.$$

Fix $\varepsilon > 0$, and use the continuity of g to show that, if δ is small enough, then

$$|I| < \varepsilon.$$

3. Show that, if $|\mathbf{x} - \boldsymbol{\xi}| < \delta/2$ and $|\boldsymbol{\sigma} - \boldsymbol{\xi}| > \delta$, then $|\mathbf{x} - \boldsymbol{\sigma}| > \delta/2$ and therefore

$$\lim_{\mathbf{x} \to \boldsymbol{\xi}} II = 0.$$

3.21. Consider the equation

$$Lu \equiv \Delta u + k^2 u = 0 \qquad \text{in } \mathbb{R}^3$$

called *Helmoltz* or *reduced wave equation*.

(a) Show that the radial solutions $u = u(r)$, $r = |\mathbf{x}|$, satisfying the *outgoing Sommerfeld* condition

$$u_r + iku = O\left(\frac{1}{r^2}\right) \qquad \text{as } r \to \infty,$$

are of the form

$$\varphi(r; k) = c\frac{e^{-ikr}}{r} \qquad c \in \mathbb{C}.$$

(b) For f smooth and compactly supported in $\mathbb{R}^3$ define the potential

$$U(\mathbf{x}) = c_0 \int_{\mathbb{R}^3} f(\mathbf{y}) \frac{e^{-ik|\mathbf{x} - \mathbf{y}|}}{|\mathbf{x} - \mathbf{y}|} d\mathbf{y}.$$

Select the constant c_0 such that

$$LU(\mathbf{x}) = -f(\mathbf{x}).$$

[*Answer* (b): $c_0 = (4\pi)^{-1}$].

4

Scalar Conservation Laws and First Order Equations

Introduction – Linear Transport Equation – Traffic Dynamics – Integral (or Weak) Solutions – The Method of Characteristics For Quasilinear Equations – General First Order Equations

4.1 Introduction

In the first part of this chapter we consider equations of the form

$$u_t + q(u)_x = 0, \qquad x \in \mathbb{R},\ t > 0. \tag{4.1}$$

In general, $u = u(x,t)$ represents the *density* or the *concentration* of a physical quantity Q and $q(u)$ is its *flux function*[1]. Equation (4.1) constitutes a *link* between density and flux and expresses a (**scalar**) **conservation law** for the following reason. If we consider a control interval $[x_1, x_2]$, the integral

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

gives the amount of Q between x_1 and x_2 at time t. A *conservation law* states that, without sources or sinks, the rate of change of Q in the interior of $[x_1, x_2]$ is determined by the net flux through the end points of the interval. If the flux is modelled by a function $q = q(u)$, the law translates into the equation

$$\frac{d}{dt}\int_{x_1}^{x_2} u(x,t)\,dx = -q(u(x_2,t)) + q(u(x_1,t)), \tag{4.2}$$

where we assume that $q > 0$ ($q < 0$) for a flux along the positive (negative) direction of the x axes. If u and q are smooth functions, equation (4.2) can be rewritten in the form

$$\int_{x_1}^{x_2} [u_t(x,t) + q(u(x,t))_x]\,dx = 0$$

which implies (4.1), due to the arbitrariness of the interval $[x_1, x_2]$.

---

[1] The dimensions of q are $[mass] \times [time]^{-1}$.

At this point we have to decide which type of flux function we are dealing with, or, in other words, we have to establish a *constitutive relation for q*.

In the next section we go back to the model of pollution in a channel, considered in section 2.5.3, neglecting the diffusion and choosing for q a linear function of u, namely:
$$q\left(u\right) = vu,$$
where v is constant. The result is a *pure transport* model, in which the vector $v\mathbf{i}$ is the *advection*[2] *speed*. In the sequel, to introduce and motivate some important concepts and results, we shall use a nonlinear model from traffic dynamics, with speed v depending on u.

The conservation law (4.1) occurs, for instance, in $1-$dimensional fluid dynamics where it often describes the formation and propagation of the so called *shock waves*. Along a shock curve a solution undergoes a *jump discontinuity* and an important question is how to reinterpret the differential equation (4.1) in order to admit discontinuous solutions.

A typical problem associated with equation (4.1) is the *initial value problem*:

$$\begin{cases} u_t + q\left(u\right)_x = 0 \\ u\left(x,0\right) = g\left(x\right) \end{cases} \tag{4.3}$$

where $x \in \mathbb{R}$. Sometimes x varies in a half-line or in a finite interval; in these cases some other conditions have to be added to obtain a well posed problem.

4.2 Linear Transport Equation

4.2.1 Pollution in a channel

We go back to the simple model for the evolution of a pollutant in a narrow channel, considered in section 2.5.3. When diffusion and transport are both relevant we have derived the equation
$$c_t = Dc_{xx} - vc_x,$$
where c is the concentration of the pollutant and $v\mathbf{i}$ is the stream speed ($v > 0$, constant). We want to discuss here the case of the *pure transport* equation

$$c_t + vc_x = 0 \tag{4.4}$$

i.e. when $D = 0$. Introducing the vector

$$\mathbf{v} = v\mathbf{i} + \mathbf{j}$$

equation (4.4) can be written in the form

$$vc_x + c_t = \nabla c \cdot \mathbf{v} = 0,$$

---

[2] Advection is usually synonymous of *linear convection*.

pointing out the orthogonality of ∇c and $\mathbf{v}$. But ∇c is orthogonal to the level lines of c, along which c is constant. Therefore the level lines of c are the straight lines parallel to $\mathbf{v}$, of equation

$$x = vt + x_0.$$

These straight lines are called **characteristics**. Let us compute c along the characteristic $x = vt + x_0$ letting

$$w(t) = c(x_0 + vt, t).$$

Since[3]

$$\dot{w}(t) = vc_x(x_0 + vt, t) + c_t(x_0 + vt, t),$$

equation (4.4) becomes the *ordinary differential equation* $\dot{w}(t) = 0$ which implies that c is constant along the characteristic.

We want to determine the evolution of the concentration c, by knowing its initial profile

$$c(x, 0) = g(x). \tag{4.5}$$
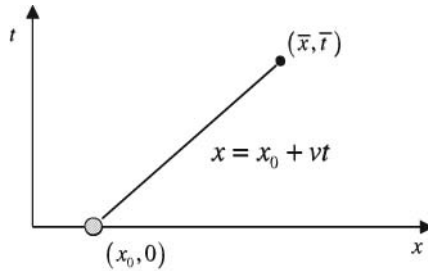
Fig. 4.1. Characteristic line for the linear transport problem

The method to compute the solution at a point $(\bar{x}, \bar{t})$, $t > 0$, is very simple. Let $x = vt + x_0$ be the equation of the characteristic passing through $(\bar{x}, \bar{t})$.

Go back in time along this characteristic from $(\bar{x}, \bar{t})$ until the point $(x_0, 0)$, of intersection with the $x-$ axes (see Fig. 4.1).

Since c is constant along the characteristic and $c(x_0, 0) = g(x_0)$, it must be

$$c(\bar{x}, \bar{t}) = g(x_0) = g(\bar{x} - v\bar{t}).$$

Thus, if $g \in C^1(\mathbb{R})$, the solution of the initial value problem (4.4), (4.5) is given by

$$c(x, t) = g(x - vt). \tag{4.6}$$

The solution (4.6) represents a *travelling wave*, moving with speed v in the positive $x-$direction. In figure 4.2, an initial profile $g(x) = \sin(\pi x)\chi_{[0,1]}(x)$ is *transported* in the plane x, t along the straight-lines $x - t = $ constant, i.e. with speed $v = 1$.

---

[3] The *dot* denotes derivative with respect to time.
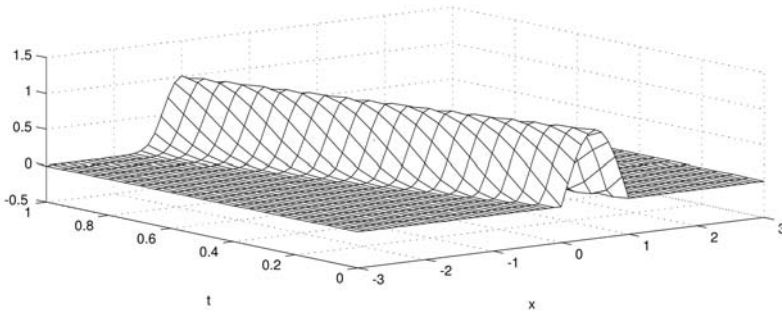
Fig. 4.2. Travelling wave solution of the linear transport equation

4.2.2 Distributed source

The equation

$$c_t + vc_x = f(x, t), \tag{4.7}$$

with the initial condition

$$c(x, 0) = g(x), \tag{4.8}$$

describes the effect of an external distributed source along the channel. The function f represents the intensity of the source, measured in concentration per unit time.

Again, to compute the value of the solution u at a point $(\bar{x}, \bar{t})$ is not difficult. Let $x = x_0 + vt$ be the characteristic passing through $(\bar{x}, \bar{t})$ and compute u along this characteristic, setting $w(t) = c(x_0 + vt, t)$. From (4.7), w satisfies the *ordinary differential equation*

$$\dot{w}(t) = vc_x(x_0 + vt, t) + c_t(x_0 + vt, t) = f(x_0 + vt, t)$$

with the initial condition

$$w(0) = g(x_0).$$

Thus

$$w(t) = g(x_0) + \int_0^t f(x_0 + vs, s)\, ds.$$

Letting $t = \bar{t}$ and recalling that $x_0 = \bar{x} - v\bar{t}$, we get

$$c(\bar{x}, \bar{t}) = w(\bar{t}) = g(\bar{x} - v\bar{t}) + \int_0^{\bar{t}} f(\bar{x} - v(\bar{t} - s), s)\, ds. \tag{4.9}$$

Since $(\bar{x}, \bar{t})$ is arbitrary, if g and f are reasonably smooth functions, (4.9) is our solution.

Proposition 4.1. *Let $g \in C^1(\mathbb{R})$ and $f, f_x \in C(\mathbb{R} \times \mathbb{R}_+)$. The solution of the initial value problem*

$$\begin{cases} c_t + vc_x = f(x, t) & x \in \mathbb{R},\, t > 0 \\ c(x, 0) = g(x) & x \in \mathbb{R} \end{cases}$$

is given by the formula

$$c(x,t) = g(x - vt) + \int_0^t f(x - v(t-s), s)\, ds. \tag{4.10}$$

Remark 4.1. Formula (4.10) can be derived using the *Duhamel method,* as in section 2.2.8 (see Problem 4.1.)

4.2.3 Decay and localized source

Suppose that, due to *biological decomposition*, the pollutant decays at the rate

$$r(x,t) = -\gamma c(x,t) \qquad \gamma > 0.$$

Without external sources and diffusion, the mathematical model is

$$c_t + vc_x = -\gamma c,$$

with the initial condition

$$c(x,0) = g(x).$$

Setting

$$u(x,t) = c(x,t)\, e^{\frac{\gamma}{v}x}, \tag{4.11}$$

we have

$$u_x = \left(c_x + \frac{\gamma}{v}c\right) e^{\frac{\gamma}{v}x} \quad \text{and} \quad u_t = c_t e^{\frac{\gamma}{v}x}$$

and therefore the equation for u is

$$u_t + vu_x = 0$$

with the initial condition

$$u(x,0) = g(x)\, e^{\frac{\gamma}{v}x}.$$

From Proposition 4.1, we get

$$u(x,t) = g(x - vt)\, e^{\frac{\gamma}{v}(x-vt)}$$

and from (4.11)

$$c(x,t) = g(x - vt)\, e^{-\gamma t}$$

which is a *damped travelling wave.*

We now examine the effect of a source of pollutant placed at a certain point of the channel, e.g. at $x = 0$. Typically, one can think of waste material from industrial machineries. Before the machines start working, for instance before time $t = 0$, we assume that the channel is clean. We want to determine the pollutant concentration, supposing that at $x = 0$ it is kept at a constant level $\beta > 0$, for $t > 0$.

To model this source we introduce the Heaviside function

$$\mathcal{H}(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0, \end{cases}$$

with the *boundary condition*

$$c(0, t) = \beta \mathcal{H}(t) \qquad (4.12)$$

and the initial condition

$$c(x, 0) = 0 \qquad \text{for } x > 0. \qquad (4.13)$$

As before, let $u(x, t) = c(x, t) e^{\frac{\gamma}{v}x}$, which is a solution of $u_t + vu_x = 0$. Then:

$$u(x, 0) = c(x, 0) e^{\frac{\gamma}{v}x} = 0 \qquad x > 0$$
$$u(0, t) = c(0, t) = \beta \mathcal{H}(t).$$

Since u is constant along the characteristics it must be of the form

$$u(x, t) = u_0(x - vt) \qquad (4.14)$$

where u_0 is to be determined from the boundary condition (4.12) and the initial condition (4.13).

To compute u for $x < vt$, observe that a characteristic leaving the $t-$axis from a point $(0, t)$ carries the data $\beta \mathcal{H}(t)$. Therefore, we must have

$$u_0(-vt) = \beta \mathcal{H}(t).$$

Letting $s = -vt$ we get

$$u_0(s) = \beta \mathcal{H}\left(-\frac{s}{v}\right)$$

and from (4.14),

$$u(x, t) = \beta \mathcal{H}\left(t - \frac{x}{v}\right).$$

This formula gives the solution also in the sector

$$x > vt, \quad t > 0,$$

since the characteristics leaving the $x-$axis carry *zero data* and hence we deduce $u = c = 0$ there. This means that the pollutant has not yet reached the point x at time t, if $x > vt$.

Finally, recalling (4.11), we find

$$c(x, t) = \beta \mathcal{H}\left(t - \frac{x}{v}\right) e^{-\frac{\gamma}{v}x}.$$

Observe that in $(0, 0)$ there is a *jump discontinuity which is transported along the characteristic $x = vt$. Figure 4.3 shows the solution for $\beta = 3$, $\gamma = 0.7$, $v = 2$.
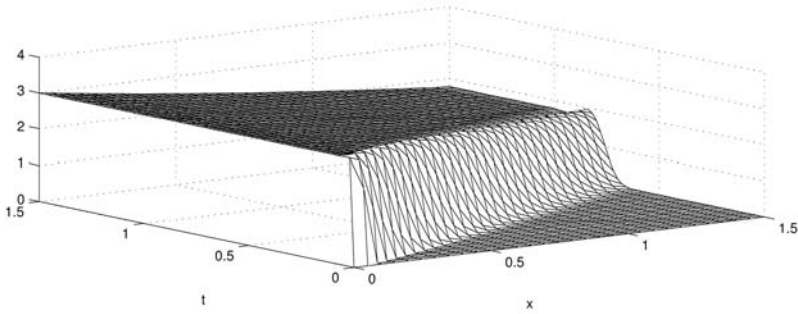
Fig. 4.3. Propagation of a discontinuity

4.2.4 Inflow and outflow characteristics. A stability estimate

The domain in the localized source problem is the quadrant $x > 0, t > 0$. To uniquely determine the solution we have used the initial data on the $x-$axis, $x > 0$, and the boundary data on the $t-$axis, $t > 0$. The problem is therefore well posed. This is due to the fact that, since $v > 0$, when time increases, *all* the characteristics carry the information (the data) *towards the interior* of the quadrant $x > 0, t > 0$. In other words the characteristics are **inflow characteristics**.

More generally, consider the equation

$$u_t + au_x = f(x, t)$$

in the domain $x > 0$, $t > 0$, where a is a constant $(a \neq 0)$. The characteristics are the lines

$$x - at = \text{constant}$$

as shown in figure 4.4. If $a > 0$, we are in the case of the pollutant model: all the characteristics **are inflow** and **the data must be assigned on both semi-axes.**

If $a < 0$, the characteristics leaving the $x-$ axis are **inflow**, while those leaving the $t-$ axis are **outflow**. In this case the initial data alone are sufficient to uniquely determine the solution, while **no data has to be assigned on the semi-axis** $x = 0$, $t > 0$.

Coherently, a problem in the half-strip $0 < x < R$, $t > 0$, besides the initial data, requires a data assignment on the inflow boundary, namely (Fig. 4.4):

$$\begin{cases} u(0, t) = h_0(t) & \text{if } a > 0 \\ u(R, t) = h_R(t) & \text{if } a < 0. \end{cases}$$

The resulting initial-boundary value problem is well posed, since the solution is uniquely determined at every point in the strip by its values along the characteristics. Moreover, a stability estimate can be proved as follows. Consider, for instance,
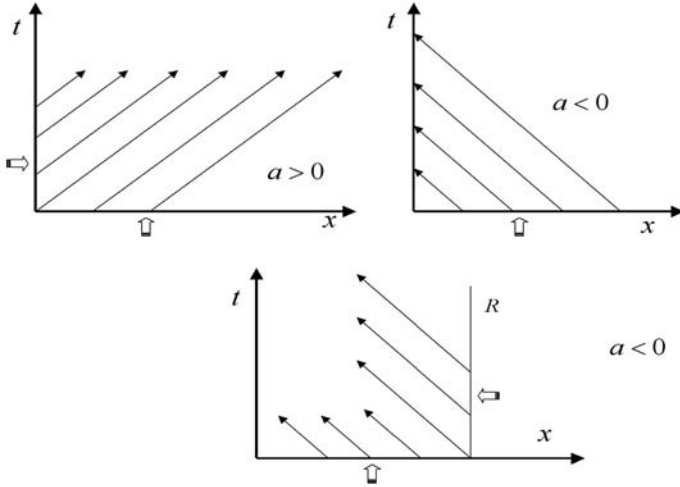
Fig. 4.4. The arrows indicate where the data should be assigned

the case $a > 0$ and the problem[4]

$$\begin{cases} u_t + a u_x = 0 & 0 < x < R, t > 0 \\ u(0, t) = h(t) & t > 0 \\ u(x, 0) = g(x) & 0 < x < R. \end{cases} \qquad (4.15)$$

Multiply the differential equation by u and write

$$u u_t + a u u_x = \frac{1}{2} \frac{d}{dt} u^2 + \frac{a}{2} \frac{d}{dx} u^2 = 0.$$

Integrating in x over $(0, R)$ we get:

$$\frac{d}{dt} \int_0^R u^2(x, t)\, dx + a\left[u^2(R, t) - u^2(0, t)\right] = 0.$$

Now use the data $u(0, t) = h(t)$ and the positivity of a to obtain

$$\frac{d}{dt} \int_0^R u^2(x, t)\, dx \le a h^2(t).$$

Integrating in t we have, using the initial condition $u(x, 0) = g(x)$,

$$\int_0^R u^2(x, t)\, dx \le \int_0^R g^2(x)\, dx + a \int_0^t h^2(s)\, ds. \qquad (4.16)$$

Now, let u_1 and u_2 be solutions of problem (4.15) with initial data g_1, g_2 and boundary data h_1, h_2 on $x = 0$. Then, by linearity, $w = u_1 - u_2$ is a solution

---

[4] For the case $u_t + a u_x = f \ne 0$, see Problem 4.2.

of problem (4.15) with initial data $g_1 - g_2$ and boundary data $h_1 - h_2$ on $x = 0$. Applying the inequality (4.16) to w we have

$$\int_0^R [u_1(x,t) - u_2(x,t)]^2 dx \le \int_0^R [g_1(x) - g_2(x)]^2 dx + a \int_0^t [h_1(s) - h_2(s)]^2 ds.$$

Thus, a least-squares approximation of the data controls a least-squares approximation of the corresponding solutions. In this sense, the solution of problem (4.15) depends continuously on the initial data and on the boundary data on $x = 0$. We point out that the values of u on $x = R$ do not appear in (4.16).

4.3 Traffic Dynamics

4.3.1 A macroscopic model

From far away, an intense traffic on a highway can be considered as a fluid flow and described by means of macroscopic variables such as the *density* of cars[5] ρ, their *average speed* v and their *flux*[6] q. The three (more or less regular) functions ρ, u and q are linked by the simple convection relation

$$q = v\rho.$$

To construct a model for the evolution of ρ we assume the following hypotheses.

1. *There is only one lane and overtaking is not allowed.* This is realistic for instance for traffic in a tunnel (see Problem 4.7). Multi-lanes models with overtaking are beyond the scope of this introduction. However the model we will present is often in agreement with observations also in this case.

2. *No car "sources" or "sinks".* We consider a road section without exit/entrance gates.

3. *The average speed is not constant and depends on the density alone*, that is

$$v = v(\rho).$$

This rather controversial assumption means that at a certain density the speed is uniquely determined and that a density change causes an immediate speed variation. Clearly

$$v'(\rho) = \frac{dv}{d\rho} \le 0$$

since we expect the speed to decrease as the density increases.

As in Section 4.1, from hypotheses **2** and **3** we derive the conservation law:

$$\rho_t + q(\rho)_x = 0 \tag{4.17}$$

---

[5] Number of cars per unit length.
[6] Cars per unit time.

where

$$q(\rho) = v(\rho)\,\rho.$$

We need a constitutive relation for $v = v(\rho)$. When ρ is small, it is reasonable to assume that the average speed v is more or less equal to the maximal velocity v_m, given by the speed limit. When ρ increases, traffic slows down and stops at the maximum density ρ_m (bumper-to-bumper traffic). We adopt the simplest model consistent with the above considerations[7], namely

$$v(\rho) = v_m\left(1 - \frac{\rho}{\rho_m}\right),$$

so that

$$q(\rho) = v_m\rho\left(1 - \frac{\rho}{\rho_m}\right). \tag{4.18}$$

Since

$$q(\rho)_x = q'(\rho)\,\rho_x = v_m\left(1 - \frac{2\rho}{\rho_m}\right)\rho_x$$

equation (4.17) becomes

$$\rho_t + \underbrace{v_m\left(1 - \frac{2\rho}{\rho_m}\right)}_{q'(\rho)}\rho_x = 0. \tag{4.19}$$

According to the terminology in Section 1.1, this is a *quasilinear* equation. We also point out that

$$q''(\rho) = -\frac{2v_m}{\rho_m} < 0$$

so that q is strictly *concave*. We couple the equation (4.19) with the initial condition

$$\rho(x, 0) = g(x). \tag{4.20}$$

4.3.2 The method of characteristics

We want to solve the initial value problem (4.19), (4.20). To compute the density ρ at a point (x, t) we follow the idea we used in the linear transport case: *to connect the point (x, t) with a point $(x_0, 0)$ on the $x-$axis, through a curve along which ρ is constant* (Fig. 4.5).

Clearly, if we manage to find such a curve, that we call **characteristic based at** $(x_0, 0)$, the value of ρ at (x, t) is given by $\rho(x_0, 0) = g(x_0)$. Moreover, if this procedure can be repeated for every point (x, t), $x \in \mathbb{R}$, $t > 0$, then we can compute ρ at every point and the problem is completely solved. This is the *method of characteristics*.

---

[7] And in good agreement with experimental data.
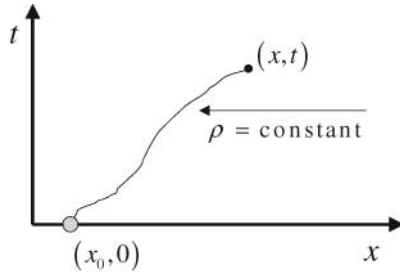
Fig. 4.5. Characteristic curve

Adopting a slightly different point of view, we can implement the above idea as follows: assume that $x = x(t)$ is the equation of the characteristic based at the point $(x_0, 0)$; along $x = x(t)$ we *observe always the same initial density $g(x_0)$*. In other words

$$\rho(x(t), t) = g(x_0) \tag{4.21}$$

for every $t > 0$. If we differentiate identity (4.21), we get

$$\frac{d}{dt}\rho(x(t), t) = \rho_x(x(t), t) x'(t) + \rho_t(x(t), t) = 0 \qquad (t > 0). \tag{4.22}$$

On the other hand, (4.19) yields

$$\rho_t(x(t), t) + q'(g(x_0)) \rho_x(x(t), t) = 0$$

so that, subtracting (4.23) from (4.22), we obtain

$$\rho_x(x(t), t) [\dot{x}(t) - q'(g(x_0))] = 0. \tag{4.23}$$

Assuming $\rho_x(x(t), t) \neq 0$, we deduce

$$\dot{x}(t) = q'(g(x_0)).$$

Since $x(0) = x_0$ we find

$$x(t) = q'(g(x_0)) t + x_0. \tag{4.24}$$

Thus, the characteristics are **straight lines** with slope $q'(g(x_0))$. Different values of x_0 give, in general, different values of the slope.

We can now derive a formula for ρ. To compute $\rho(x, t)$, $t > 0$, go back in time along the characteristic through (x, t) until its base point $(x_0, 0)$. Then $\rho(x, t) = g(x_0)$. From (4.24) we have, since $x(t) = x$,

$$x_0 = x - q'(g(x_0)) t$$

and finally

$$\rho(x, t) = g(x - q'(g(x_0)) t). \tag{4.25}$$
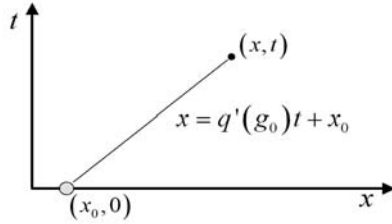
Fig. 4.6. Characteristic straight line ($g_0 = g(x_0)$)

Formula (4.25) represents **a travelling wave propagating with speed** $q'(g(x_0))$ along the positive $x-$direction.

We emphasize that $q'(g(x_0))$ is the *local wave speed* and it must not be confused with the traffic velocity. In fact, in general,

$$\frac{dq}{d\rho} = \frac{d(\rho v)}{d\rho} = v + \rho\frac{dv}{d\rho} \leq v$$

since $\rho \geq 0$ and $\frac{dv}{d\rho} \leq 0$.

The different nature of the two speeds becomes more evident if we observe that the wave speed *may be negative* as well. This means that, while the traffic advances along the positive $x-$direction, the disturbance given by the travelling wave may propagate in the opposite direction. Indeed, in our model (4.18), $\frac{dq}{d\rho} < 0$ when $\rho > \frac{\rho_m}{2}$.

Formula (4.25) seems to be rather satisfactory, since, apparently, it gives the solution of the initial value problem (4.19), (4.20) at every point. Actually, a more accurate analysis shows that, even if the initial data g are smooth, the solution may develop a singularity in finite time (e.g. a jump discontinuity). When this occurs, the method of characteristics does not work anymore and formula (4.25) is not effective. A typical case is described in figure 4.7: two characteristics based at different points $(x_1, 0)$ e $(x_2, 0)$ intersect at the point (x, t) and the value $u(x, t)$ is not uniquely determined as soon as $g(x_1) \neq g(x_2)$.
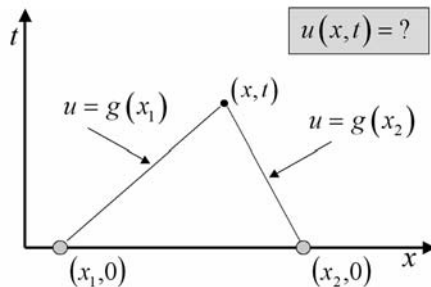
Fig. 4.7. Intersection of characteristics

In this case we have to weaken the concept of solution and the computation technique. We will come back on these questions later. For the moment, we analyze the method of characteristics in some particularly significant cases.

4.3.3 The green light problem

Suppose that bumper-to-bumper traffic is standing at a red light, placed at $x = 0$, while the road ahead is empty. Accordingly, the initial density profile is

$$g\left(x\right) = \begin{cases} \rho_m & \text{for } x \leq 0 \\ 0 & \text{for } x > 0. \end{cases}$$

At time $t = 0$ the traffic light turns green and we want to describe the car flow evolution for $t > 0$. At the beginning, only the cars nearer to the light start moving while most remain standing.

Since $q'\left(\rho\right) = v_m \left(1 - \frac{2\rho}{\rho_m}\right)$, the local wave speed is given by

$$q'\left(g\left(x_0\right)\right) = \begin{cases} -v_m & \text{for } x_0 \leq 0 \\ v_m & \text{for } x_0 > 0 \end{cases}$$

and the characteristics are the straight lines

$$x = -v_m t + x_0 \qquad \text{if } x_0 < 0$$
$$x = v_m t + x_0 \qquad \text{if } x_0 > 0.$$

The lines $x = v_m t$ and $x = -v_m t$ partition the upper half-plane in the three regions R, S and T, shown in figure 4.8.
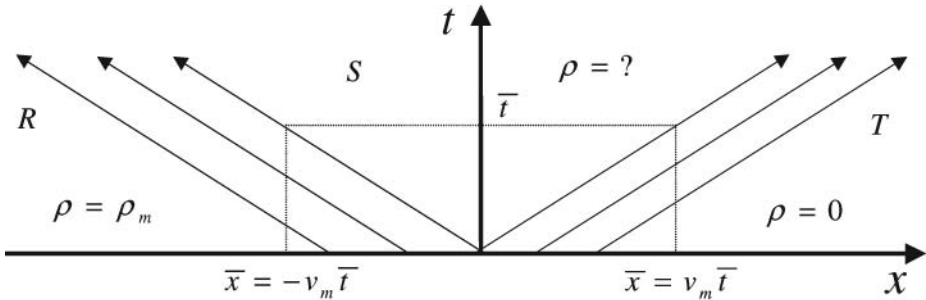
Fig. 4.8. Characteristics for the green light problem

Inside R we have $\rho\left(x, t\right) = \rho_m$, while inside T we have $\rho\left(x, t\right) = 0$. Consider the points on the horizontal line $t = \bar{t}$. At the points $\left(x, \bar{t}\right) \in T$ the density is zero: the traffic has not yet arrived in x at time $t = \bar{t}$. The front car is located at the point

$$\bar{x} = v_m \bar{t}$$

which moves at the maximum speed, since ahead the road is empty.

The cars placed at the points $(x, \bar{t}) \in R$ are still standing. The first car that starts moving at time $t = \bar{t}$ is at the point

$$\bar{x} = -v_m \bar{t}.$$

In particular, it follows that *the green light signal propagates back through the traffic at the speed v_m.*

What is the value of the density inside the sector S? No characteristic extends into S due to the discontinuity of the initial data at the origin, and the method as it stands does not give any information on the value of ρ inside S.

A strategy that may give a reasonable answer is the following:

a) approximate the initial data by a continuous function g_ε, which converges to g as $\varepsilon \to 0$ at every point x, except 0;

b) construct the solution ρ_ε of the ε−problem by the method of characteristics;

c) let $\varepsilon \to 0$ and check that the limit of ρ_ε is a solution of the original problem.

Clearly we run the risk of constructing many solutions, each one depending on the way we regularize the initial data, but for the moment we are satisfied if we construct at least one solution.

a) Let us choose as g_ε the function (Fig. 4.9)

$$g_\varepsilon(x) = \begin{cases} \rho_m & x \le 0 \\ \rho_m(1 - \dfrac{x}{\varepsilon}) & 0 < x < \varepsilon \\ 0 & x \ge \varepsilon. \end{cases}$$

When $\varepsilon \to 0$, $g_\varepsilon(x) \to g(x)$ for every $x \ne 0$.
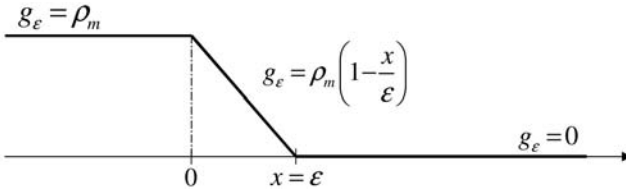
Fig. 4.9. Smoothing of the initial data in the green light problem

b) The characteristics for the ε−problem are:

$$\begin{array}{ll} x = -v_m t + x_0 & \text{if } x_0 < 0 \\ x = -v_m \left(1 - 2\dfrac{x_0}{\varepsilon}\right) t + x_0 & \text{if } 0 \le x_0 < \varepsilon \\ x = v_m t + x_0 & \text{if } x_0 \ge \varepsilon \end{array}$$

since, for $0 \le x_0 < \varepsilon$,

$$q'(g_\varepsilon(x_0)) = v_m \left(1 - \frac{2g_\varepsilon(x_0)}{\rho_m}\right) = -v_m \left(1 - 2\frac{x_0}{\varepsilon}\right).$$

The characteristics in the region $-v_m t < x < v_m t + \varepsilon$ form a *rarefaction fan* (Fig. 4.10).
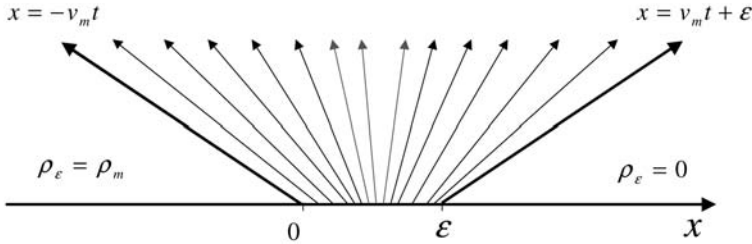
Fig. 4.10. Fanlike characteristics

Clearly, $\rho_\varepsilon (x, t) = 0$ for $x \geq v_m t + \varepsilon$ and $\rho_\varepsilon (x, t) = \rho_m$ for $x \leq -v_m t$. Let now (x, t) belong to the region

$$-v_m t < x < v_m t + \varepsilon.$$

Solving for x_0 in the equation of the characteristic $x = -v_m \left(1 - 2\dfrac{x_0}{\varepsilon}\right) t + x_0$, we find

$$x_0 = \varepsilon \frac{x + v_m t}{2 v_m t + \varepsilon}.$$

Then

$$\rho_\varepsilon (x, t) = g_\varepsilon (x_0) = \rho_m (1 - \frac{x_0}{\varepsilon}) = \rho_m \left(1 - \frac{x + v_m t}{2 v_m t + \varepsilon}\right). \tag{4.26}$$

c) Letting $\varepsilon \to 0$ in (4.26) we obtain

$$\rho (x, t) = \begin{cases} \rho_m & \text{for } x \leq -v_m t \\ \dfrac{\rho_m}{2} \left(1 - \dfrac{x}{v_m t}\right) & \text{for } -v_m t < x < v_m t \\ 0 & \text{for } x \geq v_m t \end{cases} . \tag{4.27}$$

It is easy to check that ρ is a solution of the equation (4.19) in the regions R, S, T. For fixed t, the function ρ decreases linearly from ρ_m to 0 as x varies from $-v_m t$ to $v_m t$. Moreover, ρ is constant on the fan of straight lines

$$x = h t \qquad -v_m < h < v_m.$$

These type of solutions are called **rarefaction** or **simple waves** (centered at the origin).

The formula for $\rho (x, t)$ in the sector S can be obtained, a posteriori, by a formal procedure that emphasizes its structure. The equation of the characteristics can be written in the form

$$x = v_m \left(1 - \frac{2 g (x_0)}{\rho_m}\right) t + x_0 = v_m \left(1 - \frac{2 \rho (x, t)}{\rho_m}\right) t + x_0.$$
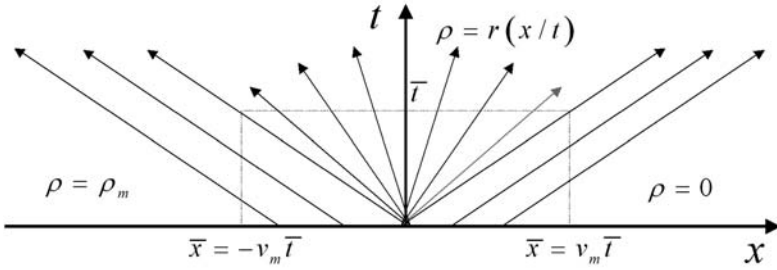
Fig. 4.11. Characteristics in a rarefaction wave
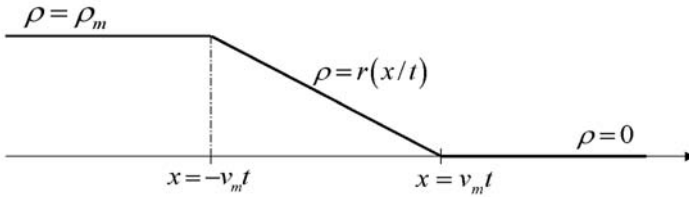
Fig. 4.12. Profile of a rarefaction wave at time t

because $\rho(x,t) = g(x_0)$. Inserting $x_0 = 0$ we obtain

$$x = v_m \left(1 - \frac{2\rho(x,t)}{\rho_m} \right) t.$$

Solving for ρ we find exactly

$$\rho(x,t) = \frac{\rho_m}{2} \left(1 - \frac{x}{v_m t} \right) \qquad (t > 0). \tag{4.28}$$

Since $v_m \left(1 - \frac{2\rho}{\rho_m} \right) = q'(\rho)$, we see that (4.28) is equivalent to

$$\rho(x,t) = r\left(\frac{x}{t} \right)$$

where $r = (q')^{-1}$ is the inverse function of q'. Indeed this is the general form of a rarefaction wave (centered at the origin) for a conservation law.

We have constructed a continuous solution ρ of the green light problem, connecting the two constant states ρ_m and 0 by a rarefaction wave. However, it is not clear in which sense ρ is a solution across the lines $x = \pm v_m t$, since, there, its derivatives undergo a jump discontinuity. Also, it is not clear whether or not (4.27) is the only solution. We will return on these important points.

4.3.4 Traffic jam ahead

Suppose that the initial density profile is

$$g(x) = \begin{cases} \frac{1}{8}\rho_m & \text{for } x < 0 \\ \rho_m & \text{for } x > 0. \end{cases}$$

For $x > 0$, the density is maximal and therefore the traffic is bumper-to-bumper. The cars on the left move with speed $v = \frac{7}{8}v_m$ so that we expect congestion propagating back into the traffic. We have

$$q'(g(x_0)) = \begin{cases} \frac{3}{4}v_m & \text{if } x_0 < 0 \\ -v_m & \text{if } x_0 > 0 \end{cases}$$

and therefore the characteristics are

$$x = \frac{3}{4}v_m t + x_0 \qquad \text{if } x_0 < 0$$
$$x = -v_m t + x_0 \qquad \text{if } x_0 > 0.$$

The characteristics configuration (Fig. 4.13) shows that the latter intersect somewhere in finite time and the theory predicts that ρ becomes a "multivalued" function of position. In other words, ρ should assume two different values at the same point, which clearly makes no sense in our situation. Therefore we have to admit solutions with jump discontinuities (**shocks**), but then we have to reexamine the derivation of the conservation law, because the smoothness assumption for ρ does not hold anymore.

Thus, let us go back to the conservation of cars in integral form (see (4.2)):

$$\frac{d}{dt}\int_{x_1}^{x_2} \rho(x,t)\ dx = -q(\rho(x_2,t)) + q(\rho(x_1,t)), \tag{4.29}$$

valid in any control interval $[x_1, x_2]$. Suppose now that ρ is a smooth function except along a curve

$$x = s(t) \qquad t \in [t_1, t_2],$$

that we call **shock curve**, on which ρ undergoes a *jump discontinuity*.

For fixed t, let $[x_1, x_2]$ be an interval containing the discontinuity point

$$x = s(t).$$

From(4.29) we have

$$\frac{d}{dt}\left\{\int_{x_1}^{s(t)} \rho(y,t)\,dy + \int_{s(t)}^{x_2} \rho(y,t)\,dy\right\} + q[\rho(x_2,t)] - q[\rho(x_1,t)] = 0. \tag{4.30}$$

The fundamental theorem of calculus gives

$$\frac{d}{dt}\int_{x_1}^{s(t)} \rho(y,t)\,dy = \int_{x_1}^{s(t)} \rho_t(y,t)\,dy + \rho^-(s(t),t)\frac{ds}{dt}$$
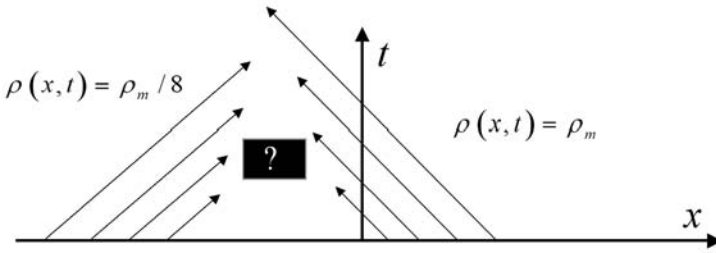
Fig. 4.13. Expecting a shock

and

$$\frac{d}{dt} \int_{s(t)}^{x_2} \rho(y, t)\, dy = \int_{s(t)}^{x_2} \rho_t(y, t)\, dy - \rho^+(s(t), t)\frac{ds}{dt},$$

where

$$\rho^-(s(t), t) = \lim_{y \uparrow s(t)} \rho(y, t), \qquad \rho^+(s(t), t) = \lim_{y \downarrow s(t)} \rho(y, t).$$

Hence, equation (4.30) becomes

$$\int_{x_1}^{x_2} \rho_t(y, t)\, dy + \left[\rho^-(s(t), t) - \rho^+(s(t), t)\right]\dot{s}(t) = q\left[\rho(x_1, t)\right] - q\left[\rho(x_2, t)\right].$$

Letting $x_2 \downarrow s(t)$ and $x_1 \uparrow s(t)$ we obtain

$$\left[\rho^-(s(t), t) - \rho^+(s(t), t)\right]\dot{s}(t) = q\left[\rho^-(s(t), t)\right] - q\left[\rho^+(s(t), t)\right]$$

that is:

$$\dot{s} = \frac{q\left[\rho^+(s, t)\right] - q\left[\rho^-(s, t)\right]}{\rho^+(s, t) - \rho^-(s, t)}. \tag{4.31}$$

The relation (4.31) is an ordinary differential equation for s and it is known as **Rankine-Hugoniot condition**. The discontinuity propagating along the shock curve is called **shock wave.** The Rankine-Hugoniot condition gives the *shock speed* $\dot{s}(t)$ as the quotient of the flux jump over the density jump. To determine the shock curve we need to know *its initial point* and the values of ρ from both sides of the curve.

Let us apply the above considerations to our traffic problem[8]. We have

$$\rho^+ = \rho_m, \qquad \rho^- = \frac{\rho_m}{8}$$

---

[8] In the present case the following simple formula holds:

$$\frac{q(w) - q(z)}{w - z} = v_m \left(1 - \frac{w + z}{\rho_m}\right).$$

while

$$q\left[\rho^+\right] = 0 \qquad q\left[\rho^-\right] = \frac{7}{64}v_m\rho_m$$

and (4.31) gives

$$\dot{s}\left(t\right) = \frac{q\left[\rho^+\right] - q\left[\rho^-\right]}{\rho^+ - \rho^-} = -\frac{1}{8}v_m.$$

Since clearly $s\left(0\right) = 0$, the shock curve is the straight line

$$x = -\frac{1}{8}v_m t.$$

Note that *the slope is negative*: the shock propagates back with speed $-\frac{1}{8}v_m$, as it is revealed by the braking of the cars, slowing down because of a traffic jam ahead.

As a consequence, the solution of our problem is given by the following formula (Fig. 4.14)

$$\rho\left(x, t\right) = \begin{cases} \frac{1}{8}\rho_m & x < -\frac{1}{8}v_m t \\ \rho_m & x > -\frac{1}{8}v_m t. \end{cases}$$

This time the two constant states $\frac{1}{8}\rho_m$ and ρ_m are connected by a **shock wave**.
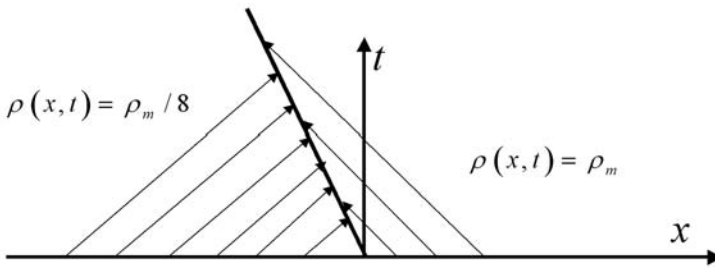
Fig. 4.14. Shock wave

4.4 Integral (or Weak) Solutions

4.4.1 The method of characteristics revisited

The method of characteristics applied to the problem

$$\begin{cases} u_t + q\left(u\right)_x = 0 \\ u\left(x, 0\right) = g\left(x\right) \end{cases} \tag{4.32}$$

gives the travelling wave (see (4.25) with $x_0 = \xi$)

$$u(x, t) = g\left[x - q'\left(g\left(\xi\right)\right)t\right] \qquad \left(q' = \frac{dq}{du}\right) \tag{4.33}$$

with local speed $q'(g(\xi))$, in the positive $x-$direction. Since $u(x,t) \equiv g(\xi)$ along the characteristic based at $(\xi, 0)$, from (4.33) we obtain that u is implicitly defined by the equation

$$G(x,t,u) \equiv u - g[x - q'(u)t] = 0. \tag{4.34}$$

If g and q' are smooth, the *Implicit Function Theorem,* implies that equation (4.34) defines u as a function of (x,t), as long as the condition

$$G_u(x,t,u) = 1 + tq''(u)g'[x - q'(u)t] \neq 0 \tag{4.35}$$

holds. An immediate consequence is that if q'' and g' have the same sign, the solution given by the method of characteristics is defined and smooth for all times $t \geq 0$. Precisely, we have:

Proposition 4.2. *Suppose that $q \in C^2(\mathbb{R})$, $g \in C^1(\mathbb{R})$ and $g'q'' \geq 0$ in $\mathbb{R}$. Then formula (4.34) defines the unique solution u of problem (4.32) in the half-plane $t \geq 0$. Moreover, $u(x,t) \in C^1(\mathbb{R} \times [0, \infty))$.*

Thus, if q'' and g' have the same sign, the characteristics do not intersect. Note that in the $\varepsilon-$approximation of the green light problem, q is concave and g_ε is decreasing. Although g_ε is not smooth, the characteristics do not intersect and ρ_ε is well defined for all times $t > 0$. In the limit as $\varepsilon \to 0$, the discontinuity of g reappears and the fan of characteristics produces a rarefaction wave.

What happens if q'' and g' have a different sign in an interval $[a, b]$? Proposition 4.2 still holds for small times, since $G_u \sim 1$ if $t \sim 0$, but when time goes on we expect the formation of a shock. Indeed, suppose, for instance, that q is concave and g is increasing. The family of characteristics based on a point in the interval $[a, b]$ is

$$x = q'(g(\xi))t + \xi \qquad \xi \in [a, b]. \tag{4.36}$$

When ξ increases, g increases as well, while $q'(g(\xi))$ decreases so that we expect intersection of characteristics along a shock curve. The main question is to find the positive time t_s (*breaking time*) and the location x_s of **first appearance of the shock**.

According to the above discussion, the *breaking time* must coincide with the first time t at which the expression

$$G_u(x,t,u) = 1 + tq''(u)g'[x - q'(u)t]$$

becomes zero. Computing G_u along the characteristic (4.36), we have $u = g(\xi)$ and

$$G_u(x,t,u) = 1 + tq''(g(\xi))g'(\xi).$$

Assume that the nonnegative function

$$z(\xi) = -q''(g(\xi))g'(\xi)$$

attains its maximum only at the point $\xi_M \in [a, b]$. Then $z(\xi_M) > 0$ and

$$t_s = \min_{\xi \in [a,b]} \frac{1}{z(\xi)} = \frac{1}{z(\xi_M)}. \tag{4.37}$$

Since x_s belongs to the characteristics $x = q'(g(\xi_M))\, t + \xi_M$, we find

$$x_s = \frac{q'(g(\xi_M))}{z(\xi_M)} + \xi_M. \tag{4.38}$$

The point (x_s, t_s) has an interesting geometrical meaning. In fact, it turns out that if $q''g' < 0$, the family of characteristics (4.36) admits an *envelope*[9] and (x_s, t_s) is the point on the envelope with minimum time coordinate (see Problem 4.8).
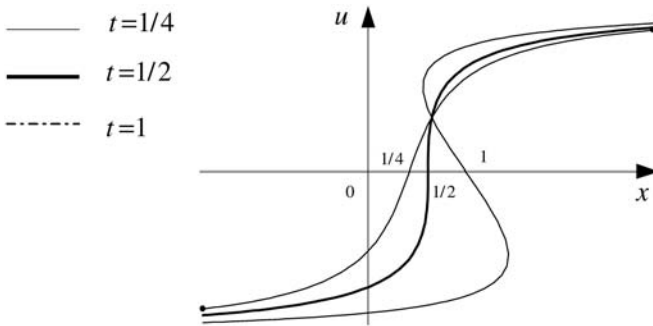
Fig. 4.15. Breaking time for problem (4.39)

Example 4.1. Consider the initial value problem

$$\begin{cases} u_t + (1 - 2u)u_x = 0 \\ u(x, 0) = \arctan x. \end{cases} \tag{4.39}$$

We have $q(u) = u - u^2$, $q'(u) = 1 - 2u$, $q''(u) = -2$, and $g(\xi) = \arctan \xi$, $g'(\xi) = 1/(1 + \xi^2)$. Therefore, the function

$$z(\xi) = -q''(g(\xi))g'(\xi) = \frac{2}{(1 + \xi^2)}$$

---

[9] Recall that the *envelope* of a family of curves $\phi(x, t, \xi) = 0$, depending on the parameter ξ, is a curve $\psi(x, t) = 0$ *tangent at each one of its points to a curve of the family*. If the family of curves $\phi(x, t, \xi) = 0$ has an envelope, its parametric equations are obtained by solving the system

$$\begin{cases} \phi(x, t, \xi) = 0 \\ \phi_\xi(x, t, \xi) = 0 \end{cases}$$

with respect to x and t.

has a maximum at $\xi_M = 0$ and $z(0) = 2$. The breaking-time is $t_S = 1/2$ and $x_S = 1/2$. Thus, the shock curve starts from $(1/2, 1/2)$. For $0 \leq t < 1/2$ the solution u is smooth and implicitly defined by the equation

$$u - \arctan[x - (1 - 2u)t] = 0. \tag{4.40}$$

After $t = 1/2$, equation (4.40) defines u as a multivalued function of (x, t) and does not define a solution anymore. Figure 4.15 shows what happens for $t = 1/4$, $1/2$ and 1. Note that the common point of intersection is $(1/2, \tan 1/2)$ which is not the first shock point.

How does the solution evolve after $t = 1/2$? We have to insert a shock wave into the multivalued graph in figure 4.15 in such a way the conservation law is preserved. We will see that the correct insertion point is prescribed by the Rankine-Hugoniot condition. It turns out that this corresponds to cutting off from the multivalued profile two **equal area** lobes A and B as described in figure 4.16 (*G. B. Whitham equal area rule*[10]).
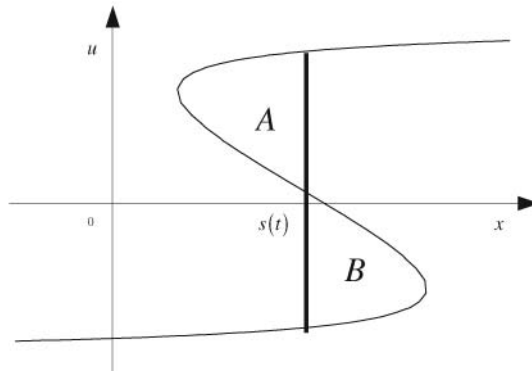
Fig. 4.16. Inserting a shock wave by the Whitham *equal-area rule*

4.4.2 Definition of integral solution

We have seen that the method of characteristics is not sufficient, in general, to determine the solution of an initial value problem for all times $t > 0$. In the green light problem a rarefaction wave was used to construct the solution in a region not covered by characteristics. In the traffic jam case the solution undergoes a shock, propagating according to the Rankine-Hugoniot condition.

---

[10] The *equal-area* rule holds for a general conservation law (see *Whitham*, 1974).

Some questions arise naturally.

Q1. *In which sense is the differential equation satisfied across a shock or, more generally, across a separation curve where the constructed solution is not differentiable?*

One way to solve the problem is simply to not care about those points. However, in this case it would be possible to construct solutions that do not have anything to do with the physical meaning of the conservation law.

Q2. *Is the solution unique?*

Q3. *If there is no uniqueness, is there a criterion to select the "physically correct" solution?*

To answer, we need first of all to introduce a more flexible notion of solution, in which the derivatives of the solution are not directly involved. Let us go back to the problem

$$\begin{cases} u_t + q\left(u\right)_x = 0 & x \in \mathbb{R},\ t > 0 \\ u\left(x, 0\right) = g\left(x\right) & x \in \mathbb{R} \end{cases} \tag{4.41}$$

and assume for the moment that u is a smooth solution, at least of class C^1 in $\mathbb{R} \times [0, \infty)$. We say that u is a **classical solution**.

Let v be a smooth function in $\mathbb{R} \times [0, \infty)$, with compact support. We call v a *test function*. Multiply the differential equation by v and integrate on $\mathbb{R} \times (0, \infty)$. We get

$$\int_0^\infty \int_{\mathbb{R}} \left[u_t + q\left(u\right)_x\right] v\ dxdt = 0. \tag{4.42}$$

The idea is to carry the derivatives onto the test function v via an integration by parts. If we integrate by parts the first term with respect to t we obtain[11]:

$$\int_0^\infty \int_{\mathbb{R}} u_t v\ dxdt = -\int_0^\infty \int_{\mathbb{R}} u v_t\ dxdt - \int_{\mathbb{R}} u\left(x, 0\right) v\left(x, 0\right) dx$$
$$= -\int_0^\infty \int_{\mathbb{R}} u v_t\ dxdt - \int_{\mathbb{R}} g\left(x\right) v\left(x, 0\right) dx.$$

Integrating by parts the second term in (4.42) with respect to x, we have:

$$\int_0^\infty \int_{\mathbb{R}} q\left(u\right)_x v\ dxdt = -\int_0^\infty \int_{\mathbb{R}} q(u) v_x\ dxdt.$$

Then, equation (4.42) becomes

$$\int_0^\infty \int_{\mathbb{R}} \left[u v_t + q(u) v_x\right] dxdt + \int_{\mathbb{R}} g\left(x\right) v\left(x, 0\right) dx = 0. \tag{4.43}$$

We have obtained an integral equation, valid **for every test function** v. Observe that *no derivative of u* appears in (4.43).

---

[11] Since v is compactly supported and u, v are smooth, there is no problem in exchanging the order of integration.

On the other hand, suppose that a smooth function u satisfies (4.43) *for every test function v*. Integrating by parts in the reverse order, we arrive to the equation

$$\int_0^\infty \int_{\mathbb{R}} [u_t + q(u)_x] v \; dxdt + \int_{\mathbb{R}} [g(x) - u(x,0)] v(x,0) \, dx = 0, \qquad (4.44)$$

still true for every test function v. Choose v vanishing for $t = 0$; then the second integral is zero and the arbitrariness of v implies

$$u_t + q(u)_x = 0 \qquad \text{in } \mathbb{R} \times (0, +\infty). \qquad (4.45)$$

Choosing now v non vanishing for $t = 0$, from (4.44) and (4.45), we get

$$\int_{\mathbb{R}} [g(x) - u(x,0)] v(x,0) \, dx = 0.$$

Once more, the arbitrariness of v implies

$$u(x,0) = g(x) \qquad \text{in } \mathbb{R}.$$

Therefore u is a solution of problem (4.41).

Conclusion: *a function $u \in C^1 (\mathbb{R} \times [0 \times \infty))$ is a solution of problem (4.41) if and only if the equation (4.43) holds for every test function v.*

But (4.43) makes perfect sense for u merely bounded, so that it constitutes an alternative **integral or weak** formulation of problem (4.41). This motivates the following definition.

Definition 4.1. *A function u, bounded in $\mathbb{R} \times [0, \infty)$, is called integral (or weak) solution of problem (4.41) if equation (4.43) holds for every test function v in $\mathbb{R} \times [0, \infty)$, with compact support.*

We point out that an integral solution may be discontinuous, since the definition requires only boundedness.

4.4.3 The Rankine-Hugoniot condition

Definition 4.1 looks rather satisfactory, because of its flexibility. However we have to understand which information about the weak solutions behavior at a singularity, e.g. across a shock curve, is hidden in the integral formulation.

Consider an open set V, contained in the half-plane $t > 0$, partitioned into two disjoint domains V^+ and V^- by a smooth (shock) curve Γ of equation $x = s(t)$ (Fig. 4.17).

Suppose u is a weak solution in V, of class C^1 in both $\overline{V^+}$ and $\overline{V^-}$, separately[12]. We have seen that u is a classical solution of $u_t + q(u)_x = 0$ in V^+ and V^-.

---

[12] That is, u and its the first derivatives extend continuously up to Γ, from both sides, separately.
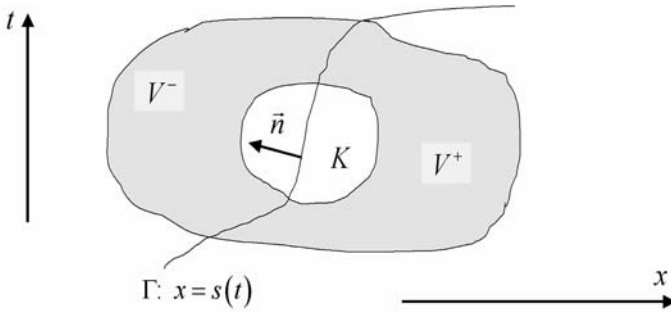
Fig. 4.17. A shock curve dividing a domain V

Choose now a test function v, supported in a compact set $K \subset V$, such that $K \cap \Gamma$ is non empty. Since $v(x, 0) = 0$, we can write:

$$0 = \int_0^\infty \int_{\mathbb{R}} [uv_t + q(u)v_x] \, dxdt$$

$$= \int_{V^+} [uv_t + q(u)v_x] \, dxdt + \int_{V^-} [uv_t + q(u)v_x] \, dxdt.$$

Integrating by parts and observing that $v = 0$ on $\partial V^+ \backslash \Gamma$, we have:

$$\int_{V^+} [uv_t + q(u)v_x] \, dxdt =$$

$$= -\int_{V^+} [u_t + q(u)_x] \, v \, dxdt + \int_{\Gamma} [u_+ \, n_2 + q(u_+)n_1] \, v \, dl$$

$$= \int_{\Gamma} [u_+ \, n_2 + q(u_+)n_1] \, v \, dl$$

where u_+ denotes the value of u on Γ from the V^+ side, $\mathbf{n} = (n_1, n_2)$ is the outward unit normal vector on ∂V^+ and dl denotes the arc length on Γ. Similarly, since $\mathbf{n}$ is inward with respect to V^-:

$$\int_{V^-} [uv_t + q(u)v_x] \, dxdt = -\int_{\Gamma} [u_- \, n_2 + q(u_-)n_1] \, v \, dl$$

where u_- denotes the value of u on Γ from the V^- side. Therefore we deduce that

$$\int_{\Gamma} \{[q(u_+) - q(u_-)] \, n_1 + [u_+ - u_-] \, n_2\} \, v \, dl = 0.$$

The arbitrariness of v yields

$$[q(u_+) - q(u_-)] \, n_1 + [u_+ - u_-] \, n_2 = 0 \tag{4.46}$$

on Γ. If u is continuous across Γ, (4.46) is automatically satisfied. If $u_+ \neq u_-$ we write the relation (4.46) more explicitly. Since $x = s(t)$ on Γ, we have

$$\mathbf{n} = (n_1, n_2) = \frac{1}{\sqrt{1 + (\dot{s}(t))^2}} (-1, \dot{s}(t)).$$

Hence (4.46) becomes, after simple calculations,

$$\dot{s} = \frac{q[u_+(s,t)] - q[u_-(s,t)]}{u_+(s,t) - u_-(s,t)} \tag{4.47}$$

which is the **Rankine-Hugoniot condition** for the shock curve Γ.

In general, functions constructed by connecting classical solutions and rarefaction waves in a continuous way are weak solutions. The same is true for shock waves satisfying the Rankine-Hugoniot condition. Then, the solutions of the green light and of the traffic jam problems are precisely integral solutions.

Thus, Definition 4.1 gives a satisfactory answer to question **Q1**. The other two questions require a deeper analysis as the following example shows.
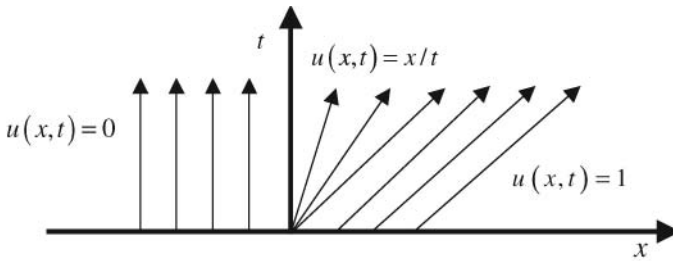
Fig. 4.18. The rarefaction wave of example 4.3

Example 4.2. (Non uniqueness). Imagine a flux of particles along the $x-$axis, each one moving with constant speed. Suppose that $u = u(x,t)$ represents *the velocity field*, which gives the speed of the particle located at x at time t. If $x = x(t)$ is the *path* of a particle, its velocity at time t is given by

$$\dot{x}(t) = u(x(t), t) \equiv \text{constant}.$$

Thus, we have

$$0 = \frac{d}{dt} u(x(t), t) = u_t(x(t), t) + u_x(x(t), t) \dot{x}(t)$$
$$= u_t(x(t), t) + u_x(x(t), t) \ u(x(t), t).$$

Therefore $u = u(x,t)$ satisfies *Burger's equation*

$$u_t + u u_x = u_t + \left(\frac{u^2}{2}\right)_x = 0 \tag{4.48}$$

which is a conservation law with $q(u) = u^2/2$. Note that q is strictly convex: $q'(u) = u$ and $q''(u) = 1$. We couple (4.48) with the initial condition $u(x,0) = g(x)$, where

$$g(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases}$$

The characteristics are the straight lines

$$x = g(x_0)t + x_0. \tag{4.49}$$

Therefore, $u = 0$ if $x < 0$ and $u = 1$ if $x > t$. The region $S = \{0 < x < t\}$ is not covered by characteristics. As in the green light problem, we connect the states 0 and 1 through a *rarefaction wave*. Since $q'(u) = u$, we have $r(s) = (q')^{-1}(s) = s$, so that we construct the weak solution.

$$u(x,t) = \begin{cases} 0 & x \le 0 \\ \dfrac{x}{t} & 0 < x < t \\ 1 & x \ge t. \end{cases} \tag{4.50}$$

However, u **is not the unique weak solution**! There exists also a *shock wave* solution. In fact, since

$$u_- = 0, \; u_+ = 1, \; q(u_-) = 0, \; q(u_+) = \frac{1}{2},$$

the Rankine-Hugoniot condition yields

$$\dot{s}(t) = \frac{q(u_+) - q(u_-)}{u_+ - u_-} = \frac{1}{2}.$$

Given the discontinuity at $x = 0$ of the initial data, the shock curve starts at $s(0) = 0$ and it is the straight line

$$x = \frac{t}{2}.$$

Hence, the function

$$w(x,t) = \begin{cases} 0 & x < \frac{t}{2} \\ 1 & x > \frac{t}{2} \end{cases}$$

is another weak solution (Fig. 4.19). As we shall see, this shock wave has to be considered not physically acceptable.

The example shows that the answer to question **Q2** is *negative* and question **Q3** becomes relevant. We need a criterion to establish which one is the physically correct solution.
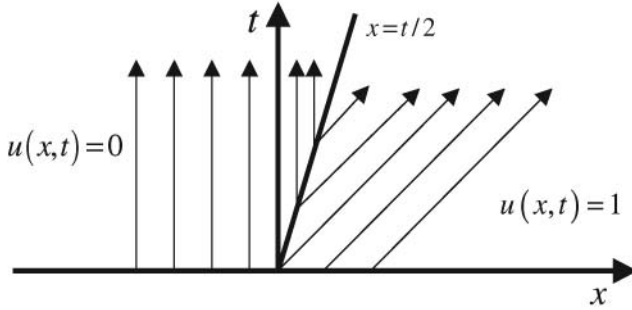
Fig. 4.19. A *non physical* shock

4.4.4 The entropy condition

From Proposition 4.2 we have seen that the equation

$$G(x, t, u) \equiv u - g[x - q'(u)t] = 0$$

defines the unique classical solution u of problem (4.41), at least for small times. The Implicit Function Theorem gives

$$u_x = -\frac{G_x}{G_u} = \frac{g'}{1 + tg'q''}.$$

If we assume $g' > 0$, $q'' \geq C > 0$, we get

$$u_x \leq \frac{E}{t}$$

where[13] $E = \frac{1}{C}$. Using the mean value theorem we deduce the following condition, called **entropy condition**: *there exists $E \geq 0$ such that*[14], *for every $x, z \in \mathbb{R}$, $z > 0$, and every $t > 0$,*

$$u(x + z, t) - u(x, t) \leq \frac{E}{t}z. \tag{4.51}$$

The denomination comes from an analogy with gas dynamics, where a condition like (4.51) implies that entropy increases across a shock. The entropy condition does not involve any derivative of u and makes perfect sense for discontinuous solutions as well. A weak solution satisfying (4.51) is said to be an **entropy solution.** A number of consequences follows directly from (4.51).

---

[13] In the case $g' < 0$ and $q'' \leq C < 0$ we already have $u_x < 0$.
[14]

$$u(x + z, t) - u(x, t) = u_x(x + z^*)z$$

with a suitable z^* between 0 e z.

- The function

$$x \longmapsto u(x,t) - \frac{E}{t}x$$

  is *decreasing*. In fact, let $x + z = x_2$, $x = x_1$ and $z > 0$. Then $x_2 > x_1$ and (4.51) is equivalent to

$$u(x_2,t) - \frac{E}{t}x_2 \leq u(x_1,t) - \frac{E}{t}x_1. \tag{4.52}$$

- If x is a discontinuity point for $u(\cdot,t)$, then

$$u_+(x,t) < u_-(x,t) \tag{4.53}$$

  where $u_\pm(x,t) = \lim_{y \to x^\pm} u(y,t)$. In fact, choose $x_1 < x < x_2$ and let x_1 and x_2 both go to x in (4.52).

If q is **strictly convex**, (4.53) yields

$$q'(u_+) < \frac{q(u_+) - q(u_-)}{u_+ - u_-} < q'(u^-).$$

Then, the Rankine-Hugoniot implies that, if $x = s(t)$ is a shock curve,

$$q'(u_+(x,t)) < \dot{s}(t) < q'(u_-(x,t)) \tag{4.54}$$

that is called **entropy inequality**. The geometrical meaning of (4.54) is remarkable: *the slope of a shock curve is less than the slope of the left-characteristics and greater than the slope of the right-characteristics.* Roughly, the characteristics **hit forward** in time the shock line, so that it is not possible to go back in time along characteristics and hit a shock line, expressing a sort of irreversibility after a shock.

The above considerations lead us to select the entropy solutions as the only physically meaningful ones. On the other hand, if the characteristics hit a shock curve backward in time, the shock wave is to be considered *non-physical*.

Thus, in the non-uniqueness Example 4.3, the solution w represents a non-physical shock since it does not satisfy the entropy condition. The correct solution is therefore the simple wave (4.50). The following important result holds (see e.g. *Smoller*, 1983).

Theorem 4.1. *If $q \in C^2(\mathbb{R})$ is convex (or concave) and g is bounded, there exists a unique entropy solution of the problem*

$$\begin{cases} u_t + q(u)_x = 0 & x \in \mathbb{R}, \ t > 0 \\ u(x,0) = g(x) & x \in \mathbb{R}. \end{cases} \tag{4.55}$$

4.4.5 The Riemann problem

We apply Theorem 4.1 to solve explicitly problem (4.55) with initial data

$$g(x) = \begin{cases} u_+ & x > 0 \\ u_- & x < 0, \end{cases} \tag{4.56}$$

where u_+ and u_- are constants, $u_+ \neq u_-$. This problem is known as **Riemann problem**, and it is particularly important for the numerical approximation of more complex problems.

Theorem 4.2. Let $q \in C^2(\mathbb{R})$ be strictly convex and $q'' \geq h > 0$.
a) If $u_+ < u_-$, the unique entropy solution is the shock wave

$$u(x,t) = \begin{cases} u_+ & \frac{x}{t} > s'(t) \\ u_- & \frac{x}{t} < s'(t) \end{cases} \tag{4.57}$$

where

$$\dot{s}(t) = \frac{q(u_+) - q(u_-)}{u_+ - u_-}.$$

b) If $u_+ > u_-$, the unique entropy solution is the rarefaction wave

$$u(x,t) = \begin{cases} u_- & \frac{x}{t} < q'(u_-) \\ r\left(\frac{x}{t}\right) & q'(u_-) < \frac{x}{t} < q'(u_+) \\ u_+ & \frac{x}{t} > q'(u_+) \end{cases}$$

where $r = (q')^{-1}$, is the inverse function of q'.

Proof. a) The shock wave (4.57) satisfies the Rankine Hugoniot condition and therefore it is clearly a weak solution. Moreover, since $u_+ < u_-$ the entropy condition holds as well, and u is the unique entropy solution of problem (4.56) by Theorem 4.1.
b) Since

$$r(q'(u_+)) = u_- \quad \text{and} \quad r(q'(u_-)) = u_-,$$

u is continuous in the half-plane $t > 0$ and we have only to check that u satisfies the equation $u_t + q(u)_x = 0$ in the region

$$S = \left\{ (x,t) : q'(u_-) < \frac{x}{t} < q'(u_+) \right\}.$$

Let $u(x,t) = r\left(\frac{x}{t}\right)$. We have:

$$u_t + q(u)_x = -r'\left(\frac{x}{t}\right)\frac{x}{t^2} + q'(r)r'\left(\frac{x}{t}\right)\frac{1}{t} = r'\left(\frac{x}{t}\right)\frac{1}{t}\left[q'(r) - \frac{x}{t}\right] \equiv 0.$$

Thus, u is a weak solution in the upper half-plane.

Let us check the entropy condition. We consider only the case

$$q'\left(u_-\right)t \le x < x + z \le q'\left(u_+\right)t$$

leaving the others to the reader. Since $q'' \ge h > 0$, we have

$$r'\left(s\right) = \frac{1}{q''\left(r\right)} \le \frac{1}{h} \qquad (s = q'\left(r\right))$$

so that $(0 < z^* < z)$

$$u\left(x + z\right) - u\left(x, t\right) = r\left(\frac{x + z}{t}\right) - r\left(\frac{x}{t}\right)$$

$$= R'\left(\frac{x + z^*}{t}\right)\frac{z}{t} \le \frac{1}{h}\frac{z}{t}$$

which is the entropy condition with $E = 1/h$. $\square$

4.4.6 Vanishing viscosity method

There is another instructive and perhaps more natural way to construct discontinuous solutions of the conservation law

$$u_t + q\left(u\right)_x = 0, \tag{4.58}$$

the so called *vanishing viscosity method.* This method consists in viewing equation (4.58) as the limit for $\varepsilon \to 0^+$ of the equation

$$u_t + q\left(u\right)_x = \varepsilon u_{xx}, \tag{4.59}$$

that corresponds to choosing the flux function

$$\tilde{q}\left(u\right) = q\left(u\right) - \varepsilon u_x, \tag{4.60}$$

where ε is a *small positive* number. Although we recognize εu_{xx} as a diffusion term, this kind of model arises mostly in fluid dynamics where u is the fluid velocity and ε its *viscosity*, from which comes the name of the method.

There are several good reasons in favor of this approach. First of all, a small amount of diffusion or viscosity makes the mathematical model more realistic in most applications. Note that εu_{xx} becomes relevant only when u_{xx} is large, that is in a region where u_x changes rapidly and a shock occurs. For instance in our model of traffic dynamics, it is natural to assume that drivers would slow down when they see increased (relative) density ahead. Thus, an appropriate model for their velocity is

$$\tilde{v}\left(\rho\right) = v\left(\rho\right) - \varepsilon\frac{\rho_x}{\rho}$$

which corresponds to $\tilde{q}\left(\rho\right) = \rho v\left(\rho\right) - \varepsilon\rho_x$ for the flow-rate of cars.

Another reason comes from the fact that shocks constructed by the vanishing viscosity method are *physical shocks*, since they satisfy the entropy inequality.

As for the heat equation, in principle we expect to obtain smooth solutions even with discontinuous initial data. On the other hand, the nonlinear term may force the evolution towards a shock wave.

Here we are interested in solutions of (4.59) connecting two constant states u_L and u_R, that is, satisfying the conditions

$$\lim_{x \to -\infty} u(x,t) = u_L, \qquad \lim_{x \to +\infty} u(x,t) = u_R. \tag{4.61}$$

Since we are looking for shock waves, it is reasonable to seek a solution depending only on a coordinate $\xi = x - vt$ moving with the (unknown) shock speed v. Thus, let us look for *bounded travelling waves* solution of (4.59) of the form

$$u(x,t) = U(x - vt) \equiv U(\xi)$$

with

$$U(-\infty) = u_L \quad \text{and} \quad U(+\infty) = u_R \tag{4.62}$$

and $u_L \neq u_R$. We have

$$u_t = -v\frac{dU}{d\xi}, \quad u_x = \frac{dU}{d\xi}, \quad u_{xx} = \frac{d^2U}{d\xi^2}$$

so that we obtain for U the ordinary differential equation

$$(q'(U) - v)\frac{dU}{d\xi} = \varepsilon\frac{d^2U}{d\xi^2}$$

which can be integrated to yield

$$q(U) - vU + A = \varepsilon\frac{dU}{d\xi}$$

where A is an arbitrary constant. Assuming that $\dfrac{dU}{d\xi} \to 0$ as $\xi \to \pm\infty$ and using (4.62) we get

$$q(u_L) - vu_L + A = 0 \quad \text{and} \quad q(u_R) - vu_R + A = 0. \tag{4.63}$$

Subtracting these two equations we find

$$v = \frac{q(u_R) - q(u_L)}{u_R - u_L} \equiv \bar{v}. \tag{4.64}$$

and then $A = \dfrac{-q(u_R)u_L + q(u_L)u_R}{u_R - u_L} \equiv \bar{A}$.

Thus, if there exists a travelling wave solution satisfying conditions (4.61), it moves with a speed $\bar{v}$ predicted by the Rankine-Hugoniot formula. Still it is not

clear whether such travelling wave solution exists. To verifies this, examine the equation

$$\varepsilon \frac{dU}{d\xi} = q(U) - \bar{v}U + \bar{A}. \tag{4.65}$$

From (4.63), equation (4.65) has the two equilibria $U = u_R$ and $U = u_L$. A bounded travelling wave connecting u_R and u_L corresponds to a solution of (4.65) starting from a point ξ_0 between u_R and u_L. On the other hand, conditions (4.62) require u_R to be *asymptotically stable* and u_L *unstable*. At this point, we need to have information on the shape of q.

Assume $q'' < 0$. Then the phase diagram for equation (4.65) is described in Fig. 4.20 for the two cases $u_L > u_R$ and $u_L < u_R$.

Between u_L and u_R, $q(U) - \bar{v}U + \bar{A} > 0$ and, as the arrows indicate, U is *increasing*. We see that only the case $u_L < u_R$ is compatible with conditions (4.62) and this corresponds precisely to a shock formation for the non diffusive conservation law. Thus,

$$q'(u_L) - \bar{v} > 0 \quad \text{and} \quad q'(u_R) - \bar{v} < 0$$

or

$$q'(u_R) < \bar{v} < q'(u_L) \tag{4.66}$$

which is *the entropy inequality*.

Similarly, if $q'' > 0$, a travelling wave solution connecting the two states u_R and u_L exists only if $u_L > u_R$ and (4.66) holds.

$$z = q(U) - \bar{v}U + \bar{A}$$
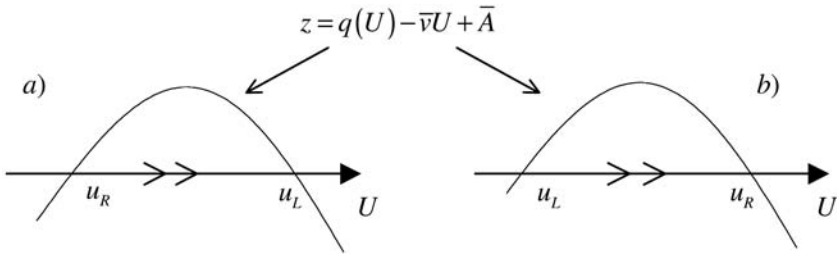
Fig. 4.20. Case b) only is compatible with conditions (4.61)

Let us see what happens when $\varepsilon \to 0$. Assume $q'' < 0$. For ε small, we expect that our travelling wave increases abruptly from a value $U(\xi_1)$ close to u_L to a value $U(\xi_2)$ close to u_R within a narrow region called the *transition layer*. For instance we may choose ξ_1 and ξ_2 such that

$$U(\xi_2) - U(\xi_1) \geq (1 - \beta)(u_R - u_L)$$

with a positive β, very close to 0. We call the number $\varkappa = \xi_2 - \xi_1$ *thickness* of the transition layer. To compute it, we separate the variables U and ξ in (4.65) and

integrate over (ξ_1, ξ_2); this yields

$$\xi_2 - \xi_1 = \varepsilon \int_{U(\xi_1)}^{U(\xi_2)} \frac{ds}{q(s) - vs + \bar{A}}.$$

Thus, the thickness of the transition layer is proportional to ε. As $\varepsilon \to 0$, the transition region becomes more and more narrow and eventually a shock wave that satisfies the entropy inequality is obtained.

This phenomenon is clearly seen in the important case of *viscous Burger's* equation that we examine in more details in the next subsection.

Example 4.3. Burger's shock solution. Let us determine a travelling wave solution of the viscous Burger equation

$$u_t + uu_x = \varepsilon u_{xx} \tag{4.67}$$

connecting the states $u_L = 1$ and $u_R = 0$. Note that $q(u) = u^2/2$ is convex. Then $\bar{v} = 1/2$ and $\bar{A} = 0$. Equation (4.65) becomes

$$2\varepsilon \frac{dU}{d\xi} = U^2 - U$$

that can be easily integrated to give

$$U(\xi) = \frac{1}{1 + \exp\left(\dfrac{\xi}{2\varepsilon}\right)}.$$

Thus the travelling wave is given by

$$u(x, t) = U\left(x - \frac{t}{2}\right) = \frac{1}{1 + \exp\left(\dfrac{2x - t}{4\varepsilon}\right)}. \tag{4.68}$$

When $\varepsilon \to 0$,

$$u(x, t) \to w(x, t) = \begin{cases} 0 & x > t/2 \\ 1 & x < t/2 \end{cases}$$

which is the entropy shock solution for the non viscous Burger equation with initial data 1 if $x < 0$ and 0 if $x > 0$.

4.4.7 The viscous Burger equation

The viscous Burger equation is one of the most celebrated examples of nonlinear diffusion equation. It arose (Burger, 1948) as a simplified form of the Navier-Stokes equation, in an attempt to study some aspects of turbulence. It appears also in gas dynamics, in the theory of sound waves and in traffic flow modelling and
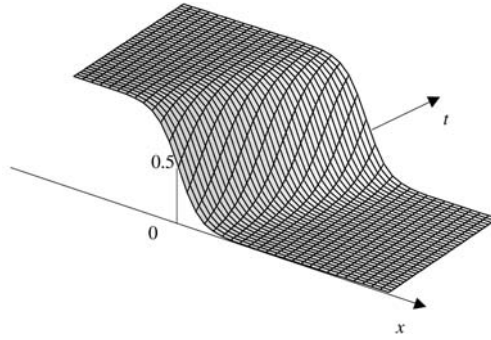
Fig. 4.21. The travelling wave in Example 4.3

it constitutes a basic example of competition between *dissipation* (due to linear diffusion) and *steepening* (shock formation due to the nonlinear transport term uu_x).

The success of Burger's equation is in large part due to the rather surprising fact that the initial value problem can be solved analytically. In fact, via the so called *Hopf-Cole transformation*, Burger's equation is converted into the heat equation. Let us see how this can be done. Write the equation in the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2 - \varepsilon u_x\right) = 0.$$

Then, the planar vector field $\left(-u, \frac{1}{2}u^2 - \varepsilon u_x\right)$ is curl-free and therefore there exists a potential $\psi = \psi(x, t)$ such that

$$\psi_x = -u \quad \text{and} \quad \psi_t = \frac{1}{2}u^2 - \varepsilon u_x.$$

Thus, ψ solves the equation

$$\psi_t = \frac{1}{2}\psi_x^2 + \varepsilon\psi_{xx}. \tag{4.69}$$

Now we try to get rid of the quadratic term letting $\psi = g(\varphi)$, with g to be chosen. We have

$$\psi_t = g'(\varphi)\varphi_t, \quad \psi_x = g'(\varphi)\varphi_x, \quad \psi_{xx} = g''(\varphi)(\varphi_x)^2 + g'(\varphi)\varphi_{xx}.$$

Substituting into (4.69) we find

$$g'(\varphi)[\varphi_t - \varepsilon\varphi_{xx}] = [\frac{1}{2}(g'(\varphi))^2 + \varepsilon g''(\varphi)](\varphi_x)^2.$$

Hence, if we choose $g(s) = 2\varepsilon\log s$, then the right hand side vanishes and we are left with

$$\varphi_t - \varepsilon\varphi_{xx} = 0. \tag{4.70}$$

Thus
$$\psi = 2\varepsilon \log \varphi$$

and from $u = -\psi_x$ we obtain
$$u = -2\varepsilon \frac{\varphi_x}{\varphi} \tag{4.71}$$

which is the *Hopf-Cole transformation*. An initial data
$$u(x, 0) = u_0(x) \tag{4.72}$$

transforms into an initial data of the form[15]
$$\varphi_0(x) = \exp\left\{-\int_a^x \frac{u_0(z)}{2\varepsilon} dz\right\} \qquad (a \in \mathbb{R}). \tag{4.73}$$

If (see Theorem 2.4)
$$\frac{1}{x^2} \int_a^x u_0(z)\, dz \to 0 \quad \text{as } |x| \to \infty,$$

the initial value problem (4.70), (4.73) has a unique smooth solution in the half-plane $t > 0$, given by formula (2.137):
$$\varphi(x, t) = \frac{1}{\sqrt{4\pi\varepsilon t}} \int_{-\infty}^{+\infty} \varphi_0(y) \exp\left(-\frac{(x - y)^2}{4\varepsilon t}\right) dy.$$

This solution is continuous with its $x-$derivative[16] up to $t = 0$ at any continuity point of u_0. Consequently, from (4.71), problem (4.67) has a unique smooth solution in the half-plane $t > 0$, continuous up to $t = 0$ at any continuity point of u_0, given by
$$u(x, t) = \frac{\displaystyle\int_{-\infty}^{+\infty} \frac{x - y}{t} \varphi_0(y) \exp\left(-\frac{(x - y)^2}{4\varepsilon t}\right) dy}{\displaystyle\int_{-\infty}^{+\infty} \varphi_0(y) \exp\left(-\frac{(x - y)^2}{4\varepsilon t}\right) dy}. \tag{4.74}$$

We use formula (4.74) to solve an initial pulse problem.

Example 4.4. Initial pulse. Consider problem (4.67), (4.67) with the initial condition
$$u_0(x) = M\delta(x)$$

where δ denotes the Dirac density at the origin. We have, choosing $a = 1$,
$$\varphi_0(x) = \exp\left\{-\int_1^x \frac{u_0(y)}{2\varepsilon} dy\right\} = \begin{cases} 1 & x > 0 \\ \exp\left(\frac{M}{2\varepsilon}\right) & x < 0. \end{cases}$$

---

[15] The choice of a is arbitrary and does not affect the value of u.
[16] Check it.

Formula (4.74), gives, after some routine calculations,

$$u\left(x,t\right)=\sqrt{\frac{4\varepsilon}{\pi t}}\frac{\exp\left(-\dfrac{x^2}{4\varepsilon t}\right)}{\dfrac{2}{\exp\left(M/2\varepsilon\right)-1}+\dfrac{\sqrt{\pi}}{2}\left[1-\operatorname{erf}\left(\dfrac{x}{\sqrt{4\varepsilon t}}\right)\right]}$$

where

$$\operatorname{erf}(x)=\int_0^x e^{-z^2}dz$$

is the *error function*.
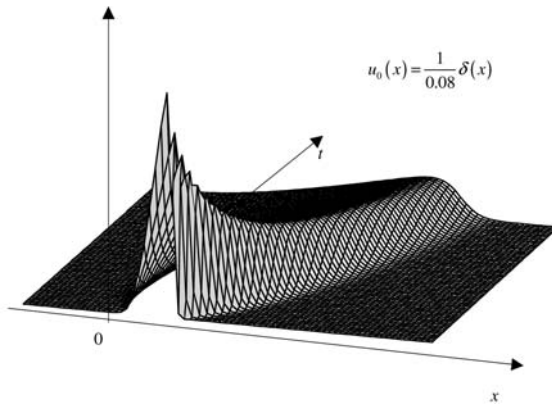
Fig. 4.22. Evolution of an initial pulse for the viscous Burger's equation ($M=1, \varepsilon = 0.04$)

4.5 The Method of Characteristics for Quasilinear Equations

In this section we apply the characteristics method to general quasilinear equations. We consider mainly the case of two independent variables, where the intuition is supported by the geometric interpretation. However, the generalization to any number of dimensions should not be too difficult for the reader.

4.5.1 Characteristics

We consider equations of the form

$$a\left(x,y,u\right)u_x+b\left(x,y,u\right)u_y=c\left(x,y,u\right) \tag{4.75}$$

where

$$u = u(x, y)$$

and a, b, c are *continuously differentiable functions*.

The solutions of (4.75) can be constructed via geometric arguments. The tangent plane to the graph of a solution u at a point (x_0, y_0, z_0) has equation

$$u_x(x_0, y_0)(x - x_0) + u_y(x_0, y_0)(y - y_0) - (z - z_0) = 0$$

and the vector

$$\mathbf{n}_0 = (u_x(x_0, y_0), u_y(x_0, y_0), -1)$$

is *normal* to the plane. Introducing the vector

$$\mathbf{v}_0 = (a(x_0, y_0, z_0), b(x_0, y_0, z_0), c(x_0, y_0, z_0)),$$

equation (4.75) implies that

$$\mathbf{n}_0 \cdot \mathbf{v}_0 = 0.$$

Thus, $\mathbf{v}_0$ is tangent to the graph of u (Fig. 4.23). In other words, (4.75) says that, at every point (x, y, z), the graph of any solution is tangent to the vector field

$$\mathbf{v}(x, y, z) = (a(x, y, z), b(x, y, z), c(x, y, z)).$$

In this case we say that the graph of a solution is an **integral surface** of the vector field $\mathbf{v}$.
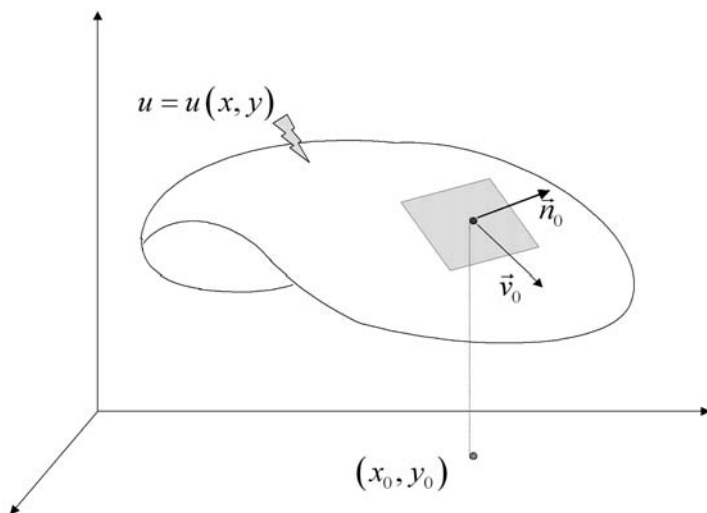
Fig. 4.23. Integral surface

Now, we may construct integral surfaces of $\mathbf{v}$ as union of **integral curves of $\mathbf{v}$**, that is curves tangent to $\mathbf{v}$ at every point. These curves are solutions of the

system

$$\frac{dx}{dt} = a\,(x, y, z)\,, \qquad \frac{dy}{dt} = b\,(x, y, z)\,, \qquad \frac{dz}{dt} = c\,(x, y, z) \qquad (4.76)$$

and are called **characteristics**. Note that

$$z = z\,(t)$$

gives the values u along a characteristic, that is

$$z\,(t) = u\,(x\,(t)\,, y\,(t))\,. \qquad (4.77)$$

In fact, differentiating (4.77) and using (4.76) and (4.75), we have

$$\begin{aligned}
\frac{dz}{dt} &= u_x\,(x\,(t)\,, y\,(t))\,\frac{dx}{dt} + u_y\,(x\,(t)\,, y\,(t))\,\frac{dy}{dt} \\
&= a\,(x\,(t)\,, y\,(t)\,, z\,(t))\,u_x\,(x\,(t)\,, y\,(t)) + b((x\,(t)\,, y\,(t)\,, z\,(t))\,u_y\,(x\,(t)\,, y\,(t)) \\
&= c\,(x\,(t)\,, y\,(t)\,, z\,(t))\,.
\end{aligned}$$

Thus, along a characteristic the partial differential equation (4.75) degenerates into an ordinary differential equation.

In the case of a conservation law (with $t = y$)

$$u_y + q'\,(u)\,u_x = 0 \qquad\qquad \left(q'\,(u) = \frac{dq}{du}\right),$$

we have introduced the notion of characteristic in a slightly different way, but we shall see later that there is no contradiction.

The following proposition is a consequence of the above geometric reasoning and of the existence and uniqueness theorem for system of ordinary differential equations[17].

Proposition 4.3. *a) Let the surface S be the graph of a C^1 function $u = u\,(x, y)$. If S is union of characteristics then u is a solution of the equation (4.75).*

b) Every integral surface S of the vector field $\mathbf{v}$ is union of characteristics. Namely: every point of S belongs exactly to one characteristic, entirely contained in S.

c) Two integral surfaces intersecting at one point intersect along the whole characteristic passing through that point.

4.5.2 The Cauchy problem

Proposition 4.3 gives a characterization of the integral surfaces as a union of characteristics. The problem is how to construct such unions to get a smooth surface. One way to proceed is to look for solutions u *whose values are prescribed on a curve γ_0, contained in the x, y plane.*

---

[17] We leave the details of the proof to the reader.

In other words, suppose that

$$x(s) = f(s), \qquad y(s) = g(s) \qquad s \in I \subseteq \mathbb{R}$$

is a parametrization of γ_0. We look for a solution u of (4.75) such that

$$u(f(s), g(s)) = h(s), \qquad s \in I, \tag{4.78}$$

where $h = h(s)$ is a given function. We assume that I is a neighborhood of $s = 0$, and that f, g, h are *continuously differentiable in* I.

The system (4.75), (4.78) is called **Cauchy problem**. If we consider the three-dimensional curve Γ_0 given by the parametrization

$$x(s) = f(s), \qquad y(s) = g(s), \qquad z(s) = h(s),$$

then, solving the Cauchy problem (4.75), (4.78) amounts to *determining an integral surface containing* Γ_0 (Fig. 4.24).

The data are often assigned in the form of *initial values*

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

with y playing the role of "time". In this case, γ_0 is the axis $y = 0$ and x plays the role of the parameter s. Then a parametrization of Γ_0 is given by

$$x = x, \qquad y = 0, \qquad z(x) = h(x).$$

By analogy, we often refer to Γ_0 as to the *initial curve*. The strategy to solve a Cauchy problem comes from its geometric meaning: since the graph of a solution
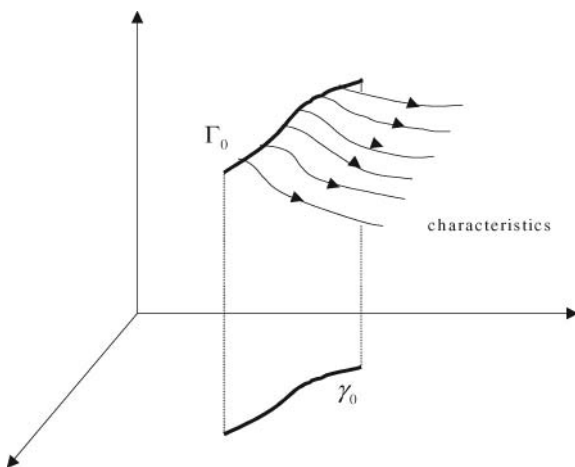
Fig. 4.24. Characteristics flowing out of Γ_0

$u = u(x, y)$ is a union of characteristics, we determine those flowing out from Γ_0, by solving the system

$$\frac{dx}{dt} = a(x, y, z), \qquad \frac{dx}{dt} = b(x, y, z), \qquad \frac{dx}{dt} = c(x, y, z), \qquad (4.79)$$

with the family of initial conditions

$$x(0) = f(s), \quad y(0) = g(s), \quad z(0) = h(s), \qquad (4.80)$$

parametrized by $s \in I$. The union of the characteristics of this family should give[18] u. Why only *should*? We will come back to this question.

Under our hypotheses, the Cauchy problem (4.79), (4.80) has exactly one solution

$$x = X(s, t), \quad y = Y(s, t), \quad z = Z(s, t). \qquad (4.81)$$

in a neighborhood of $t = 0$, for every $s \in I$.

A couple of questions arise:

a) Do the three equations (4.81) define a function $z = u(x, y)$?

b) Even if the answer to question **a)** is positive, is $z = u(x, y)$ the unique solution of the Cauchy problem?

Let us reason in a neighborhood of $s = t = 0$, setting

$$X(0, 0) = f(0) = x_0, \quad Y(0, 0) = g(0) = y_0, \quad Z(0, 0) = h(0) = z_0.$$

The answer to question **a)** is positive if we can solve for s and t the first two equations in (4.81), and find $s = S(x, y)$ e $t = T(x, y)$ of class C^1 in a neighborhood of (x_0, y_0), such that

$$S(x_0, y_0) = 0, \qquad T(x_0, y_0) = 0.$$

Then, from the third equation $z = Z(s, t)$, we get

$$z = Z(S(x, y), T(x, y)) = u(x, y). \qquad (4.82)$$

From the *Inverse Function Theorem*, the system

$$\begin{cases} X(s, t) = x \\ Y(s, t) = y \end{cases}$$

defines

$$s = S(x, y) \quad \text{and} \quad t = T(x, y)$$

in a neighborhood of (x_0, y_0) if

$$J(0, 0) = \begin{vmatrix} X_s(0, 0) & Y_s(0, 0) \\ X_t(0, 0) & Y_t(0, 0) \end{vmatrix} \neq 0. \qquad (4.83)$$

---

[18] Identifying u with its graph.

From (4.79) and (4.80) we have

$$X_s(0,0) = f'(0), \quad Y_s(0,0) = g'(0)$$

and

$$X_t(0,0) = a(x_0, y_0, z_0), \quad Y_t(0,0) = b(x_0, y_0, z_0),$$

so that (4.83) becomes

$$J(0,0) = \begin{vmatrix} f'(0) & g'(0) \\ a(x_0, y_0, z_0) & b(x_0, y_0, z_0) \end{vmatrix} \neq 0 \tag{4.84}$$

or

$$b(x_0, y_0, z_0) f'(0) \neq a(x_0, y_0, z_0) g'(0). \tag{4.85}$$

Condition (4.85) means that *the vectors*

$$(a(x_0, y_0, z_0), b(x_0, y_0, z_0)) \quad \text{and} \quad (f'(0), g'(0))$$

are not parallel.

In conclusion: *if condition (4.84) holds, then (4.82) is a well defined C^1−function.*

Now consider question **b)**. The above construction of u implies that the surface $z = u(x,y)$ contains Γ_0 and all the characteristics flowing out from Γ_0, so that u is a solution of the Cauchy problem. Moreover, by Proposition 4.5 *c)*, two integral surfaces containing Γ_0 must contain the same characteristics and therefore coincide.

We summarize everything in the following theorem, recalling that

$$(x_0, y_0, z_0) = (f(0), g(0), h(0)).$$

Theorem 4.3. *Let a, b, c be C^1−functions in a neighborhood of (x_0, y_0, z_0) and f, g, h be C^1−functions in I. If $J(0,0) \neq 0$, then, in a neighborhood of (x_0, y_0), there exists a unique C^1−solution $u = u(x,y)$ of the Cauchy problem*

$$\begin{cases} a(x, y, u) u_x + b(x, y, u) u_y = c(x, y, u) \\ u(f(s), g(s)) = h(s). \end{cases} \tag{4.86}$$

Moreover, u is defined by the parametric equations (4.81).

Remark 4.2. If a, b, c and f, g, h are C^k−functions, $k \geq 2$, then u is a C^k−function as well.

It remains to examine what happens when $J(0,0) = 0$, that is when the vectors $(a(x_0, y_0, z_0), b(x_0, y_0, z_0))$ and $(f'(0), g'(0))$ are parallel.

Suppose that there exists a C^1−solution u of the Cauchy problem (4.86). Differentiating the second equation in (4.86) we get

$$h'(s) = u_x(f(s), g(s)) f'(s) + u_y(f(s), g(s)) g'(s). \tag{4.87}$$

Computing at $x = x_0$, $y = y_0$, $z = z_0$ and $s = 0$, we obtain

$$\begin{cases} a\left(x_0, y_0, z_0\right) u_x\left(x_0, y_0\right) + b\left(x_0, y_0, z_0\right) u_y\left(x_0, y_0\right) = c\left(x_0, y_0, z_0\right) \\ f'\left(0\right) u_x\left(x_0, y_0\right) + g'\left(0\right) u_y\left(x_0, y_0\right) = h'\left(0\right). \end{cases} \tag{4.88}$$

Since u is a solution of the Cauchy problem, the vector $\left(u_x\left(x_0, y_0\right), u_y\left(x_0, y_0\right)\right)$ is a solution of the algebraic system (4.88). But then, from Linear Algebra, we know that the condition

$$\operatorname{rank} \begin{pmatrix} a\left(x_0, y_0, z_0\right) & b\left(x_0, y_0, z_0\right) & c\left(x_0, y_0, z_0\right) \\ f'\left(0\right) & g'\left(0\right) & h'\left(0\right) \end{pmatrix} = 1 \tag{4.89}$$

must hold and therefore the two vectors

$$\left(a\left(x_0, y_0, z_0\right) b\left(x_0, y_0, z_0\right), c\left(x_0, y_0, z_0\right)\right) \quad \text{and} \quad \left(f'\left(0\right), g'\left(0\right), h'\left(0\right)\right) \tag{4.90}$$

are parallel. This is equivalent to saying that Γ_0 is parallel to the characteristic curve at (x_0, y_0, z_0). When this occurs, we say that Γ_0 **is characteristic at the point** (x_0, y_0, z_0).

Conclusion: *If $J\left(0, 0\right) = 0$, a necessary condition for the existence of a C^1-solution $u = u\left(x, y\right)$ of the Cauchy problem in a neighborhood of (x_0, y_0) is that Γ_0 be characteristic at (x_0, y_0, z_0).*

Now, assume Γ_0 itself is a characteristic and let $P_0 = (x_0, y_0, z_0) \in \Gamma_0$. If we choose a curve Γ^* transversal[19] to Γ_0 at P_0, by Theorem 4.3 there exists a unique integral surface containing Γ^* and, by Proposition 4.3 **c)**, this surface contains Γ_0. In this way we can construct infinitely many smooth solutions.

We point out that the condition (4.89) is compatible with the existence of a C^1-solution only if Γ_0 is characteristic at P_0. On the other hand, it may occur that $J\left(0, 0\right) = 0$, that Γ_0 is non characteristic at P_0 and that solutions of the Cauchy problem exist anyway; clearly, these solutions **cannot** be of class C^1.

Let us summarize the steps to solve the Cauchy problem (4.86):

Step **1.** Determine the solution

$$x = X\left(s, t\right), \quad y = Y\left(s, t\right), \quad z = Z\left(s, t\right) \tag{4.91}$$

of the characteristic system

$$\frac{dx}{dt} = a\left(x, y, z\right), \qquad \frac{dy}{dt} = b\left(x, y, z\right), \qquad \frac{dz}{dt} = c\left(x, y, z\right) \tag{4.92}$$

with initial conditions

$$X\left(s, 0\right) = f\left(s\right), \quad Y\left(s, 0\right) = g\left(s\right), \quad Z\left(s, 0\right) = h\left(s\right), \qquad s \in I.$$

---

[19] Not tangent.

Step **2**. Compute $J(s, t)$ on the initial curve Γ_0 i.e.

$$J(s, 0) = \begin{vmatrix} f'(s) & g'(s) \\ X_t(s, 0) & Y_t(s, 0) \end{vmatrix}.$$

The following cases may occur:

2a. $J(s, 0) \neq 0$, for every $s \in I$. This means that Γ_0 does not have characteristic points. Then, in a neighborhood of Γ_0, there exists a unique solution $u = u(x, y)$ of the Cauchy problem, defined by the parametric equations (4.91).

2b. $J(s_0, 0) = 0$ for some $s_0 \in I$ and Γ_0 is characteristic at the point $P_0 = (f(s_0), g(s_0), h(s_0))$. A C^1−solution may exist in a neighborhood of P_0 only if the rank condition (4.89) holds at P_0.

2c. $J(s_0, 0) = 0$ for some $s_0 \in I$ and Γ_0 is **not** characteristic at P_0. There are no C^1−solutions in a neighborhood of P_0. There may exist less regular solutions.

2d. Γ_0 is a characteristic. Then there exist infinitely many C^1−solutions in a neighborhood of Γ_0.

Example 4.5. Consider the non-homogeneous Burger equation

$$uu_x + u_y = 1. \tag{4.93}$$

As in Example 4.2, if y is the time variable y, $u = u(x, y)$ represents *a velocity field* of a flux of particles along the x−axis. Equation (4.93) states that the acceleration of each particle is equal to 1. Assume

$$u(x, 0) = h(x), \qquad x \in \mathbb{R}.$$

The characteristics are solutions of the system

$$\frac{dx}{dt} = z, \quad \frac{dy}{dt} = 1, \quad \frac{dz}{dt} = 1$$

and the initial curve Γ_0 has the parametrization

$$x = f(s) = s, \quad y = g(s) = 0, \quad z = h(s) \qquad s \in \mathbb{R}.$$

The characteristics flowing out from Γ_0 are

$$X(s, t) = s + \frac{t^2}{2} + th(s), \qquad Y(s, t) = t, \qquad Z(s, t) = t + h(s).$$

Since

$$J(s, t) = \begin{vmatrix} 1 + th'(s) & 0 \\ t + h(s) & 1 \end{vmatrix} = 1 + th'(s),$$

we have $J(s,0) = 1$ and we are in the case **2a**: in a neighborhood of Γ_0 there exists a unique C^1-solution. If, for instance, $h(s) = s$, we find the solution

$$u = y + \frac{2x - y^2}{2 + 2y} \qquad (x \in \mathbb{R},\, y \geq -1).$$

Now consider the same equation with initial condition

$$u\left(\frac{y^2}{4}, y\right) = \frac{y}{2},$$

equivalent to assigning the values of u on the parabola $x = \frac{y^2}{4}$. A parametrization of Γ_0 is given by

$$x = s^2, \quad y = 2s, \quad z = s, \qquad s \in \mathbb{R}.$$

Solving the characteristic system with these initial conditions, we find

$$X(s,t) = s^2 + ts + \frac{t^2}{2}, \qquad Y(s,t) = 2s + t, \qquad Z(s,t) = s + t. \qquad (4.94)$$

Observe that Γ_0 does not have any characteristic point, since its tangent vector $(2s, 2, 1)$ is never parallel to the characteristic direction $(s, 1, 1)$. However

$$J(s,t) = \begin{vmatrix} 2s + t & 2 \\ s + t & 1 \end{vmatrix} = -t$$

which vanishes for $t = 0$, i.e. exactly on Γ_0. We are in the case **2c**. Solving for s and t, $t \neq 0$, in the first two equations (4.94), and substituting into the third one, we find

$$u(x, y) = \frac{y}{2} \pm \sqrt{x - \frac{y^2}{4}}.$$

We have found two solutions of the Cauchy problem, satisfying the differential equation in the region $x > \frac{y^2}{4}$. However, these solutions are not smooth in a neighborhood of Γ_0, since on Γ_0 they are not differentiable.

• *Conservation laws.* According to the new definition, the characteristics of the equation

$$u_y + q'(u)\, u_x = 0 \qquad\qquad \left(q'(u) = \frac{dq}{du}\right),$$

with initial conditions

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

are the three-dimensional solution curves of the system

$$\frac{dx}{dt} = q'(z), \qquad \frac{dy}{dt} = 1, \qquad \frac{dz}{dt} = 0$$

with initial conditions

$$x(s, 0) = s, \quad y(s, 0) = 0, \quad z(s, 0) = g(s), \qquad s \in \mathbb{R}.$$

Integrating, we find

$$z = g(s), \qquad x = q'(g(s))t + s, \qquad y = t.$$

The *projections* of these straight-lines on the x, y plane are

$$x = q'(g(s))y + s.$$

which are the "old characteristics", called *projected characteristics* in the general quasilinear context.

- *Linear equations.* Consider a linear equation

$$a(x, y)u_x + b(x, y)u_y = 0. \tag{4.95}$$

Introducing the vector $\mathbf{w} = (a, b)$, we may write the equation (4.95) in the form

$$D_{\mathbf{w}}u = \nabla u \cdot \mathbf{w} = 0.$$

Thus, every solution of the (4.95) is constant along the integral lines of the vector $\mathbf{w}$, i.e. along the *projected characteristics*, solutions of the reduced characteristic system

$$\frac{dx}{dt} = a(x, y), \qquad \frac{dy}{dt} = b(x, y), \tag{4.96}$$

locally equivalent to the ordinary differential equation

$$b(x, y)dx - a(x, y)dy = 0.$$

If a *first integral*[20] $\psi = \psi(x, y)$ of the system (4.96) is known, then the family of the projected characteristics is given in implicit form by

$$\psi(x, y) = k, \qquad k \in \mathbb{R}$$

and the general solution of (4.95) is given by the formula

$$u(x, y) = G(\psi(x, y)),$$

where $G = G(r)$ is an arbitrary C^1−function, that can be determined by the Cauchy data.

Example 4.6. Let us solve the problem

$$\begin{cases} yu_x + xu_y = 0 \\ u(x, 0) = x^4. \end{cases}$$

---

[20] We recall that a *first integral* (also called *constant of motion*) for a system of o.d.e. $\dfrac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x})$, is a C^1−function $\varphi = \varphi(\mathbf{x})$ which is constant along the trajectories of the system, i.e. such that $\nabla\varphi \cdot \mathbf{f} \equiv 0$.

Here $\mathbf{w} = (y, x)$ and the reduced characteristic system is

$$\frac{dx}{dt} = y, \qquad \frac{dy}{dt} = x$$

locally equivalent to

$$x\,dx - y\,dy = 0.$$

Integrating we find that the projected characteristics are the hyperbolas

$$\psi\,(x, y) = x^2 - y^2 = k$$

and therefore $\psi\,(x, y) = x^2 - y^2$ is a first integral. Then, the general solution of the equation is

$$u\,(x, y) = G(x^2 - y^2).$$

Imposing the Cauchy condition, we have

$$G(x^2) = x^4$$

from which $G\,(r) = r^2$. The solution of the Cauchy problem is

$$u\,(x, y) = (x^2 - y^2)^2.$$

Example 4.7. An interesting situation occurs when we want to find a solution of the equation (4.95) in a smooth domain $\Omega \subset \mathbb{R}^2$, that assumes prescribed values on a subset of the boundary $\gamma = \partial\Omega$. Figure 4.25 shows a possible situation. In analogy with the problems in subsection 4.2.4, for the solvability of the problem we have to assign the Cauchy data only on the so called **inflow boundary** γ_i defined by

$$\gamma_i = \{\sigma \in \gamma : \mathbf{w} \cdot \boldsymbol{\nu} < 0\}$$

where $\boldsymbol{\nu}$ is the outward unit normal to γ. If a smooth Cauchy data is prescribed on γ_i, a smooth solution is obtained by defining u to be constant along the projected characteristics like l_1 and piecewise constant on those like l_2. Observe that the points at which $\mathbf{w}$ is tangent to γ are characteristic.

4.5.3 Lagrange method of first integrals

We have seen that, in the linear case, we can construct a *general solution*, depending on an arbitrary function from the knowledge of a first integral for the reduced characteristic system. The method can be extended to equations of the form

$$a\,(x, y, u)\,u_x + b\,(x, y, u)\,u_y = c\,(x, y, u). \tag{4.97}$$

We say that two *first integrals* $\varphi = \varphi\,(x, y, u)$ are *independent* if $\nabla\varphi$ and $\nabla\psi$ are nowhere colinear. Then:
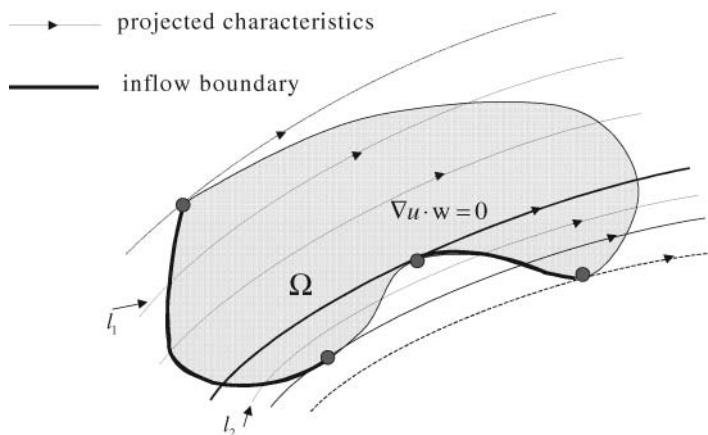
Fig. 4.25. Cauchy problem in a domain

Theorem 4.4. *Let $\varphi = \varphi(x, y, u)$, $\psi = \psi(x, y, u)$ be two independent first integrals of the characteristic system and $F = F(h, k)$ be a C^1-function. If*

$$F_h \varphi_u + F_k \psi_u \neq 0,$$

the equation

$$F(\varphi(x, y, u), \psi(x, y, u)) = 0$$

defines the general solution of (4.97) in implicit form.

Proof. It is based on the following two observations. First, the function

$$w = F(\varphi(x, y, u), \psi(x, y, u)) \tag{4.98}$$

is a first integral. In fact,

$$\nabla w = F_h \nabla \varphi + F_k \nabla \psi$$

so that

$$\nabla w \cdot (a, b, c) = F_h \nabla \varphi \cdot (a, b, c) + F_k \nabla \psi \cdot (a, b, c) \equiv 0$$

since φ and ψ are *first integrals*. Moreover, by hypothesis,

$$w_u = F_h \varphi_u + F_k \psi_u \neq 0.$$

Second, if w is a first integral and $w_u \neq 0$, then equation

$$w(x, y, u) = 0 \tag{4.99}$$

defines implicitly an integral surface $u = u(x, y)$ of (4.97). In fact, since w is a first integral it satisfies the equation

$$a(x, y, u) w_x + b(x, y, u) w_y + c(x, y, u) w_u = 0. \tag{4.100}$$

Moreover, from the Implicit Function Theorem, we have

$$u_x = -\frac{w_x}{w_u}, \qquad u_y = -\frac{w_y}{w_u}$$

and from (4.100) we get easily (4.97). □

Remark 4.3. As a by-product of the proof, we have that the general solution of the three-dimensional homogeneous equation (4.100) is given by (4.98).

Remark 4.4. The search for first integrals is sometimes simplified by writing the characteristic system in the form

$$\frac{dx}{a\,(x,y,u)} = \frac{dy}{b\,(x,y,u)} = \frac{du}{c\,(x,y,u)}.$$

Example 4.8. Consider again the nonhomogeneous Burger equation

$$uu_x + u_y = 1$$

with initial condition

$$u\left(\frac{1}{2}y^2, y\right) = y.$$

A parametrization of the initial curve Γ_0 is

$$x = \frac{1}{2}s^2, \quad y = s, \quad z = s$$

and therefore Γ_0 is a characteristic. We are in the case **2d**.

Let us use the Lagrange method. To find two independent first integrals, we write the characteristic system in the form

$$\frac{dx}{z} = dy = dz$$

or

$$dx = z\,dz, \qquad dy = dz.$$

Integrating the two equations, we get

$$x - \frac{1}{2}z^2 = c_2, \qquad y - z = c_1$$

so that

$$\varphi\,(x,y,z) = x - \frac{1}{2}z^2, \qquad \psi\,(x,y,z) = y - z$$

are two first integral. Since

$$\nabla\varphi\,(x,y,z) = (1,0,-z)$$

and
$$\nabla\psi\left(x,y,z\right) = (0,1,-1)$$

we see that they are also independent.

Thus, the general solution of Burger equation is given by

$$F\left(x - \frac{1}{2}z^2, y - z\right) = 0$$

where F is an arbitrary C^1-function.

Finally, to satisfy the Cauchy condition, it is enough to choose F such that $F(0,0) = 0$. As expected, there exist infinitely many solutions of the Cauchy problem.

4.5.4 Underground flow

Consider the underground flow of a fluid (like water). In the wet region, only a fraction of any control volume is filled with fluid. This fraction, denoted by ϕ, is called *porosity* and, in general, it depends on position, temperature and pressure. Here, we assume ϕ depends on position only: $\phi = \phi\left(x,y,z\right)$.

If ρ is the fluid density and $\mathbf{q} = (q_1, q_2, q_3)$ is the flux vector (the volumetric flow rate of the fluid), the conservation of mass yields, in this case,

$$\phi\rho_t + \mathrm{div}(\rho\mathbf{q}) = 0.$$

For $\mathbf{q}$ the following modified *Darcy law* is commonly adopted:

$$\mathbf{q} = -\frac{k}{\mu}\left(\nabla p + \rho\mathbf{g}\right)$$

where p is the pressure and $\mathbf{g}$ is the gravity acceleration; $k > 0$ is the *permeability* of the medium and μ is the fluid *viscosity*[21]. Thus, we have:

$$\phi\rho_t - \mathrm{div}\left[\frac{\rho k}{\mu}\left(\nabla p + \rho\mathbf{g}\right)\right] = 0. \tag{4.101}$$

Now, suppose that two *immiscible* fluids, of density ρ_1 and ρ_2, flow underground. Immiscible means that the two fluids cannot dissolve one into the other or chemically interact. In particular, the conservation law holds for the mass of each fluid. Thus, if we denote by S_1 and S_2 the fractions (*saturations*) of the available space filled by the two fluids, respectively, we can write

$$\phi\left(S_1\rho_1\right)_t + \mathrm{div}(\rho_1\mathbf{q_1}) = 0 \tag{4.102}$$

$$\phi\left(S_2\rho_2\right)_t + \mathrm{div}(\rho_2\mathbf{q_2}) = 0. \tag{4.103}$$

---

[21] For water: $\mu = 1.14 \cdot 10^3 \; Kg \times m \times s^{-1}$.

We assume that $S_1 + S_2 = 1$, i.e. that the medium is completely saturated and that capillarity effects are negligible. We set $S_1 = S$ and $S_2 = 1 - S$. The Darcy law for the two fluids becomes

$$\mathbf{q}_1 = - k \frac{k_1}{\mu_1} \left(\nabla p + \rho_1 \mathbf{g} \right)$$

$$\mathbf{q}_2 = - k \frac{k_2}{\mu_2} \left(\nabla p + \rho_2 \mathbf{g} \right)$$

where k_1, k_2 are the *relative permeability coefficients*, in general depending on S.

We make now some other simplifying assumptions:

– gravitational effects are negligible,

– $k, \phi, \rho_1, \rho_2, \mu_1, \mu_2$ are constant,

– $k_1 = k_1(S), k_2 = k_2(S)$ are known.

Equations (4.103) and (4.102) become:

$$\phi S_t + \operatorname{div} \mathbf{q}_1 = 0, \qquad -\phi S_t + \operatorname{div} \mathbf{q}_2 = 0 \qquad (4.104)$$

while the Darcy laws take the form

$$\mathbf{q}_1 = - k \frac{k_1}{\mu_1} \nabla p, \qquad \mathbf{q}_2 = - k \frac{k_2}{\mu_2} \nabla p. \qquad (4.105)$$

Letting $\mathbf{q} = \mathbf{q}_1 + \mathbf{q}_2$ and adding the two equations in (4.104) we have

$$\operatorname{div} \mathbf{q} = 0.$$

Adding the two equations in (4.105) yields

$$\nabla p = -\frac{1}{k} \left(\frac{k_1}{\mu_1} + \frac{k_2}{\mu_2} \right)^{-1} \mathbf{q}$$

from which

$$\operatorname{div} \nabla p = \Delta p = -\frac{1}{k} \mathbf{q} \cdot \nabla \left(\frac{k_1}{\mu_1} + \frac{k_2}{\mu_2} \right)^{-1}.$$

From the first equations in (4.104) and (4.105) we get

$$\phi S_t = - \operatorname{div} \mathbf{q}_1 = k \nabla \left(\frac{k_1}{\mu_1} \right) \cdot \nabla p + k \frac{k_1}{\mu_1} \Delta p$$

$$= - \left(\frac{k_1}{\mu_1} + \frac{k_2}{\mu_2} \right)^{-1} \mathbf{q} \cdot \nabla \left(\frac{k_1}{\mu_1} \right) - \frac{k_1}{\mu_1} \mathbf{q} \cdot \nabla \left(\frac{k_1}{\mu_1} + \frac{k_2}{\mu_2} \right)^{-1}$$

$$= \mathbf{q} \cdot \nabla H(S) = H'(S) \, \mathbf{q} \cdot \nabla S$$

where

$$H(S) = -\frac{k_1(S)}{\mu_1} \left(\frac{k_1(S)}{\mu_1} + \frac{k_2(S)}{\mu_2} \right)^{-1}.$$

When $\mathbf{q}$ is known, the resulting *quasilinear* equation for the saturation S is

$$\phi S_t = H'(S)\,\mathbf{q}\cdot\nabla S,$$

known as the *Bukley-Leverett* equation.

In particular, if $\mathbf{q}$ can be considered one-dimensional and constant, i.e. $\mathbf{q} = q\mathbf{i}$, we have

$$qH'(S)\,S_x + \phi S_t = 0$$

which is of the form (4.97), with $u = S$ and $y = t$. The characteristic system is (see Remark 4.10)

$$\frac{dx}{qH'(S)} = \frac{dt}{\phi} = \frac{dS}{0}.$$

Two first integrals are

$$w_1 = \phi x - qH'(S)\,t$$

and $w_2 = S$. Thus, the general solution is given by

$$F\left(\phi x - qH'(S)\,t, S\right) = 0.$$

The choice

$$F\left(w_1, w_2\right) = w_2 - f\left(w_1\right),$$

yields

$$S = f\left(\phi x - qH'(S)\,t\right)$$

that satisfies the initial condition $S(x,0) = f(\phi x)$.

4.6 General First Order Equations

4.6.1 Characteristic strips

We extend the characteristic method to *nonlinear* equations of the form

$$F\left(x, y, u, u_x, u_y\right) = 0 \tag{4.106}$$

We assume that $F = F(x, y, u, p, q)$ is a smooth function of its arguments and, to avoid trivial cases, that $F_p^2 + F_q^2 \neq 0$. In the quasilinear case,

$$F\left(x, y, u, p, q\right) = a\left(x, y, u\right)p + b\left(x, y, u\right)q - c\left(x, y, u\right)$$

and

$$F_p = a\left(x, y, u\right), \quad F_q = b\left(x, y, u\right) \tag{4.107}$$

so that $F_p^2 + F_q^2 \neq 0$ says that a and b do not vanish simultaneously.

Equation (4.106) has a geometrical interpretation as well. Let $u = u(x,y)$ be a smooth solution and consider a point (x_0, y_0, z_0) on its graph. Equation (4.106) constitutes a link between the components u_x and u_y of the normal vector

$$\mathbf{n}_0 = \left(-u_x\left(x_0, y_0\right), -u_y\left(x_0, y_0\right), 1\right)$$

but it is a little more complicated than in the quasilinear case[22] and it is not a priori clear what a characteristic system for equation (4.106) should be. Reasoning by analogy, from (4.107) we are lead to the choice

$$\frac{dx}{dt} = F_p(x, y, z, p, q)$$
$$\frac{dy}{dt} = F_q(x, y, z, p, q). \tag{4.108}$$

where $z(t) = u(x(t), y(t))$ and

$$p = p(t) = u_x(x(t), y(t)), \qquad q = q(t) = u_y(x(t), y(t)). \tag{4.109}$$

Thus, taking account of (4.108), the equation for z is:

$$\frac{dz}{dt} = u_x \frac{dx}{dt} + u_y \frac{dy}{dt} = pF_p + qF_q \tag{4.110}$$

Equations (4.110) and (4.108) correspond to the characteristic system (4.76), but **with two more unknown functions**: $p(t)$ e $q(t)$.

We need two more equations. Proceeding formally, from (4.108) we can write

$$\frac{dp}{dt} = u_{xx}(x(t), y(t)) \frac{dx}{dt} + u_{xy}(x(t), y(t)) \frac{dy}{dt}$$
$$= u_{xx}(x(t), y(t)) F_p + u_{xy}(x(t), y(t)) F_q.$$

We have to get rid of the second order derivatives. Since u is a solution of (4.106), the identity

$$F(x, y, u(x, y), u_x(x, y), u_y(x, y)) \equiv 0.$$

holds. Partial differentiation with respect to x yields, since $u_{xy} = u_{yx}$:

$$F_x + F_u u_x + F_p u_{xx} + F_q u_{xy} \equiv 0.$$

Computing along $x = x(t)$, $y = y(t)$, we get

$$u_{xx}(x(t), y(t)) F_p + u_{xy}(x(t), y(t)) F_q = -F_x - p(t) F_u. \tag{4.111}$$

---

[22] If, for instance $F_q \neq 0$, by the Implicit Function Theorem, the equation $F(x_0, y_0, z_0, p, q) = 0$ defines $q = q(p)$ so that

$$F(x_0, y_0, z_0, p, q(p)) = 0.$$

Therefore, the possible tangent planes to u at (x_0, y_0, z_0) form a one parameter family of planes, given by

$$p(x - x_0) + q(p)(y - y_0) - (z - z_0) = 0.$$

This family, in general, envelopes a cone with vertex at (x_0, y_0, z_0), called *Monge cone*. Each possible tangent plane touches the Monge cone along a generatrix.

Thus, we deduce for p the following differential equation:

$$\frac{dp}{dt} = -F_x\left(x, y, z, p, q\right) - pF_u\left(x, y, z, p, q\right).$$

Similarly, we find

$$\frac{dq}{dt} = -F_y\left(x, y, z, p, q\right) - qF_u\left(x, y, z, p, q\right).$$

In conclusion, we are lead to the following **characteristic system** of five autonomous equations

$$\frac{dx}{dt} = F_p, \qquad \frac{dy}{dt} = F_q, \qquad \frac{dz}{dt} = pF_p + qF_q \tag{4.112}$$

and

$$\frac{dp}{dt} = -F_x - pF_u, \qquad \frac{dq}{dt} = -F_y - qF_u. \tag{4.113}$$

Observe that $F = F\left(x, y, u, p, q\right)$ is a **first integral of** (4.112), (4.113). In fact

$$\frac{d}{dt}F\left(x\left(t\right), y\left(t\right), z\left(t\right), p\left(t\right), q\left(t\right)\right)$$
$$= F_x\frac{dx}{dt} + F_y\frac{dy}{dt} + F_u\frac{dz}{dt} + F_p\frac{dp}{dt} + F_q\frac{dq}{dt}$$
$$= F_xF_p + F_yF_q + F_u\left(pF_p + qF_q\right) + F_p\left(-F_x - pF_p\right) + F_q\left(-F_y - qF_q\right)$$
$$\equiv 0$$

and therefore, if $F\left(x\left(t_0\right), y\left(t_0\right), z\left(t_0\right), p\left(t_0\right), q\left(t_0\right)\right) = 0$ at some t_0, then

$$F\left(x\left(t\right), y\left(t\right), z\left(t\right), p\left(t\right), q\left(t\right)\right) \equiv 0. \tag{4.114}$$

Thus, the curve,

$$x = x\left(t\right), \quad y = y\left(t\right), \quad z = z\left(t\right),$$

still called a *characteristic curve*, is contained in an integral surface, while

$$p = p\left(t\right), \quad q = q\left(t\right)$$

give the normal vector at each point, and can be associated with a piece of the tangent plane, as shown in figure 4.26.

For this reason, a solution $\left(x\left(t\right), y\left(t\right), z\left(t\right), p\left(t\right), q\left(t\right)\right)$ of (4.112), (4.113) is called *characteristic strip* (Fig. 4.26).
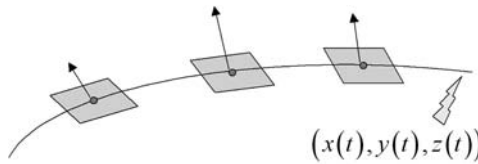
Fig. 4.26. Characteristic strip

4.6.2 The Cauchy Problem

As usual, the *Cauchy problem* consists in looking for a solution u of (4.106), assuming prescribed values on a given curve γ_0 in the x, y plane. If γ_0 has the parametrization

$$x = f(s), \qquad y = g(s) \qquad s \in I \subseteq \mathbb{R}$$

we want that

$$u(f(s), g(s)) = h(s), \qquad s \in I,$$

where $h = h(s)$ is a given function. We assume that $0 \in I$ and that f, g, h are smooth functions in I.

Let Γ_0 be the *initial curve*, given by the parametrization

$$x = f(s), \qquad y = g(s), \qquad z = h(s). \tag{4.115}$$

Equations (4.115) only specify the "initial" points for x, y and z. To solve the characteristic system, we have first to complete Γ_0 into a *strip*

$$(f(s), g(s), h(s), \varphi(s), \psi(s))$$

where

$$\varphi(s) = u_x(f(s), g(s)) \quad \text{and} \quad \psi(s) = u_y(f(s), g(s)).$$

The two functions $\varphi(s)$ and $\psi(s)$ represent the initial values for p and q and cannot be chosen arbitrarily. In fact, a first condition that $\varphi(s)$ and $\psi(s)$ have to satisfy is (recall (4.114)):

$$F(f(s), g(s), h(s), \varphi(s), \psi(s)) \equiv 0. \tag{4.116}$$

A second condition comes from differentiating $h(s) = u(f(s), g(s))$. The result is the so called *strip condition*

$$h'(s) = \varphi(s) f'(s) + \psi(s) g'(s). \tag{4.117}$$

Now we are in position to give a (formal) procedure to construct a solution of our Cauchy problem: *Determine a solution $u = u(x, y)$ of*

$$F(x, y, u, u_x, u_y) = 0,$$

containing the initial curve $(f(s), g(s), h(s))$:

1. Solve for $\varphi(s)$ and $\psi(s)$ the (nonlinear) system

$$\begin{cases} F(f(s), g(s), h(s), \varphi(s), \psi(s)) = 0 \\ \varphi(s) f'(s) + \psi(s) g'(s) = h'(s). \end{cases} \tag{4.118}$$

2. Solve the characteristic system (4.112), (4.113) with initial conditions

$$x(0) = f(s), y(0) = g(s), z(0) = h(s), p(0) = \varphi(s), q(0) = \psi(s).$$

Suppose we find the solution

$$x = X\left(s, t\right), y = Y\left(s, t\right), z = Z\left(s, t\right), p = P\left(s, t\right), q = Q\left(s, t\right).$$

3. Solve $x = X\left(s, t\right), y = Y\left(s, t\right)$ for s, t in terms of x, y. Substitute $s = S\left(x, y\right)$ and $t = T\left(x, y\right)$ into $z = Z\left(t, s\right)$ to yield a solution $z = u\left(x, y\right)$.

Example 4.9. We want to solve the equation

$$u = u_x^2 - 3u_y^2$$

with initial condition $u\left(x, 0\right) = x^2$. We have $F\left(p, q\right) = p^2 - 3q^2 - u$ and the characteristic system is

$$\frac{dx}{dt} = 2p, \quad \frac{dy}{dt} = -6q, \quad \frac{dz}{dt} = 2p^2 - 6q^2 = 2z \tag{4.119}$$

$$\frac{dp}{dt} = p, \qquad \frac{dq}{dt} = q. \tag{4.120}$$

A parametrization of the initial line Γ_0 is

$$f\left(s\right) = s, \ g\left(s\right) = 0, \ h\left(s\right) = s^2.$$

To complete the initial strip we solve the system (4.118):

$$\begin{cases} \varphi^2 - 3\psi^2 = s^2 \\ \varphi = 2s. \end{cases}$$

There are two solutions:

$$\varphi\left(s\right) = 2s, \qquad \psi\left(s\right) = \pm s.$$

The choice of $\psi\left(s\right) = s$ yields, integrating equations (4.120),

$$P\left(s, t\right) = 2se^t, \qquad Q\left(s, t\right) = se^t$$

whence, from (4.119),

$$X\left(s, t\right) = 4s(e^t - 1) + s, \qquad Y\left(s, t\right) = -6s(e^t - 1), \qquad Z\left(s, t\right) = s^2 e^{2t}.$$

Solving the first two equations for s, t and substituting into the third one, we get

$$u\left(x, y\right) = \left(x + \frac{y}{2}\right)^2.$$

The choice of $\psi\left(s\right) = -s$ yields

$$u\left(x, y\right) = \left(x - \frac{y}{2}\right)^2.$$

As the example shows, in general *there is no uniqueness*, unless system (4.118) has a unique solution. On the other hand, if this system has no (real) solution, then the Cauchy problem has no solution as well.

Furthermore, observe that if $(x_0, y_0, z_0) = (f(0), g(0), h(0))$ and (p_0, q_0) is a solution of the system

$$\begin{cases} F(x_0, y_0, z_0, p_0, q_0) = 0 \\ p_0 f'(0) + q_0 g'(0) = h'(0), \end{cases}$$

by the Implicit Function Theorem, the condition

$$\begin{vmatrix} f'(0) & F_p(x_0, y_0, z_0, p_0, q_0) \\ g'(0) & F_q(x_0, y_0, z_0, p_0, q_0) \end{vmatrix} \neq 0 \tag{4.121}$$

assures the existence of a solution $\varphi(s)$ and $\psi(s)$ of (4.118), in a neighborhood of $s = 0$. Condition (4.121) corresponds to (4.84) in the quasilinear case.

The following theorem summarizes the above discussion on the Cauchy problem

$$F(x, y, u, u_x, u_y) = 0 \tag{4.122}$$

with initial curve Γ_0 given by

$$x = f(s), \quad y = g(s), \quad z = h(s). \tag{4.123}$$

Theorem 4.5. *Assume that:*

i) F is twice continuously differentiable in a domain $D \subseteq \mathbb{R}^5$ and $F_p^2 + F_q^2 \neq 0$;

ii) f, g, h are twice continuously differentiable in a neighborhood of $s = 0$.

iii) (p_0, q_0) is a solution of the system (5.30) and condition (4.121) holds.

Then, in a neighborhood of (x_0, y_0), there exists a C^2 solution $z = u(x, y)$ of the Cauchy problem (4.122), (4.123).

- *Geometrical optics.* The equation

$$c^2 (u_x^2 + u_y^2) = 1 \qquad (c > 0) \tag{4.124}$$

is called *eikonal equation* and arises in (two dimensional) geometrical optics. The level lines γ_t of equation

$$u(x, y) = t \tag{4.125}$$

represent the "wave fronts" of a wave perturbation (i.e. light) moving with time t and c denotes the propagation speed, that we assume to be constant. An orthogonal trajectory to the wave fronts coincides with a *light ray*. A point $(x(t), y(t))$ on a ray satisfies the identity

$$u(x(t), y(t)) = t \tag{4.126}$$

and its velocity vector $\mathbf{v} = (\dot{x}, \dot{y})$ is parallel to ∇u. Therefore

$$\nabla u \cdot \mathbf{v} = |\nabla u| \, |\mathbf{v}| = c \, |\nabla u|.$$

On the other hand, differentiating (4.126) we get

$$\nabla u \cdot \mathbf{v} = u_x \dot{x} + u_y \dot{y} = 1$$

from which

$$c^2 |\nabla u|^2 = 1.$$

Geometrically, if we fix a point (x_0, y_0, z_0), equation $c^2(p^2 + q^2) = 1$ states that the family of planes

$$z - z_0 = p(x - x_0) + q(y - y_0),$$

tangent to an integral surface at (x_0, y_0, z_0), all make a fixed angle $\theta = \arctan |\nabla u|^{-1} = \arctan c$ with the $z-$axis. This family envelopes the circular cone (exercise)

$$(x - x_0)^2 + (y - y_0)^2 = c^2(z - z_0)^2$$

called the *light cone*, with opening angle 2θ.

The *eikonal equation* is of the form (4.106) with[23]

$$F(x, y, u, p, q) = \frac{1}{2} \left[c^2(p^2 + q^2) - 1 \right].$$

The characteristic system is[24]:

$$\frac{dx}{d\tau} = c^2 p, \qquad \frac{dy}{d\tau} = c^2 q, \qquad \frac{dz}{d\tau} = c^2 p^2 + c^2 q^2 = 1 \tag{4.127}$$

and

$$\frac{dp}{d\tau} = 0, \qquad \frac{dq}{d\tau} = 0. \tag{c3}$$

An initial curve Γ_0

$$x = f(s), \qquad y = g(s), \qquad z = h(s),$$

can be completed into an initial strip by solving for ϕ and ψ the system

$$\begin{cases} \phi(s)^2 + \psi(s)^2 = c^{-2} \\ \phi(s) f'(s) + \psi(s) g'(s) = h'(s). \end{cases} \tag{4.128}$$

This system has *two real distinct solutions* if

$$f'(s)^2 + g'(s)^2 > c^2 h'(s)^2 \tag{4.129}$$

while it has *no real solutions* if[25]

$$f'(s)^2 + g'(s)^2 < c^2 h'(s)^2. \tag{4.130}$$

---

[23] The factor $\frac{1}{2}$ is there for esthetic reasons.

[24] Using τ as a parameter along the characteristics.

[25] System (4.128) is equivalent to finding the intersection between the circle $\xi^2 + \eta^2 = c^{-2}$ and the straight line $f'\xi + g'\eta = h'$. The distance of the center $(0, 0)$ from the line is given by

$$d = \frac{|h'|}{\sqrt{(f')^2 + (g')^2}}$$

so that there are 2 real intersections if $d < c^{-1}$, while there is no real intersection if $d > c^{-1}$.

If (4.129) holds, Γ_0 forms an angle greater than θ with the z−axis and therefore it is exterior to the light cone (Fig. 4.22). In this case we say that Γ_0 is *space-like* and we can find two different solutions of the Cauchy problem. If (4.130) holds, Γ_0 is contained in the light cone and we say it is *time-like*. The Cauchy problem does not have any solution.

Given a space-like curve Γ_0 and ϕ, ψ solutions of the system (4.128), the corresponding characteristic strip is, for s fixed,

$$x\left(t\right) = f\left(s\right) + c^2\phi\left(s\right)t, \qquad y\left(t\right) = g\left(s\right) + c^2\psi\left(s\right)t, \qquad z\left(t\right) = h\left(s\right) + t$$

$$p\left(t\right) = \phi\left(s\right), \qquad q\left(t\right) = \psi\left(s\right).$$

Observe that the point $\left(x\left(t\right), y\left(t\right)\right)$ moves along the characteristic with speed

$$\sqrt{\dot{x}^2\left(t\right) + \dot{y}^2\left(t\right)} = \sqrt{\phi^2\left(s\right) + \psi^2\left(s\right)} = c$$

with direction $\left(\phi\left(s\right), \psi\left(s\right)\right) = \left(p\left(t\right), q\left(t\right)\right)$. Therefore, the characteristic lines are coincident with the light rays. Moreover, we see that the fronts γ_t can be constructed from γ_0 by shifting any point on γ_0 along a ray at a distance ct. Thus, the wave fronts constitute a family of "parallel" curves.
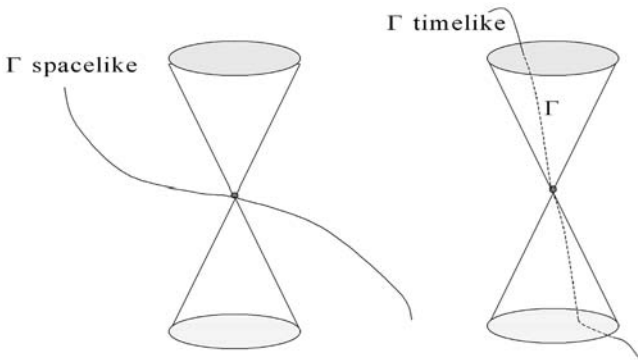
Fig. 4.27. Space-like and time-like initial curves

Problems

4.1. Using *Duhamel's method* (see subsection 2.8.3) solve the problem

$$\begin{cases} c_t + vc_x = f\left(x,t\right) & x \in \mathbb{R}, t > 0 \\ c(x,0) = 0 & x \in \mathbb{R}. \end{cases}$$

Find an explicit formula when $f\left(x,t\right) = e^{-t}\sin x$.

[*Hint.* For fixed $s \geq 0$ and $t > s$, solve

$$\begin{cases} w_t + v w_x = 0 \\ w(x, s; s) = f(x, s). \end{cases}$$

Then, integrate w with respect to s over $(0, t)$].

4.2. Consider the following problem $(a > 0)$:

$$\begin{cases} u_t + a u_x = f(x, t) & 0 < x < R, t > 0 \\ u(0, t) = 0 & t > 0 \\ u(x, 0) = 0 & 0 < x < R. \end{cases}$$

Prove the stability estimate

$$\int_0^R u^2(x, t)\, dx \leq e^t \int_0^t \int_0^R f^2(x, s)\, dx ds, \qquad t > 0.$$

[*Hint.* Multiply by u the equation. Use $a > 0$ and the inequality $2fu \leq f^2 + u^2$ to obtain

$$\frac{d}{dt} \int_0^R u^2(x, t)\, dx \leq \int_0^R f^2(x, t)\, dx + \int_0^R u^2(x, t)\, dx.$$

Prove that if $E(t)$ satisfies

$$E'(t) \leq G(t) + E(t), \qquad E(0) = 0$$

then $E(t) \leq e^t \int_0^t G(s)\, ds$].

4.3. Solve Burger's equation $u_t + u u_x = 0$ with initial data

$$g(x) = \begin{cases} 1 & x \leq 0 \\ 1 - x & 0 < x < 1 \\ 0 & x \geq 1. \end{cases}$$

[*Answer*: See figure 4.28].

4.4. In the Green light problem (subsection 4.3.3) compute:

a) the car density at the light for $t > 0$.

b) The time that a car located at the point $x_0 = -v_m t_0$ at time t_0 takes to reach the light.

[*Hint.* b) If $x = x(t)$ is the position of the car at time t, show that $\frac{dx}{dt} = \frac{v_m}{2} + \frac{x(t)}{2t}$].

4.5. Consider equation (4.19) in section 4.3.1, with initial density

$$\rho_0(x) = \begin{cases} \rho_1 & x < 0 \\ \rho_m & x > 0. \end{cases}$$
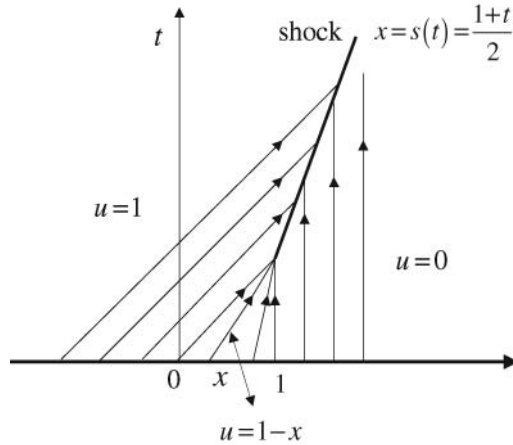
Fig. 4.28. The solution of problem 4.3

Assuming $0 < \rho_1 < \rho_m$, determine the characteristics and construct a solution in the whole half-plane $t > 0$. Give an interpretation of your result.

4.6. Solve the problem

$$
\begin{cases}
u_t + u u_x = 0 & x \in \mathbb{R}, \ t > 0 \\
u(x, 0) = g(x) & x \in \mathbb{R}
\end{cases}
$$

where.

$$
g(x) = \begin{cases}
0 & x < 0 \\
1 & 0 < x < 1 \\
0 & x > 1.
\end{cases}
$$

4.7. *Traffic in a tunnel.* A rather realistic model for the car speed in a very long tunnel is the following:

$$
v(\rho) = \begin{cases}
v_m & 0 \le \rho \le \rho_c \\
\lambda \log\left(\frac{\rho_m}{\rho}\right) & \rho_c \le \rho \le \rho_m
\end{cases}
$$

where

$$
\lambda = \frac{v_m}{\log(\rho_m/\rho_c)}.
$$

Observe that v is continuous also at $\rho_c = \rho_m e^{-v_m/\lambda}$, which represents a *critical density*: if $\rho \le \rho_c$ the drivers are free to reach the speed limit. Typical values are: $\rho_c = 7$ car/Km, $v_m = 90$ Km/h, $\rho_m = 110$ car/Km, $v_m/\lambda = 2.75$.

Assume the entrance is placed at $x = 0$ and that cars are waiting (with maximum density) the tunnel to open to the traffic at time $t = 0$. Thus, the initial density is.

$$
\rho = \begin{cases}
\rho_m & x < 0 \\
0 & x > 0.
\end{cases}
$$

a) Determine density and car speed; draw their graphs as a function of time.

b) Determine and draw in the x, t plane the trajectory of a car initially at $x = x_0 < 0$, and compute the time it takes to enter the tunnel.

4.8. Consider the equation

$$u_t + q'(u) u_x = 0$$

with initial condition $u(x, 0) = g(x)$. Assume that $g, q' \in C^1([a, b])$ and $g'q'' < 0$ in $[a, b]$. Show that the family of characteristics

$$x = q'(u) t + \xi, \qquad \xi \in [a, b] \tag{4.131}$$

admits an *envelope* and that the point (x_s, t_s) of shock formation, given by formulas (4.37) and (4.38), is the point on this envelope with minimum time coordinate (Fig. 4.29).
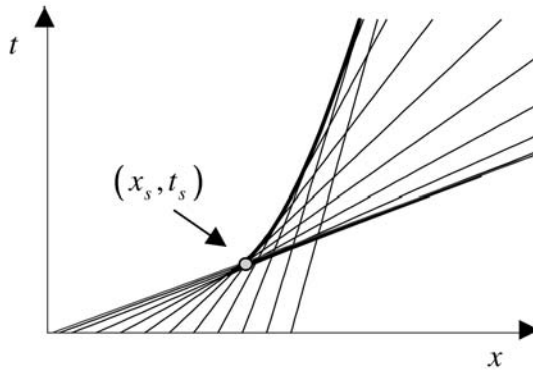
Fig. 4.29. Envelope of characteristics and point of shock formation

4.9. Find the solutions of the problems

$$\begin{cases} u_t \pm u u_x = 0 & t > 0, x \in \mathbb{R} \\ u(x, 0) = x & x \in \mathbb{R}. \end{cases}$$

4.10. Draw the characteristics and describe the evolution for $t \to +\infty$ of the solution of the problem

$$\begin{cases} u_t + u u_x = 0 & t > 0, x \in \mathbb{R} \\ u(x, 0) = \begin{cases} \sin x & 0 < x < \pi \\ 0 & x \leq 0 \text{ or } x \geq \pi \end{cases} \end{cases}$$

4.11. Show that, for every $\alpha > 1$, the function

$$u_\alpha(x, t) = \begin{cases} -1 & 2x < (1 - \alpha) t \\ -\alpha & (1 - \alpha) t < 2x < 0 \\ \alpha & 0 < 2x < (\alpha - 1) t \\ 1 & (\alpha - 1) t < 2x \end{cases}$$

is a weak solution of the problem

$$\begin{cases} u_t + uu_x = 0 & t > 0, x \in \mathbb{R} \\ u(x, 0) = \begin{cases} -1 & x < 0 \\ 1 & x > 0. \end{cases} \end{cases}$$

Is it also an entropy solution, at least for some α?

4.12. Using the Hopf-Cole transformation, solve the following problem for the viscous Burger equation

$$\begin{cases} u_t + uu_x = \varepsilon u_{xx} & t > 0, x \in \mathbb{R} \\ u(0, x) = \mathcal{H}(x) & x \in \mathbb{R}. \end{cases}$$

where $\mathcal{H}$ is the Heaviside function.

Show that, as $t \to +\infty$, $u(x, t)$ converges to a travelling wave similar to (4.68).
[*Answer.* The solution is

$$u(x, t) = \frac{1}{1 + \dfrac{\text{erfc}\left(-x/\sqrt{4\varepsilon t}\right)}{\text{erfc}\left((x - t)/\sqrt{4\varepsilon t}\right)} \exp\left(\dfrac{x - t/2}{2\varepsilon}\right)}$$

where

$$\text{erfc}(s) = \int_s^{+\infty} \exp\left(-z^2\right) dz$$

is the *complementary error function*].

4.13. Find the solution of the linear equation

$$u_x + xu_y = y$$

satisfying the initial condition $u(0, y) = g(y)$, $y \in \mathbb{R}$, with

$$(a)\ g(y) = \cos y \quad \text{and} \quad (b)\ g(y) = y^2.$$

[*Answer of (a)*:

$$u = xy - \frac{x^3}{3} + \cos\left(y - \frac{x^2}{2}\right)].$$

4.14. Consider the linear equation

$$au_x + bu_y = c(x, y),$$

where a, b are constants ($b \neq 0$), and the initial condition $u(x, 0) = h(x)$.

1) Show that

$$u(x, y) = h(x - \gamma y) + \int_0^{y/b} c(a\tau + x - \gamma y, b\tau)\, d\tau \qquad \gamma = a/b.$$

2) Deduce that a jump discontinuity at x_0 of h, propagates into a jump of the same size for u along the characteristic of equation $x - \gamma y = x_0$.

4.15. Let $D = \{(x, y) : y > x^2\}$ and $a = a(x, y)$ be a continuous function in $\overline{D}$.

1) Check the solvability of the problem

$$\begin{cases} a(x, y) u_x - u_y = -u & (x, y) \in D \\ u(x, x^2) = g(x) & x \in \mathbb{R}. \end{cases}$$

2) Examine the case

$$a(x, y) = y/2 \quad \text{and} \quad g(x) = \exp\left(-\gamma x^2\right),$$

where γ is a real parameter.

4.16. Solve the Cauchy problem

$$\begin{cases} x u_x - y u_y = u - y & x > 0, y > 0 \\ u(y^2, y) = y & y > 0. \end{cases}$$

May a solution exist in a neighborhood of the origin?

[*Answer.*

$$u(x, y) = \left(y + x^{2/3} y^{-1/3}\right)/2.$$

In no neighborhood of $(0, 0)$ a solution can exist].

4.17. Consider a cylindrical pipe with axis along the $x-$axis, filled with a fluid moving along the positive direction. Let $\rho = \rho(x, t)$ and $q = \frac{1}{2}\rho^2$ be the fluid density and the flux function. Assume the walls of the pipe are composed by porous material, from which the fluid leaks at the rate $H = k\rho^2$.

a) Following the derivation of the conservation law given in the introduction, show that ρ satisfies the equation.

$$\rho_t + \rho \rho_x = -k\rho^2$$

b) Solve the Cauchy problem with $\rho(x, 0) = 1$.

[*Answer.* b) $\rho(x, t) = 1/(1 + kt)$].

4.18. Solve the Cauchy problem

$$\begin{cases} u_x = -(u_y)^2 & x > 0, y \in \mathbb{R} \\ u(0, y) = 3y & y \in \mathbb{R}. \end{cases},$$

4.19. Solve the Cauchy problem

$$\begin{cases} u_x^2 + u_y^2 = 4u \\ u(x, -1) = x^2 & x \in \mathbb{R}. \end{cases}$$

[*Answer:* $u(x, y) = x^2 + (y + 1)^2$]

4.20. Solve the Cauchy problem

$$\begin{cases} c^2 \left(u_x^2 + u_y^2 \right) = 1 \\ u \left(\cos s, \sin s \right) = 0 \end{cases} \qquad s \in \mathbb{R}.$$

[*Answer*: There are two solutions

$$u^{\pm} \left(x, y \right) = \frac{\pm 1}{c} \left\{ 1 - \sqrt{x^2 + y^2} \right\}$$

whose wave fronts are shown in figure 4.30].
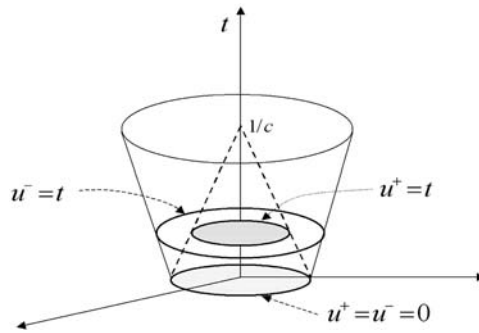
Fig. 4.30. Solutions of problem 4.20

5

Waves and Vibrations

General Concepts – Transversal Waves in a String – The One-dimensional Wave Equation – The d'Alembert Formula – Second Order Linear Equations – Hyperbolic Systems With Constant Coefficients – The Multi-dimensional Wave Equation ($n > 1$) – Two Classical Models – The Cauchy Problem – Linear Water Waves

5.1 General Concepts

5.1.1 Types of waves

Our dayly experience deals with sound waves, electromagnetic waves (as radio or light waves), deep or surface water waves, elastic waves in solid materials. Oscillatory phenomena manifest themselves also in contexts and ways less macroscopic and known. This is the case, for instance, of rarefaction and shock waves in traffic dynamics or of electrochemical waves in human nervous system and in the regulation of the heart beat. In quantum physics, everything can be described in terms of wave functions, at a sufficiently small scale.

Although the above phenomena share many similarities, they show several differences as well. For example, progressive water waves propagate a disturbance, while standing waves do not. Sound waves need a supporting medium, while electromagnetic waves do not. Electrochemical waves interact with the supporting medium, in general modifying it, while water waves do not.

Thus, it seems too hard to give a general definition of *wave*, capable of covering all the above cases, so that we limit ourselves to introducing some terminology and general concepts, related to specific types of waves. We start with one-dimensional waves.

a. **Progressive** or **travelling** waves are disturbances described by a function of the following form:
$$u(x,t) = g(x - ct).$$
For $t = 0$, we have $u(x,0) = g(x)$, which is the "initial" profile of the perturbation. This profile propagates without change of shape with speed $|c|$, in the positive

(negative) $x-$direction if $c > 0$ $(c < 0)$. We have already met this kind of waves in Chapters 2 and 4.

b. Harmonic waves are particular progressive waves of the form

$$u\left(x, t\right) = A \exp\left\{i\left(kx - \omega t\right)\right\}, \qquad A, k, \omega \in \mathbb{R}. \tag{5.1}$$

It is understood that only the *real part* (or the imaginary part)

$$A \cos\left(kx - \omega t\right)$$

is of interest, but the complex notation may often simplify the computations. In (5.1) we distinguish, considering for simplicity ω and k positive:

- The wave *amplitude* $|A|$;

- The *wave number* k, which is the number of complete oscillations in the space interval $[0, 2\pi]$, and the *wavelength*

$$\lambda = \frac{2\pi}{k}$$

  which is the distance between successive maxima (*crest*) or minima (*troughs*) of the waveform;

- The *angular frequency* ω, and the *frequency*

$$f = \frac{\omega}{2\pi}$$

  which is the number of complete oscillations in one second (Hertz) at a fixed space position;

- The *wave or phase speed*

$$c_p = \frac{\omega}{k}$$

  which is the crests (or troughs) speed;
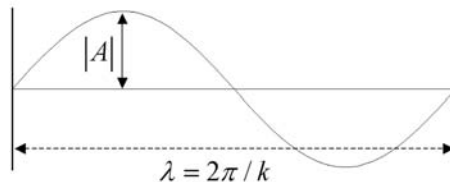
Fig. 5.1. Sinusoidal wave

   c. **Standing** waves are of the form

$$u(x,t) = B \cos kx \cos \omega t.$$

In these disturbances, the basic sinusoidal wave, $\cos kx$, is modulated by the time dependent oscillation $B \cos \omega t$. A standing wave may be generated, for instance, by superposing two harmonic waves with the same amplitude, propagating in opposite directions:

$$A \cos(kx - \omega t) + A \cos(kx + \omega t) = 2A \cos kx \cos \omega t. \qquad (5.2)$$

Consider now waves in dimension $n > 1$.

   d. **Plane waves**. *Scalar* plane waves are of the form

$$u(\mathbf{x},t) = f(\mathbf{k} \cdot \mathbf{x} - \omega t).$$

The disturbance propagates in the direction of $\mathbf{k}$ with speed $c_p = \omega/|\mathbf{k}|$. The planes of equation

$$\theta(\mathbf{x},t) = \mathbf{k} \cdot \mathbf{x} - \omega t = \text{constant}$$

constitute the *wave-fronts*.

   Harmonic or monochromatic plane waves have the form

$$u(\mathbf{x}, t) = A \exp\{i(\mathbf{k} \cdot \mathbf{x} - \omega t)\}.$$

Here $\mathbf{k}$ is the *wave number* vector and ω is the *angular frequency*. The vector $\mathbf{k}$ is orthogonal to the wave front and $|\mathbf{k}|/2\pi$ gives the number of waves per unit length. The scalar $\omega/2\pi$ still gives the number of complete oscillations in one second (Hertz) at a fixed space position.

   e. **Spherical waves** are of the form

$$u(\mathbf{x},t) = v(r, t)$$

where $r = |\mathbf{x} - \mathbf{x}_0|$ and $\mathbf{x}_0 \in \mathbb{R}^n$ is a fixed point. In particular $u(\mathbf{x},t) = e^{i\omega t} v(r)$ represents a stationary spherical wave, while $u(\mathbf{x},t) = v(r - ct)$ is a progressive wave whose wavefronts are the spheres $r - ct = \text{constant}$, moving with speed $|c|$ (outgoing if $c > 0$, incoming if $c < 0$).

5.1.2 Group velocity and dispersion relation

Many oscillatory phenomena can be modelled by linear equations whose solutions are superpositions of harmonic waves with angular frequency depending on the wave number:

$$\omega = \omega(k). \qquad (5.3)$$

A typical example is the wave system produced by dropping a stone in a pond.

   If ω is linear, e.g. $\omega(k) = ck$, $c > 0$, the crests move with speed c, independent of the wave number. However, if $\omega(k)$ is not proportional to k, the crests move with

speed $c_p = \omega(k)/k$, that *depends* on the wave number. In other words, the crests move at different speeds for different wavelengths. As a consequence, the various components in a wave packet given by the superposition of harmonic waves of different wavelengths will eventually separate or *disperse*. For this reason, (5.3) is called **dispersion relation**.

In the theory of dispersive waves, the **group velocity**, given by

$$c_g = \omega'(k)$$

is a central notion, mainly for the following three reasons.

1. *It is the speed at which an isolated wave packet moves as a whole.* A wave packet may be obtained by the superposition of dispersive harmonic waves, for instance through a Fourier integral of the form

$$u(x,t) = \int_{-\infty}^{+\infty} a(k) e^{i[kx-\omega(k)t]} dk \tag{5.4}$$

where the real part only has a physical meaning. Consider a localized wave packet, with wave number $k \approx k_0$, almost constant, and with amplitude slowly varying with x. Then, the packet contains a large number of crests and the amplitudes $|a(k)|$ of the various Fourier components are negligible except that in a small neighborhood of k_0, $(k_0 - \delta, k_0 + \delta)$, say.

Figure 5.2 shows the initial profile of a Gaussian packet,

$$\operatorname{Re} u(x,0) = \frac{3}{\sqrt{2}} \exp\left\{-\frac{x^2}{32}\right\} \cos 14x,$$

slowly varying with x, with $k_0 = 14$, and its Fourier transform:

$$a(k) = 6 \exp\{-8(k-14)^2\}.$$

As we can see, the amplitudes $|a(k)|$ of the various Fourier components are negligible except when k is near k_0.
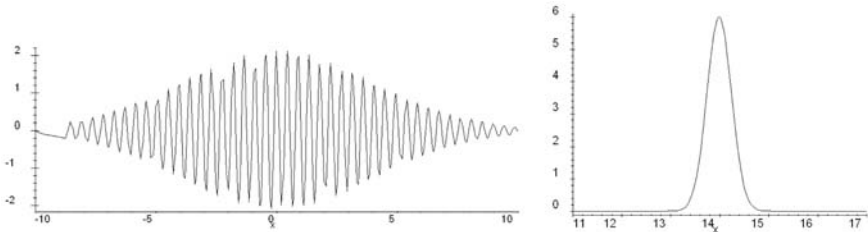
Fig. 5.2. Wave packet and its Fourier transform

Then we may write

$$\omega(k) \approx \omega(k_0) + \omega'(k_0)(k - k_0) = \omega(k_0) + c_g(k - k_0)$$

and

$$u\left(x,t\right) \approx e^{i\{k_0 x - \omega(k_0)t\}} \int_{k_0-\delta}^{k_0+\delta} a\left(k\right) e^{i(k-k_0)(x-c_g t)} dk. \tag{5.5}$$

Thus, u turns out to be well approximated by the product of two waves. The first one is a pure harmonic wave with relatively short wavelength $2\pi/k_0$ and phase speed $\omega\left(k_0\right)/k_0$. The second one depends on x, t through the combination $x - c_g t$, and is a superposition of waves of very small wavenumbers $k - k_0$, which correspond to very large wavelengths. We may interpret the second factor as a sort of envelope of the short waves of the packet, that is the packet as a whole, which therefore moves with the group speed.

2. *An observer that travels at the group velocity sees constantly waves of the same wavelength $2\pi/k$,* after the transitory effects due to a localized initial perturbation (e.g. a stone thrown into a pond). In other words, c_g is the propagation speed of the wave numbers.

Imagine dropping a stone into a pond. At the beginning, the water perturbation looks complicated, but after a sufficiently long time, the various Fourier components will be quite dispersed and the perturbation will appear as a slowly modulated wave train, almost sinusoidal near every point, with a *local wave number* $k\left(x,t\right)$ and a *local frequency* $\omega\left(x,t\right)$. If the water is deep enough, we expect that, at each fixed time t, the wavelength increases with the distance from the stone (longer waves move faster, see subsection 5.10.4) and that, at each fixed point x, the wavelength tends to decrease with time.

Thus, the essential features of the wave system can be observed at a relatively long distance from the location of the initial disturbance and after some time has elapsed.

Let us assume that the free surface displacement u is given by a Fourier integral of the form (5.4). We are interested on the behavior of u for $t \gg 1$. An important tool comes from the method of stationary phase[1] which gives an asymptotic formula for integrals of the form

$$I\left(t\right) = \int_{-\infty}^{+\infty} f\left(k\right) e^{it\varphi(k)} dk \tag{5.6}$$

as $t \to +\infty$. We can put u into the form (5.6) by writing

$$u\left(x,t\right) = \int_{-\infty}^{+\infty} a\left(k\right) e^{it[k\frac{x}{t} - \omega(k)]} dk,$$

then by moving from the origin at a fixed speed V (thus $x = Vt$) and defining

$$\varphi\left(k\right) = kV - \omega\left(k\right).$$

Assume for simplicity that φ has only one stationary point k_0, that is

$$\omega'\left(k_0\right) = V,$$

---

[1] See subsection 5.10.6

and that $\omega''(k_0) \neq 0$. Then, according to the *method of stationary phase*, we can write

$$u(Vt, t) = \sqrt{\frac{\pi}{|\omega''(k_0)|}} \frac{a(k_0)}{\sqrt{t}} \exp\left\{it\left[k_0 V - \omega(k_0)\right]\right\} + O\left(t^{-1}\right). \qquad (5.7)$$

Thus, if we allow errors of order t^{-1}, moving with speed $V = \omega'(k_0) = c_g$, *the same wave number k_0 always appears at the position $x = c_g t$*. Note that the amplitude decreases like $t^{-1/2}$ as $t \to +\infty$. This is an important attenuation effect of dispersion.

3. *Energy is transported at the group velocity by waves of wavelength $2\pi/k$.*
In a wave packet like (5.5), the energy is proportional to[2]

$$\int_{k_0-\delta}^{k_0+\delta} |a(k)|^2 \, dk \simeq 2\delta |a(k_0)|^2$$

so that it moves at the same speed of k_0, that is c_g.

Since the energy travels at the group velocity, there are significant differences in the wave system according to the sign of $c_g - c_p$, as we will see in Section 10.

5.2 Transversal Waves in a String

5.2.1 The model

We derive a classical model for the small transversal vibrations of a tightly stretched horizontal string (e.g. a string of a guitar). We assume the following hypotheses:

1. *Vibrations of the string have small amplitude.* This entails that the changes in the slope of the string from the horizontal equilibrium position are very small.
2. *Each point of the string undergoes vertical displacements only.* Horizontal displacements can be neglected, according to 1.
3. *The vertical displacement of a point depends on time and on its position on the string.* If we denote by u the vertical displacement of a point located at x when the string is at rest, then we have $u = u(x, t)$ and, according to 1, $|u_x(x, t)| \ll 1$.
4. *The string is perfectly flexible.* This means that it offers no resistance to bending. In particular, the stress at any point on the string can be modelled by a tangential[3] force **T** of magnitude τ, called *tension*. Figure 5.3 shows how the forces due to the tension acts at the end points of a small segment of the string.
5. *Friction is negligible.*

Under the above assumptions, the equation of motion of the string can be derived from *conservation of mass* and *Newton law*.

---

[2] See *A. Segel*, 1987.

[3] Consequence of absence of distributed moments along the string.
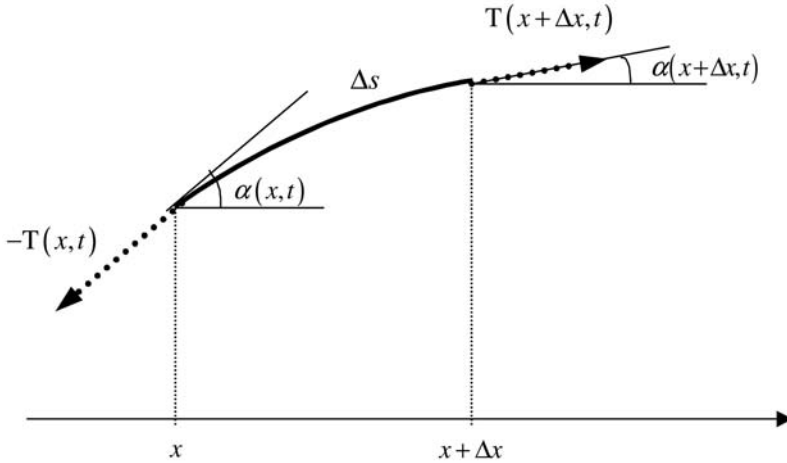
Fig. 5.3. Tension at the end points of a small segment of a string

Let $\rho_0 = \rho_0(x)$ be the linear density of the string at rest and $\rho = \rho(x,t)$ be its density at time t. Consider an arbitrary part of the string between x and $x + \Delta x$ and denote by Δs the corresponding length element at time t. Then, conservation of mass yields

$$\rho_0(x)\,\Delta x = \rho(x,t)\,\Delta s. \tag{5.8}$$

To write Newton law of motion we have to determine the forces acting on our small piece of string. Since the motion is vertical, the horizontal forces have to balance. On the other hand they come from the tension only, so that if $\tau(x,t)$ denotes the magnitude of the tension at x at time t, we can write (Fig. 5.3):

$$\tau(x+\Delta x, t)\cos\alpha(x+\Delta x, t) - \tau(x,t)\cos\alpha(x,t) = 0.$$

Dividing by Δx and letting $\Delta x \to 0$, we obtain

$$\frac{\partial}{\partial x}\left[\tau(x,t)\cos\alpha(x,t)\right] = 0$$

from which

$$\tau(x,t)\cos\alpha(x,t) = \tau_0(t) \tag{5.9}$$

where $\tau_0(t)$ is *positive*[4].

The vertical forces are given by the vertical component of the tension and by body forces such as gravity and external loads.

Using (5.9), the scalar vertical component of the tension at x, at time t, is given by:

$$\tau_{vert}(x,t) = \tau(x,t)\sin\alpha(x,t) = \tau_0(t)\tan\alpha(x,t) = \tau_0(t)\,u_x(x,t).$$

---

[4] It is the magnitude of a force.

Therefore, the (scalar) vertical component of the force acting on our small piece of string, due to the tension, is

$$\tau_{vert}\left(x + \Delta x, t\right) - \tau_{vert}\left(x, t\right) = \tau_0\left(t\right)\left[u_x\left(x + \Delta x, t\right) - u_x\left(x, t\right)\right].$$

Denote by $f\left(x, t\right)$ the magnitude of the (vertical) body forces per unit mass. Then, using (5.8), the magnitude of the body forces acting on the string segment is given by:

$$\int_x^{x+\Delta x} \rho\left(y, t\right) f\left(y, t\right) ds = \int_x^{x+\Delta x} \rho_0\left(y\right) f\left(y, t\right) dy.$$

Thus, using (5.8) again and observing that u_{tt} is the (scalar) vertical acceleration, Newton law gives:

$$\int_x^{x+\Delta x} \rho_0\left(y\right) u_{tt}\left(y, t\right) dy = \tau_0\left(t\right)\left[u_x\left(x + \Delta x, t\right) - u_x\left(x, t\right)\right] + \int_x^{x+\Delta x} \rho_0\left(y\right) f\left(y, t\right) dy.$$

Dividing by Δx and letting $\Delta x \to 0$, we obtain the equation

$$u_{tt} - c^2\left(x, t\right) u_{xx} = f\left(x, t\right) \tag{5.10}$$

where $c^2\left(x, t\right) = \tau_0\left(t\right)/\rho_0\left(x\right)$.

If the string is homogeneous then ρ_0 is constant. If moreover it is **perfectly elastic**[5] then τ_0 is constant as well, since the horizontal tension is nearly the same as for the string at rest, in the horizontal position. We shall come back to equation (5.10) shortly.

5.2.2 Energy

Suppose that a *perfectly flexible and elastic* string has length L at rest, in the horizontal position. We may identify its initial position with the segment $[0, L]$ on the x axis. Since $u_t(x, t)$ is the vertical velocity of the point at x, the expression

$$E_{cin}\left(t\right) = \frac{1}{2} \int_0^L \rho_0 u_t^2 \, dx \tag{5.11}$$

represents the total **kinetic energy during the vibrations**. The string stores **potential energy** too, due to the work of elastic forces. These forces stretch an element of string of length Δx at rest by[6]

$$\Delta s - \Delta x = \int_x^{x+\Delta x} \sqrt{1 + u_x^2} \, dx - \Delta x = \int_x^{x+\Delta x} \left(\sqrt{1 + u_x^2} - 1\right) dx \approx \frac{1}{2} u_x^2 \Delta x$$

---

[5] For instance, guitar and violin strings are nearly homogeneous, perfectly flexible and elastic.

[6] Recall that, at first order, if $\varepsilon \ll 1$, $\sqrt{1 + \varepsilon} - 1 \simeq \varepsilon/2$.

since $|u_x| \ll 1$. Thus, the work done by the elastic forces on that string element is

$$dW = \frac{1}{2} \tau_0 u_x^2 \Delta x.$$

Summing all the contributions, the total **potential energy** is given by:

$$E_{pot}(t) = \frac{1}{2} \int_0^L \tau_0 u_x^2 \, dx. \tag{5.12}$$

From (5.11) and (5.12) we find, for the total energy:

$$E(t) = \frac{1}{2} \int_0^L [\rho_0 u_t^2 + \tau_0 u_x^2] \, dx. \tag{5.13}$$

Let us compute the variation of E. Taking the time derivative under the integral, we find (remember that $\rho_0 = \rho_0(x)$ and τ_0 is constant),

$$\dot{E}(t) = \int_0^L [\rho_0 u_t u_{tt} + \tau_0 u_x u_{xt}] \, dx.$$

By an integration by parts we get

$$\int_0^L \tau_0 u_x u_{xt} \, dx = \tau_0[u_x(L,t) u_t(L,t) - u_x(0,t) u_t(0,t)] - \tau_0 \int_0^L u_t u_{xx} dx$$

whence

$$\dot{E}(t) = \int_0^L [\rho_0 u_{tt} - \tau_0 u_{xx}] u_t dx + \tau_0[u_x(L,t) u_t(L,t) - u_x(0,t) u_t(0,t)].$$

Using (5.10), we find:

$$\dot{E}(t) = \int_0^L \rho_0 f u_t \, dx + \tau_0[u_x(L,t) u_t(L,t) - u_x(0,t) u_t(0,t)]. \tag{5.14}$$

In particular, if $f = 0$ and u is constant at the end points 0 and L (therefore $u_t(L,t) = u_t(0,t) = 0$) we deduce $\dot{E}(t) = 0$. This implies

$$E(t) = E(0)$$

which expresses the *conservation of energy*.

5.3 The One-dimensional Wave Equation

5.3.1 Initial and boundary conditions

Equation (5.10) is called the *one-dimensional wave equation*. The coefficient c has the dimensions of a speed and in fact, we will shortly see that it represents the wave

propagation speed along the string. When $f \equiv 0$, the equation is *homogeneous* and the *superposition principle holds:* if u_1 and u_2 are solutions of

$$u_{tt} - c^2 u_{xx} = 0 \qquad (5.15)$$

and a, b are (real or complex) scalars, then $au_1 + bu_2$ is a solution as well. More generally, if $u_k(\mathbf{x},t)$ is a family of solutions depending on the parameter k (integer or real) and $g = g(k)$ is a function rapidly vanishing at infinity, then

$$\sum_{k=1}^{\infty} u_k(\mathbf{x},t) g(k) \quad \text{and} \quad \int_{-\infty}^{+\infty} u_k(\mathbf{x},t) g(k) \, dk$$

are still solutions of (5.15).

Suppose we are considering the space-time region $0 < x < L$, $0 < t < T$. In a well posed problem for the (one-dimensional) heat equation it is appropriate to assign the initial profile of the temperature, because of the presence of a first order time derivative, and a boundary condition at both ends $x = 0$ and $x = L$, because of the second order space derivative.

By analogy with the Cauchy problem for second order ordinary differential equations, the second order time derivative in (5.10) suggests that not only the initial profile of the string but the initial velocity has to be assigned as well.

Thus, our initial (or Cauchy) data are

$$u(x,0) = g(x), \qquad u_t(x,0) = h(x), \qquad x \in [0, L].$$

The boundary data are formally similar to those for the heat equation. Typically:

Dirichlet data describe the displacement of the end points of the string:

$$u(0,t) = a(t), \qquad u(L,t) = b(t), \qquad t > 0.$$

If $a(t) = b(t) \equiv 0$ (homogeneous data), both ends are fixed, with zero displacement.

Neumann data describe the applied (scalar) vertical tension at the end points. As in the derivation of the wave equation, we may model this tension by $\tau_0 u_x$ so that the Neumann conditions take the form

$$\tau_0 u_x(0,t) = a(t), \qquad \tau_0 u_x(L,t) = b(t), \qquad t > 0.$$

In the special case of homogeneous data, $a(t) = b(t) \equiv 0$, both ends of the string are attached to a frictionless sleeve and are free to move vertically.

Robin data describe a linear elastic attachment at the end points. One way to realize this type of boundary condition is to attach an end point to a linear spring[7] whose other end is fixed. This translates into assigning

$$\tau_0 u_x(0,t) = ku(0,t), \qquad \tau_0 u_x(L,t) = -ku(L,t), \qquad t > 0,$$

where k (positive) is the elastic constant of the spring.

---

[7] Which obeys Hooke's law: the strain is a linear function of the stress.

In several concrete situations, *mixed conditions* have to be assigned. For instance, Robin data at $x = 0$ and Dirichlet data at $x = L$.

Global Cauchy problem. We may think of a string of infinite length and assign only the initial data

$$u(x, 0) = g(x), \qquad u_t(x, 0) = h(x), \qquad x \in \mathbb{R}.$$

Although physically unrealistic, it turns out that the solution of the global Cauchy problem is of fundamental importance. We shall solve it in Section 5.4.

Under reasonable assumptions on the data, the above problems are well posed. In the next section we use separation of variables to show it for a Cauchy-Dirichlet problem.

Remark 5.1. Other kinds of problems for the wave equation are the so called *Goursat problem* and the *characteristic Cauchy problem*. Some examples are given in Problems 5.9, 5.10.

5.3.2 Separation of variables

Suppose that the vibration of a violin chord is modelled by the following Cauchy-Dirichlet problem

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0 & 0 < x < L, t > 0 \\ u(0, t) = u(L, t) = 0 & t \geq 0 \\ u(x, 0) = g(x), u_t(x, 0) = h(x) & 0 \leq x \leq L \end{cases} \tag{5.16}$$

where $c^2 = \tau_0/\rho_0$ is constant.

We want to check whether this problem is *well posed*, that is, whether a solution exists, is unique and it is stable (i.e. it depends "continuously" on the data g and h). For the time being we proceed formally, without worrying too much about the correct hypotheses on g and h and the regularity of u.

• *Existence*. Since the boundary conditions are homogeneous[8], we try to construct a solution using separation of variables.

Step **1**. We start looking for solutions of the form

$$U(x, t) = w(t) v(x)$$

with $v(0) = v(L) = 0$. Inserting U into the wave equation we find

$$0 = U_{tt} - c^2 U_{xx} = w''(t) v(x) - c^2 w(t) v''(x)$$

or, separating the variables,

$$\frac{1}{c^2} \frac{w''(t)}{w(t)} = \frac{v''(x)}{v(x)}. \tag{5.17}$$

---

[8] Remember that this is essential for using separation of variables.

We have reached a familiar situation: (5.17) is an identity between two functions, one depending on t only and the other one depending on x only. Therefore the two sides of (5.17) must be both equal to the same constant, say λ. Thus, we are lead to the equation

$$w''(t) - \lambda c^2 w(t) = 0 \tag{5.18}$$

and to the *eigenvalue problem*

$$v''(x) - \lambda v(x) = 0 \tag{5.19}$$

$$v(0) = v(L) = 0. \tag{5.20}$$

Step **2**. Solution of the eigenvalue problem. There are three possibilities for the general integral of (5.19).

a) If $\lambda = 0$, then $v(x) = A + Bx$ and (5.20) imply $A = B = 0$.

b) If $\lambda = \mu^2 > 0$, then $v(x) = Ae^{-\mu x} + Be^{\mu x}$ and again (5.20) imply $A = B = 0$.

c) If $\lambda = -\mu^2 < 0$, then $v(x) = A\sin\mu x + B\cos\mu x$. From (5.20) we get

$$v(0) = B = 0$$
$$v(1) = A\sin\mu L + B\cos\mu L = 0$$

whence

$$A \text{ arbitrary, } B = 0, \ \mu L = m\pi, \ m = 1, 2, \dots .$$

Thus, in case c) only we find non trivial solutions, of the form

$$v_m(x) = A_m \sin\mu_m x, \qquad \mu_m = \frac{m\pi}{L}. \tag{5.21}$$

Step **3**. Insert $\lambda = -\mu_m^2 = -m^2\pi^2/L^2$ into (5.18). Then, the general solution is

$$w_m(t) = C_m \cos(\mu_m ct) + D_m \sin(\mu_m ct). \tag{5.22}$$

From (5.21) and (5.22) we construct the family of solutions

$$U_m(x, t) = [a_m \cos(\mu_m ct) + b_m \sin(\mu_m ct)] \sin\mu_m x, \qquad m = 1, 2, \dots$$

where a_m and b_m are arbitrary constants.

U_m is called the $m^{th}-$**normal mode** of vibration or $m^{th} - harmonic,$ and is a *standing wave* with frequency $m/2L$. The first harmonic and its frequency $1/2L$, the lowest possible, are said to be *fundamental*. All the other frequencies are *integral multiples* of the fundamental one. Because of this reason it seems that a violin chord produces good quality tones, pleasant to the ear (this is not so, for instance, for a vibrating membrane like a drum, as we will see shortly).

Step **4**. If the initial conditions are

$$u(x, 0) = a_m \sin\mu_m x \qquad u_t(x, 0) = cb_m\mu_m \sin\mu_m x$$

then the solution of our problem is exactly U_m and the string vibrates at its mth−mode. In general, the solution is constructed by superposing the harmonics U_m through the formula

$$u(x,t) = \sum_{m=1}^{\infty} [a_m \cos(\mu_m ct) + b_m \sin(\mu_m ct)] \sin \mu_m x, \qquad (5.23)$$

where the coefficients a_m and b_m have to be chosen such that the initial conditions

$$u(x,0) = \sum_{m=1}^{\infty} a_m \sin \mu_m x = g(x) \qquad (5.24)$$

and

$$u_t(x,0) = \sum_{m=1}^{\infty} c\mu_m b_m \sin \mu_m x = h(x) \qquad (5.25)$$

are satisfied, for $0 \le x \le L$.

Looking at (5.24) and (5.25), it is natural to assume that both g and h have an expansion in Fourier sine series in the interval $[0, L]$. Let

$$\hat{g}_m = \frac{2}{L} \int_0^L g(x) \sin \mu_m x \, dx \quad \text{and} \quad \hat{h}_m = \frac{2}{L} \int_0^L h(x) \sin \mu_m x \, dx$$

be the Fourier sine coefficients of g and h. If we choose

$$a_m = \hat{g}_m, \qquad b_m = \frac{\hat{h}_m}{\mu_m c}, \qquad (5.26)$$

then (5.23) becomes

$$u(x,t) = \sum_{m=1}^{\infty} \left[\hat{g}_m \cos(\mu_m ct) + \frac{\hat{h}_m}{\mu_m c} \sin(\mu_m ct) \right] \sin \mu_m x \qquad (5.27)$$

and satisfies (5.24) and (5.25).

Although every U_m is a smooth solution of the wave equation, in principle (5.27) is only a formal solution, unless we may differentiate term by term twice with respect to both x and t, obtaining

$$(\partial_{tt} - c^2 \partial_{xx}^2) u(x,t) = \sum_{m=1}^{\infty} (\partial_{tt} - c^2 \partial_{xx}^2) U_m(x,t) = 0. \qquad (5.28)$$

This is possible if $\hat{g}_m$ and $\hat{h}_m$ vanish sufficiently fast as $m \to +\infty$. In fact, differentiating term by term twice, we have

$$u_{xx}(x,t) = -\sum_{m=1}^{\infty} \left[\mu_m^2 \hat{g}_m \cos(\mu_m ct) + \frac{\mu_m \hat{h}_m}{c} \sin(\mu_m ct) \right] \sin \mu_m x \qquad (5.29)$$

and

$$u_{tt}(x,t) = -\sum_{m=1}^{\infty} \left[\mu_m^2 \hat{g}_m c^2 \cos(\mu_m ct) + \mu_m \hat{h}_m c \sin(\mu_m ct) \right] \sin \mu_m x. \qquad (5.30)$$

Thus, if, for instance,

$$|\hat{g}_m| \leq \frac{C}{m^4} \quad \text{and} \quad \left| \hat{h}_m \right| \leq \frac{C}{m^3}, \qquad (5.31)$$

then

$$\left| \mu_m^2 \hat{g}_m \cos(\mu_m ct) \right| \leq \frac{C\pi^2}{L^2 m^2}, \quad \text{and} \quad \left| \mu_m \hat{h}_m c \sin(\mu_m ct) \right| \leq \frac{cC}{Lm^2}$$

so that, by the Weierstrass test, the series in (5.29), (5.30) converge uniformly in $[0, L] \times [0, +\infty)$. Since also the series (5.27) is clearly uniformly convergent in $[0, L] \times [0, +\infty)$, differentiation term by term is allowed and u is a C^2 solution of the wave equation.

Under which assumptions on g and h do the (5.31) hold?

Let $g \in C^4([0, L])$, $h \in C^3([0, L])$ and assume the following compatibility conditions:

$$g(0) = g(L) = g''(0) = g''(L) = 0$$
$$h(0) = h(L) = 0.$$

Then (5.31) hold[9].

Moreover, under the same assumptions, it is not difficult to check that

$$u(y, t) \to g(x), \ u_t(y, t) \to h(x), \qquad \text{as } (y, t) \to (x, 0) \qquad (5.32)$$

for every $x \in [0, L]$ and we conclude that u is a smooth solution of (5.16).

• *Uniqueness.* To show that (5.27) is the unique solution of problem (5.16), we use conservation of energy. Let u and v be solutions of (5.16). Then $w = u - v$ is a solution of the same problem with zero initial and boundary data. We want to show that $w \equiv 0$.

Formula (5.13) gives, for the total mechanical energy,

$$E(t) = E_{cin}(t) + E_{pot}(t) = \frac{1}{2} \int_0^L [\rho_0 w_t^2 + \tau_0 w_x^2] \, dx$$

---

[9] It is an exercise on integration by parts. For instance, if $f \in C^4([0, L])$ and $f(0) = f(L) = f''(0) = f''(L) = 0$, then, integrating by parts four times, we have

$$\hat{f}_m = \int_0^L f(x) \sin\left(\frac{m\pi}{L}\right) dx = \frac{1}{m^4} \int_0^L f^{(4)}(x) \sin\left(\frac{m\pi}{L}\right) dx$$

and

$$\left| \hat{f}_m \right| \leq \max \left| f^{(4)} \right| \frac{L}{m^4}.$$

and in our case we have
$$\dot{E}\left(t\right)=0$$
since $f=0$ and $w_t\left(L,t\right)=w_t\left(0,t\right)=0$, whence
$$E\left(t\right)=E\left(0\right)$$
for every $t\geq0$. Since, in particular, $w_t\left(x,0\right)=w_x\left(x,0\right)=0$, we have
$$E\left(t\right)=E\left(0\right)=0$$
for every $t\geq0$. On the other hand, $E_{cin}\left(t\right)\geq0$, $E_{pot}\left(t\right)\geq0$, so that we deduce
$$E_{cin}\left(t\right)=0,\ E_{pot}\left(t\right)=0$$
which force $w_t=w_x=0$. Therefore w is constant and since $w\left(x,0\right)=0$, we conclude that $w\left(x,t\right)=0$ for every $t\geq0$.

• *Stability.* We want to show that if the data are slightly perturbed, the corresponding solutions change only a little. Clearly, we need to establish how we intend to measure the distance for the data and for the corresponding solutions. For the initial data, we use the *least square distance*, given by[10]

$$\|g_1-g_2\|_0=\left(\int_0^L\left|g_1\left(x\right)-g_2\left(x\right)\right|^2dx\right)^{1/2}.$$

For functions depending also on time, we define

$$\|u-v\|_{0,\infty}=\sup_{t>0}\left(\int_0^L\left|u\left(x,t\right)-v\left(x,t\right)\right|^2dx\right)^{1/2}$$

which measures the maximum in time of the *least squares distance* in space.

Now, let u_1 and u_2 be solutions of problem (5.16) corresponding to the data g_1, h_1 and g_2, h_2, respectively. Their difference $w=u_1-u_2$ is a solution of the same problem with Cauchy data $g=g_1-g_2$ and $h=h_1-h_2$. From (5.27) we know that

$$w\left(x,t\right)=\sum_{m=1}^{\infty}\left[\hat{g}_m\cos(\mu_mct)+\frac{\hat{h}_m}{\mu_mc}\sin(\mu_mct)\right]\sin\mu_mx.$$

From Parseval's identity[11] and the elementary inequality $\left(a+b\right)^2\leq2(a^2+b^2)$, $a,b\in\mathbb{R}$, we can write

$$\int_0^L\left|w\left(x,t\right)\right|^2dx=\frac{L}{2}\sum_{m=1}^{\infty}\left[\hat{g}_m\cos(\mu_mct)+\frac{\hat{h}_m}{\mu_mc}\sin(\mu_mct)\right]^2$$

$$\leq L\sum_{m=1}^{\infty}\left[\hat{g}_m^2+\left(\frac{\hat{h}_m}{\mu_mc}\right)^2\right].$$

---

[10] The symbol $\|g\|$ denotes a *norm* of g. See Chapter 6.
[11] Appendix A.

Since $\mu_m \geq \pi/L$, using Parseval's equality again, we obtain

$$\int_0^L |w(x,t)|^2 \, dx \leq L \max\left\{1, \left(\frac{L}{\pi c}\right)^2\right\} \sum_{m=1}^{\infty} \left[\hat{g}_m^2 + \hat{h}_m^2\right]$$

$$= 2 \max\left\{1, \left(\frac{L}{\pi c}\right)^2\right\} \left[\|g\|_0^2 + \|h\|_0^2\right]$$

whence the stability estimate

$$\|u_1 - u_2\|_{0,\infty}^2 \leq 2 \max\left\{1, \left(\frac{L}{\pi c}\right)^2\right\} \left[\|g_1 - g_2\|_0^2 + \|h_1 - h_2\|_0^2\right]. \qquad (5.33)$$

Thus, "close" data produce "close" solutions.

Remark 5.2. From (5.27), the chord vibration is given by the superposition of harmonics corresponding to the non-zero Fourier coefficients of the initial data. The complex of such harmonics determines a particular feature of the emitted sound, known as the *timbre*, a sort of signature of the musical instrument!

Remark 5.3. The hypotheses we have made on g and h are unnaturally restrictive. For example, if we pluck a violin chord at a point, the initial profile is continuous but has a corner at that point and cannot be even C^1. A physically realistic assumption for the initial profile g is *continuity*.

Similarly, if we are willing to model the vibration of a chord set into motion by a strike of a little hammer, we should allow discontinuity in the initial velocity. Thus it is realistic to assume h *bounded*.

Under these weak hypotheses the separation of variables method does not work. On the other hand, we have already faced a similar situation in Chapter 4, where the necessity to admit discontinuous solutions of a conservation law has lead to a more general and flexible formulation of the initial value problem. Also for the wave equation it is possible to introduce suitable *weak* formulations of the various initial-boundary value problems, in order to include realistic initial data and solutions with a low degree of regularity. A first attempt is shown in subsection 5.4.2. A weak formulation more suitable for numerical methods is treated in Chapter 9.

5.4 The d'Alembert Formula

5.4.1 The homogeneous equation

In this section we establish the celebrated formula of d'Alembert for the solution of the following global Cauchy problem:

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0 & x \in \mathbb{R}, \, t > 0 \\ u(x,0) = g(x), \quad u_t(x,0) = h(x) & x \in \mathbb{R}. \end{cases} \qquad (5.34)$$

To find the solution, we first factorize the wave equation in the following way:

$$(\partial_t - c\partial_x)(\partial_t + c\partial_x)u = 0. \tag{5.35}$$

Now, let

$$v = u_t + cu_x. \tag{5.36}$$

Then v solves the linear transport equation

$$v_t - cv_x = 0$$

whence

$$v(x,t) = \psi(x + ct)$$

where ψ is a differentiable arbitrary function. From (5.36) we have

$$u_t + cu_x = \psi(x + ct)$$

and formula (4.10) in subsection 4.2.2 yields

$$u(x,t) = \int_0^t \psi(x - c(t - s) + cs)\ ds + \varphi(x - ct),$$

where φ is another arbitrary differentiable function.

Letting $x - ct + 2cs = y$, we find

$$u(x,t) = \frac{1}{2c}\int_{x-ct}^{x+ct} \psi(y)\, dy + \varphi(x - ct). \tag{5.37}$$

To determine ψ and φ we impose the initial conditions:

$$u(x,0) = \varphi(x) = g(x) \tag{5.38}$$

and

$$u_t(x,0) = \psi(x) - c\varphi'(x) = h(x)$$

whence

$$\psi(x) = h(x) + cg'(x). \tag{5.39}$$

Inserting (5.39) and (5.38) into (5.37) we get:

$$u(x,t) = \frac{1}{2c}\int_{x-ct}^{x+ct} [h(y) + cg'(y)]\ dy + g(x - ct)$$

$$= \frac{1}{2c}\int_{x-ct}^{x+ct} h(y)\, dy + \frac{1}{2}[g(x + ct) - g(x - ct)] + g(x - ct)$$

and finally the **d'Alembert** formula

$$u(x,t) = \frac{1}{2}[g(x + ct) + g(x - ct)] + \frac{1}{2c}\int_{x-ct}^{x+ct} h(y)\, dy. \tag{5.40}$$

If $g \in C^2(\mathbb{R})$ and $h \in C^1(\mathbb{R})$, formula (5.40) defines a C^2–solution in the half-plane $\mathbb{R} \times [0, +\infty)$. On the other hand, a C^2–solution u in $\mathbb{R} \times [0, +\infty)$ has to be given by (5.40), just because of the procedure we have used to solve the Cauchy problem. Thus the solution is *unique*. Observe however, that *no regularizing effect* takes place here: the solution u remains no more than C^2 for any $t > 0$. Thus, there is a striking difference with diffusion phenomena, governed by the heat equation.

Furthermore, let u_1 and u_2 be the solutions corresponding to the data g_1, h_1 and g_2, h_2, respectively. Then, the d'Alembert formula for $u_1 - u_2$ yields, for every $x \in \mathbb{R}$ and $t \in [0, T]$,

$$|u_1(x,t) - u_2(x,t)| \leq \|g_1 - g_2\|_\infty + T\|h_1 - h_2\|_\infty$$

where

$$\|g_1 - g_2\|_\infty = \sup_{x \in \mathbb{R}} |g_1(x) - g_2(x)|, \qquad \|h_1 - h_2\|_\infty = \sup_{x \in \mathbb{R}} |h_1(x) - h_2(x)|.$$

Therefore, we have stability in *pointwise uniform sense*, at least for finite time.

Rearranging the terms in (5.40), we may write u in the form[12]

$$u(x,t) = F(x + ct) + G(x - ct) \tag{5.41}$$

which gives u as a *superposition of two progressive waves moving at constant speed c in the negative and positive $x - direction$*, respectively. Thus, these waves are not dispersive.

The two terms in (5.41) are respectively constant along the two families of straight lines γ^+ and γ^- given by

$$x + ct = \text{constant}, \qquad x - ct = \text{constant}.$$

These lines are called *characteristics*[13] and carry important information, as we will see in the next subsection.

An interesting consequence of (5.41) comes from looking at figure 5.4. Consider the *characteristic parallelogram* with vertices at the point A, B, C, D. From (5.41) we have

$$F(A) = F(C), \quad G(A) = G(B)$$
$$F(D) = F(B), \quad G(D) = G(C).$$

---

[12] For instance:

$$F(x + ct) = \frac{1}{2}g(x + ct) + \frac{1}{2c}\int_0^{x+ct} h(y)\,dy$$

and

$$G(x - ct) = \frac{1}{2}g(x - ct) + \frac{1}{2c}\int_{x-ct}^0 h(y)\,dy.$$

[13] In fact they are the *characteristics* for the two first order factors in the factorization (5.35).
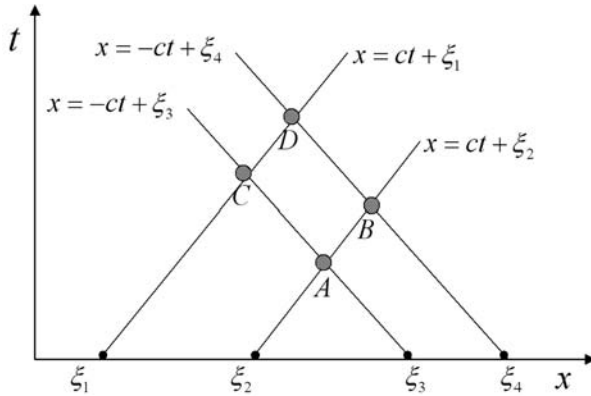
Fig. 5.4. Characteristic parallelogram

Summing these relations we get

$$[F(A) + G(A)] + [F(D) + G(D)] = [F(C) + G(C)] + [F(B) + G(B)]$$

which is equivalent to

$$u(A) + u(D) = u(C) + u(B). \tag{5.42}$$

Thus, knowing u at three points of a characteristic parallelogram, we can compute u at the fourth one.

From d'Alembert formula it follows that the value of u at the point (x, t) depends on the values of g at the points $x - ct$ e $x + ct$ and on the values of h over the whole interval $[x - ct, x + ct]$. This interval is called **domain of dependence of** (x, t) (Fig. 5.5).

From a different perspective, the values of g and h at a point z affect the value of u at the points (x, t) in the sector

$$z - ct \leq x \leq z + ct,$$

which is called **range of influence** of z (Fig. 5.5). This entails that a disturbance initially localized at z is not felt at a point x until time

$$t = \frac{|x - z|}{c}.$$

Remark 5.4. Differentiating the last term in (5.40) with respect to time we get:

$$\frac{\partial}{\partial t} \frac{1}{2c} \int_{x-ct}^{x+ct} h(y)\, dy = \frac{1}{2c} [ch(x + ct) - (-c)h(x - ct)]$$

$$= \frac{1}{2} [h(x + ct) + h(x - ct)]$$
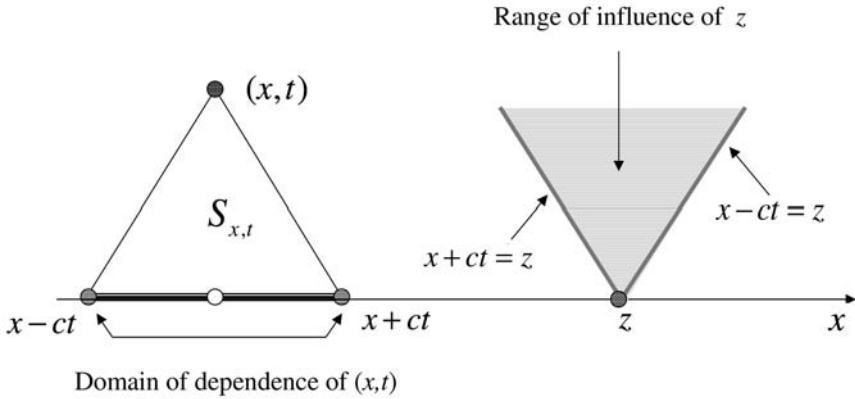
Fig. 5.5. Domain of dependence and range of influence

which has the form of the first term with g replaced by h. It follows that if w_h denotes the solution of the problem

$$\begin{cases} w_{tt} - c^2 w_{xx} = 0 & x \in \mathbb{R}, \, t > 0 \\ w(x,0) = 0, \, w_t(x,0) = h(x). & x \in \mathbb{R} \end{cases} \tag{5.43}$$

then, d'Alembert formula can be written in the form

$$u(x,t) = \frac{\partial}{\partial t} w_g(x,t) + w_h(x,t). \tag{5.44}$$

Actually, (5.44), can be established without reference to d'Alembert formula, as we will see later.

5.4.2 Generalized solutions and propagation of singularities

In Remark 5.3 we have emphasized the necessity of a weak formulation to include physically realistic data. On the other hand, observe that d'Alembert formula makes perfect sense even for g *continuous* and h *bounded.* The question is in which sense the resulting function satisfies the wave equation, since, in principle, it is not even differentiable, only continuous. There are several ways to weaken the notion of solution to include this case; here, for instance, we mimic what we did for conservation laws.

Assuming for the moment that u is a smooth solution of the global Cauchy problem, we multiply the wave equation by a C^2-*test function* v, defined in $\mathbb{R} \times [0, +\infty)$ and compactly supported. Integrating over $\mathbb{R} \times [0, +\infty)$ we obtain

$$\int_0^\infty \int_\mathbb{R} [u_{tt} - c^2 u_{xx}] v \, dx dt = 0.$$
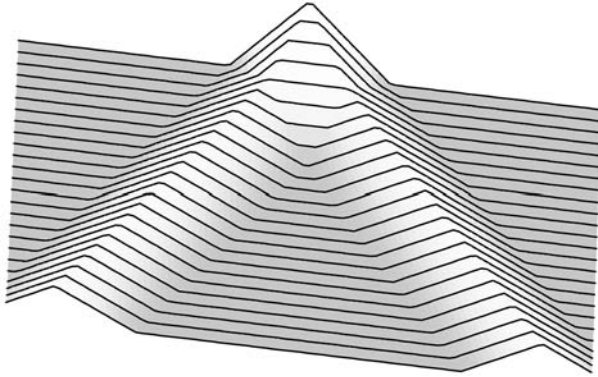
Fig. 5.6. Chord plucked at the origin $(c = 1)$

Now we integrate by parts both terms twice, to transfer all the derivatives from u to v. This yields, being v zero outside a compact subset of $\mathbb{R} \times [0, +\infty)$,

$$\int_0^\infty \int_{\mathbb{R}} c^2 u_{xx} v \; dxdt = \int_0^\infty \int_{\mathbb{R}} c^2 u v_{xx} \; dxdt$$

and

$$\int_0^\infty \int_{\mathbb{R}} u_{tt} v \; dxdt = - \int_{\mathbb{R}} u_t(x,0) v(x,0) \, dx - \int_0^\infty \int_{\mathbb{R}} u_t v_t \; dxdt$$

$$= - \int_{\mathbb{R}} [u_t(x,0) v(x,0) - u(x,0) v_t(x,0)] \, dx + \int_0^\infty \int_{\mathbb{R}} u v_{tt} dxdt.$$

Using the Cauchy data $u(x,0) = g(x)$ and $u_t(x,0) = h(x)$, we arrive to the integral equation

$$\int_0^\infty \int_{\mathbb{R}} u[v_{tt} - c^2 v_{xx}] \; dxdt - \int_{\mathbb{R}} [h(x) v(x,0) - g(x) v_t(x,0)] \, dx = 0. \qquad (5.45)$$

Note that (5.45) makes perfect sense for u continuous, g continuous and h bounded, only. Conversely, if u is a C^2 function that satisfies (5.45) **for every** test function v, then it turns out[14] that u is a solution of problem (5.34).

Thus we may adopt the following definition.

Definition 5.1. *Let* $g \in C(\mathbb{R})$ *and* h *be bounded in* $\mathbb{R}$. *We say that* $u \in C(\mathbb{R} \times [0, +\infty))$ *is a **generalized** solution of problem (5.34) if (5.45) holds **for every** test function* v.

If g is continuous and h is bounded, it can be shown that formula (5.41) constitutes precisely a generalized solution.

---

[14] Check it.

Figure 5.6 shows the wave propagation along a chord of infinite length, plucked at the origin and originally at rest, modelled by the solution of the problem

$$\begin{cases} u_{tt} - u_{xx} = 0 & x \in \mathbb{R}, \, t > 0 \\ u(x,0) = g(x), \, u_t(x,0) = 0 & x \in \mathbb{R} \end{cases}$$

where g has a triangular profile. As we see, this generalized solution displays lines of discontinuities of the first derivatives, while outside these lines it is smooth.

We want to show that these lines are *characteristics*. More generally, consider a region $G \subset \mathbb{R} \times (0, +\infty)$, divided into two domains $G^{(1)}$ e $G^{(2)}$ by a smooth curve Γ of equation $x = s(t)$, as in figure 5.7. Let

$$\boldsymbol{\nu} = \nu_1 \mathbf{i} + \nu_2 \mathbf{j} = \frac{1}{\sqrt{1 + (\dot{s}(t))^2}} \left(-\mathbf{i} + \dot{s}(t) \mathbf{j} \right) \tag{5.46}$$

be the unit normal to Γ, pointing inward to $G^{(1)}$.

Given any function f defined in G, we denote by

$$f^{(1)} \text{ and } f^{(2)}$$

its restriction to the closure of $G^{(1)}$ and $G^{(2)}$, respectively, and we use the symbol

$$[f(s(t),t)] = f^{(1)}(s(t),t) - f^{(2)}(s(t),t).$$

for the jump of f across Γ, or simply $[f]$ when there is no risk of confusion.

Now, let u be a generalized solution of our Cauchy problem, of class C^2 both in the closure[15] of $G^{(1)}$ and $G^{(2)}$, whose first derivatives undergo a jump discontinuity on Γ. We want to prove that:

Proposition 5.1. Γ *is a characteristic.*

Proof. First of all observe that, from our hypotheses, we have $[u] = 0$ and $[u_x], [u_t] \neq 0$. Moreover, the jumps $[u_x]$ and $[u_t]$ are continuous along Γ.

By analogy with conservation laws, we expect that the integral formulation (5.45) should imply a sort of Rankine-Hugoniot condition, relating the jumps of the derivatives with the slope of Γ and expressing the balance of linear momentum across Γ.

In fact, let v be a test function with compact support in G. Inserting v into (5.45), we can write

$$0 = \int_G \left(c^2 u v_{xx} - u v_{tt} \right) \, dx dt = \int_{G^{(2)}} (...) \, dx dt + \int_{G^{(1)}} (...) \, dx dt. \tag{5.47}$$

---

[15] That is, the first and second derivatives of u extend continuously up to Γ, from both sides, separately.
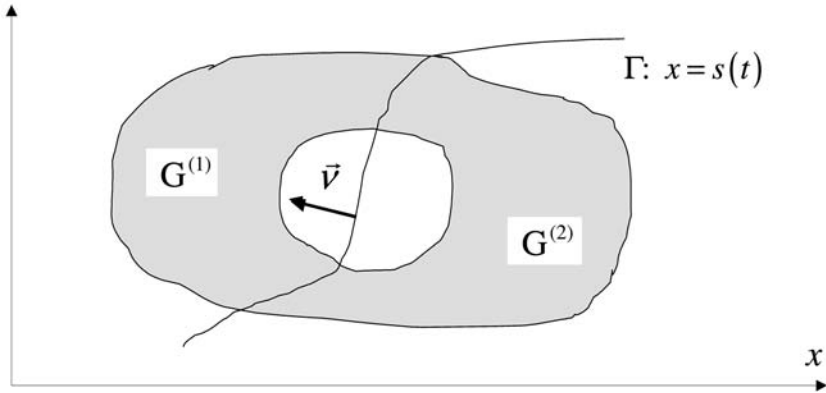
Fig. 5.7. Line of discontinuity of first derivatives

Integrating by parts, since $v = 0$ on ∂G (dl denotes arc length on Γ),

$$\int_{G^{(2)}} \left(c^2 u^{(2)} v_{xx} - u^{(2)} v_{tt} \right) dx dt$$

$$= \int_{\Gamma} (\nu_1 c^2 u^{(2)} v_x - \nu_2 u^{(2)} v_t) \, dl - \int_{G^{(2)}} (c_x^2 u^{(2)} v_x - u_t^{(2)} v_t) \, dx dt$$

$$= \int_{\Gamma} \left(\nu_1 c^2 v_x - \nu_2 v_t \right) u^{(2)} \, dl - \int_{\Gamma} (\nu_1 c^2 u_x^{(2)} - \nu_2 u_t^{(2)}) v \, dl,$$

because $\int_{G^{(2)}} [c^2 u_{xx}^{(2)} - u_{tt}^{(2)}] v \, dx dt = 0$. Similarly,

$$\int_{G^{(1)}} (c^2 u^{(1)} v_{xx} - u^{(1)} v_{tt}) \, dx dt$$

$$= -\int_{\Gamma} \left(\nu_1 c^2 v_x - \nu_2 v_t \right) u^{(1)} \, dl + \int_{\Gamma} (\nu_1 c^2 u_x^{(1)} - \nu_2 u_t^{(1)}) v \, dl,$$

because $\int_{G^{(2)}} [c^2 u_{xx}^{(1)} - u_{tt}^{(1)}] v \, dx dt = 0$ as well.

Thus, since $[u] = 0$ on Γ, or more explicitly $[u(s(t), t)] \equiv 0$, (5.47) yields

$$\int_{\Gamma} \left(c^2 [u_x] \nu_1 - [u_t] \nu_2 \right) v \, dl = 0.$$

Due to the arbitrariness of v and the continuity of $[u_x]$ and $[u_t]$ on Γ, we deduce

$$c^2 [u_x] \nu_1 - [u_t] \nu_2 = 0, \qquad \text{on } \Gamma,$$

or, recalling (5.46),

$$\dot{s} = -c^2 \frac{[u_x]}{[u_t]} \qquad \text{on } \Gamma, \tag{5.48}$$

which is the analogue of the Rankine-Hugoniot condition for conservation laws.

On the other hand, differentiating $[u\left(s\left(t\right),t\right)] \equiv 0$ we obtain

$$\frac{d}{dt}\left[u\left(s\left(t\right),t\right)\right] = [u_x\left(s\left(t\right),t\right)]\dot{s}\left(t\right) + [u_t\left(s\left(t\right),t\right)] \equiv 0$$

or

$$\dot{s} = -\frac{[u_t]}{[u_x]} \qquad \text{on } \Gamma. \tag{5.49}$$

Equations (5.48) and (5.49) entail

$$\dot{s}\left(t\right) = \pm c$$

which yields

$$s\left(t\right) = \pm ct + \text{constant}$$

showing that Γ is a characteristic. $\square$

5.4.3 The fundamental solution

It is rather instructive to solve the global Cauchy problem with $g \equiv 0$ and a special h: *the Dirac delta at a point* ξ, that is $h\left(x\right) = \delta(x - \xi)$. For instance, this models the vibrations of a violin string generated by a *unit impulse localized at* ξ (a strike of a sharp hammer). The corresponding solution is called **fundamental solution** and plays the same role of the fundamental solution for the diffusion equation.

   Certainly, the *Dirac delta* is a quite unusual data, out of reach of the theory we have developed so far. Therefore, we proceed formally.

   Thus, let $K = K\left(x,\xi,t\right)$ denote our fundamental solution and apply d'Alembert formula; we find

$$K\left(x,\xi,t\right) = \frac{1}{2c}\int_{x-ct}^{x+ct} \delta\left(y - \xi\right) dy \tag{5.50}$$

which at first glance looks like a mathematical *UFO*.

   To get a more explicit formula, we first compute $\int_{-\infty}^{x} \delta\left(y\right) dy$. To do it, recall that (subsection 2.3.3), if $\mathcal{H}$ is the *Heaviside* function and

$$I_\varepsilon\left(y\right) = \frac{\mathcal{H}\left(y + \varepsilon\right) - \mathcal{H}\left(y - \varepsilon\right)}{2\varepsilon} = \begin{cases} \frac{1}{2\varepsilon} & -\varepsilon \le y < \varepsilon \\ 0 & \text{everywhere else} \end{cases} \tag{5.51}$$

is the unit impulse of extent ε, then $\lim_{\varepsilon \downarrow 0} I_\varepsilon\left(y\right) = \delta\left(y\right)$. Then it seems appropriate to compute $\int_{-\infty}^{x} \delta\left(y\right) dy$ by means of the formula

$$\int_{-\infty}^{x} \delta\left(y\right) dy = \lim_{\varepsilon \downarrow 0}\int_{-\infty}^{x} I_\varepsilon\left(y\right) dy.$$

Now, we have:

$$\int_{-\infty}^{x} I_\varepsilon\left(y\right) dy = \begin{cases} 0 & x \le -\varepsilon \\ \left(x + \varepsilon\right)/2\varepsilon & -\varepsilon < x < \varepsilon \\ 1 & x \ge \varepsilon. \end{cases}$$

Letting $\varepsilon \to 0$ we deduce that (the value at zero is irrelevant)

$$\int_{-\infty}^{x} \delta(y)\, dy = \mathcal{H}(x), \tag{5.52}$$

which actually is not surprising, if we remember that $\mathcal{H}' = \delta$. Everything works nicely.

Let us go back to our mathematical *UFO*, by now ... identified; we write

$$\int_{x-ct}^{x+ct} \delta(y - \xi)\, dy = \lim_{\varepsilon \downarrow 0} \int_{-\infty}^{x+ct} I_\varepsilon(y - \xi)\, dy - \lim_{\varepsilon \downarrow 0} \int_{-\infty}^{x-ct} I_\varepsilon(y - \xi)\, dy.$$

Then, using (5.50), (5.51) and (5.52), we conclude:

$$K(x, \xi, t) = \frac{1}{2c} \left\{ \mathcal{H}(x - \xi + ct) - \mathcal{H}(x - \xi - ct) \right\}. \tag{5.53}$$

Figure 5.8 shows the graph of $K(x, \xi, t)$, with $c = 1$
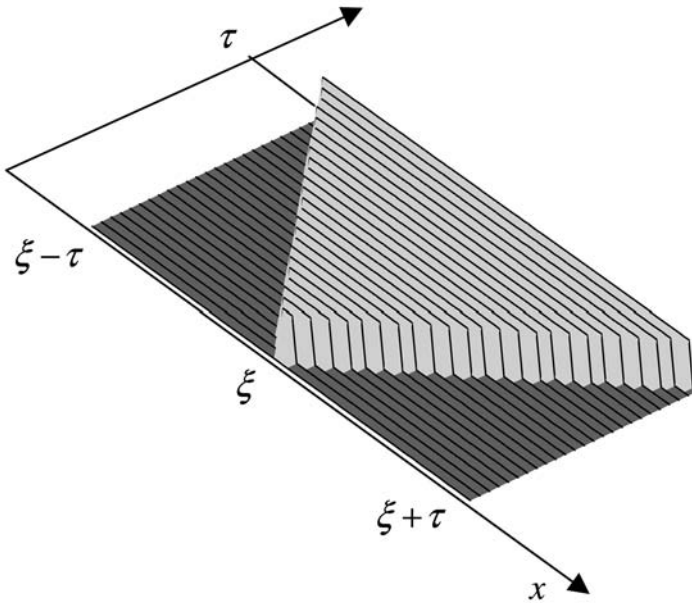
Fig. 5.8. The fundamental solution $K(x, \xi, t)$

Note how the initial discontinuity at $x = \xi$ propagates along the characteristics

$$x = \xi \pm t.$$

We have found the fundamental solution (5.53) through d'Alembert formula. Conversely, using the fundamental solution we may derive d'Alembert formula.

Namely, consider the solution w_h of the Cauchy problem (5.43), that is with data (see Remark 5.4)

$$w(x,0) = 0, \quad w_t(x,0) = h(x), \quad x \in \mathbb{R}.$$

We may write

$$h(x) = \int_{-\infty}^{+\infty} \delta(x - \xi) h(\xi) d\xi$$

looking at $h(x)$ as a superposition of impulses $\delta(x - \xi) h(\xi)$, concentrated at ξ. Then, we may construct w_h by superposing the solutions of the same problem with data $\delta(x - \xi) h(\xi)$ instead of h. But these solutions are given by

$$K(x, \xi, t) h(\xi)$$

and therefore we obtain

$$w_h(x,t) = \int_{-\infty}^{+\infty} K(x, \xi, t) h(\xi) d\xi.$$

More explicitly, from (5.53):

$$
\begin{aligned}
w_h(x,t) &= \frac{1}{2c} \int_{-\infty}^{+\infty} \{H(x - \xi + ct) - H(x - \xi - ct)\} h(\xi) d\xi \\
&= \frac{1}{2c} \int_{-\infty}^{x+ct} h(\xi) d\xi - \frac{1}{2c} \int_{-\infty}^{x-ct} h(\xi) d\xi \\
&= \frac{1}{2c} \int_{x-ct}^{x+ct} h(y) dy.
\end{aligned}
$$

At this point, (5.44) yields d'Alembert formula.

We shall use this method to construct the solution of the global Cauchy problem in dimension 3.

5.4.4 Non homogeneous equation. Duhamel's method

To solve the nonhomogeneous problem

$$
\begin{cases}
u_{tt} - c^2 u_{xx} = f(x,t) & x \in \mathbb{R}, \, t > 0 \\
u(x,0) = 0, \, u_t(x,0) = 0 & x \in \mathbb{R}.
\end{cases}
\tag{5.54}
$$

we use the Duhamel's method (see subsection 2.2.8). For $s \geq 0$ fixed, let $w = w(x,t;s)$ be the solution of problem

$$
\begin{cases}
w_{tt} - c^2 w_{xx} = 0 & x \in \mathbb{R}, \, t \geq s \\
w(x,s;s) = 0, \, w_t(x,s;s) = f(x,s) & x \in \mathbb{R}.
\end{cases}
\tag{5.55}
$$

Since the wave equation is invariant under (time) translations, from (5.40) we get

$$w\left(x,t;s\right) = \frac{1}{2c} \int_{x-c(t-s)}^{x+c(t-s)} f\left(y,s\right) dy.$$

Then, the solution of (5.54) is given by

$$u\left(x,t\right) = \int_0^t w\left(x,t;s\right)\ ds = \frac{1}{2c} \int_0^t ds \int_{x-c(t-s)}^{x+c(t-s)} f\left(y,s\right) dy.$$

In fact, $u\left(x,0\right) = 0$ and

$$u_t\left(x,t\right) = w\left(x,t;t\right) + \int_0^t w_t\left(x,t;s\right)\ ds = \int_0^t w_t\left(x,t;s\right)\ ds$$

since $w\left(x,t;t\right) = 0$. Thus $u_t\left(x,0\right) = 0$. Moreover,

$$u_{tt}\left(x,t\right) = w_t\left(x,t;t\right) + \int_0^t w_{tt}\left(x,t;s\right)\ ds = f\left(x,t\right) + \int_0^t w_{tt}\left(x,t;s\right)\ ds$$

and

$$u_{xx}\left(x,t\right) = \int_0^t w_{xx}\left(x,t;s\right)\ ds.$$

Therefore, since $w_{tt} - c^2 w_{xx} = 0$,

$$u_{tt}\left(x,t\right) - c^2 u_{xx}\left(x,t\right) = f\left(x,t\right) + \int_0^t w_{tt}\left(x,t;s\right)\ ds - c^2 \int_0^t w_{xx}\left(x,t;s\right)\ ds$$
$$= f\left(x,t\right).$$

Everything works and gives the *unique* solution in $C^2(\mathbb{R} \times [0,+\infty))$, under rather natural hypotheses on f: we require f and f_x be continuous in $\mathbb{R} \times [0,+\infty)$.

Finally note that the value of u at the point (x,t) depends on the values of the forcing term f *in all the triangular sector $S_{x,t}$* in figure 5.5.

5.4.5 Dissipation and dispersion

Dissipation and dispersion effects are quite important in wave propagation phenomena. Let us go back to our model for the vibrating string, assuming that its weight is negligible and that there are no external loads.

• *External damping.* External factors of dissipation like friction due to the medium may be included into the model through some empirical constitutive law. We may assume, for instance, a *linear law* of friction expressing a force proportional to the speed of vibration. Then, a force given by $-k\rho_0 u_t \Delta x \mathbf{j}$, where $k > 0$ is a damping constant, acts on the segment of string between x and $x + \Delta x$. The final equation takes the form

$$\rho_0 u_{tt} - \tau_0 u_{xx} + k\rho_0 u_t = 0. \tag{5.56}$$

For a string with fixed end points, the same calculations in subsection 5.2.2 yield

$$\dot{E}(t) = -\int_0^L k\rho_0 u_t^2 \, dx = -kE_{cin}(t) \leq 0 \qquad (5.57)$$

which shows a rate of energy dissipation proportional to the kinetic energy.

For equation (5.56), the usual initial-boundary value problems are still well posed under reasonable assumptions on the data. In particular, the uniqueness of the solution follows from (5.57), since $E(0) = 0$ implies $E(t) = 0$ for all $t > 0$.

• *Internal damping.* The derivation of the wave equation in subsection 5.2.1 leads to

$$\rho_0 u_{tt} = (\tau_{vert})_x$$

where τ_{vert} is the (scalar) vertical component of the tension. The hypothesis of vibrations of small amplitude corresponds to taking

$$\tau_{vert} \simeq \tau_0 u_x, \qquad (5.58)$$

where τ_0 is the (scalar) horizontal component of the tension. In other words, we assume that the vertical forces due to the tension at two end points of a string element are proportional to the relative displacement of these points. On the other hand, the string vibrations convert kinetic energy into heat, because of the friction among the particles. The amount of heat increases with the speed of vibration while, at the same time, the vertical tension decreases. Thus, the vertical tension depends not only on the relative displacements u_x, but also on how fast these displacements change with time[16]. Hence, we modify (5.58) by inserting a term proportional to u_{xt}:

$$\tau_{vert} = \tau u_x + \gamma u_{xt} \qquad (5.59)$$

where γ is a *positive* constant. The positivity of γ follows from the fact that energy dissipation lowers the vertical tension, so that the slope u_x decreases if $u_x > 0$ and increases if $u_x < 0$. Using the law (5.59) we derive the third order equation

$$\rho_0 u_{tt} - \tau u_{xx} - \gamma u_{xxt} = 0. \qquad (5.60)$$

In spite of the presence of the term u_{xxt}, the usual initial-boundary value problems are again well posed under reasonable assumptions on the data. In particular, uniqueness of the solution follows once again from dissipation of energy, since, in this case[17],

$$\dot{E}(t) = -\int_0^L \gamma\rho_0 u_{xt}^2 \leq 0.$$

• *Dispersion.* When the string is under the action of a vertical elastic restoring force proportional to u, the equation of motion becomes

$$u_{tt} - c^2 u_{xx} + \lambda u = 0 \qquad (\lambda > 0) \qquad (5.61)$$

---

[16] In the movie *The Legend of 1900* there is a spectacular demo of this phenomenon.
[17] Check it.

known as the *linearized Klein-Gordon equation*. To emphasize the effect of the zero order term λu, let us seek for *harmonic waves solutions* of the form

$$u(x,t) = Ae^{i(kx-\omega t)}.$$

Inserting u into (5.61) we find the *dispersion relation*

$$\omega^2 - c^2 k^2 = \lambda \qquad \Longrightarrow \qquad \omega(k) = \pm\sqrt{c^2 k^2 + \lambda}.$$

Thus, this waves are dispersive with phase and group velocities given respectively by

$$c_p(k) = \frac{\sqrt{c^2 k^2 + \lambda}}{|k|}, \quad c_g = \frac{d\omega}{dk} = \frac{c^2 |k|}{\sqrt{c^2 k^2 + \lambda}}.$$

Observe that $c_g < c_p$.

A wave packet solution can be obtained by an integration over all possible wave numbers k:

$$u(x,t) = \int_{-\infty}^{+\infty} A(k) e^{i[kx - \omega(k)t]} dk \tag{5.62}$$

where $A(k)$ is the Fourier transform of the initial condition:

$$A(k) = \int_{-\infty}^{+\infty} u(x,0) e^{-ikx} dx.$$

This entails that, even if the initial condition is *localized* inside a small interval, *all* the wavelength contribute to the value of u. Although we have seen in subsection 5.1.2 that we observe a decaying in amplitude of order $t^{-1/2}$ (see formula (5.7)), these dispersive waves do not dissipate energy. For example, if the ends of the string are fixed, the total mechanical energy is given by

$$E(t) = \frac{\rho_0}{2} \int_0^L \left(u_t^2 + c^2 u_x^2 + \lambda u^2 \right) dx$$

and one may check that $\dot{E}(t) = 0$, $t > 0$.

5.5 Second Order Linear Equations

5.5.1 Classification

To derive formula (5.41) we may use the characteristics in the following way. We change variables by setting

$$\xi = x + ct, \qquad \eta = x - ct \tag{5.63}$$

or

$$x = \frac{\xi + \eta}{2}, \qquad t = \frac{\xi - \eta}{2c}$$

and define

$$U\left(\xi, \eta\right) = u\left(\frac{\xi + \eta}{2}, \frac{\xi - \eta}{2c}\right).$$

Then

$$U_\xi = \frac{1}{2}u_x + \frac{1}{2c}u_t$$

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

$$U_{\xi\eta} = \frac{1}{4}u_{xx} - \frac{1}{4c}u_{xt} + \frac{1}{4c}u_{xt} - \frac{1}{4c^2}u_{tt} = 0.$$

The equation

$$U_{\xi\eta} = 0 \tag{5.64}$$

is called the *canonical* form of the wave equation; its solution is immediate:

$$U\left(\xi, \eta\right) = F\left(\xi\right) + G\left(\eta\right)$$

and going back to the original variables (5.41) follows.

Consider now a general equation of the form:

$$au_{tt} + 2bu_{xt} + cu_{xx} + du_t + eu_x + hu = f \tag{5.65}$$

with x, t varying, in general, in a domain Ω. We assume that the coefficients a, b, c, d, e, h, f are smooth functions[18] in Ω. The sum of second order terms

$$a\left(x, t\right)u_{tt} + 2b\left(x, t\right)u_{xt} + c\left(x, t\right)u_{xx} \tag{5.66}$$

is called **principal part** of equation (5.65) and determines the *type* of equation according to the following classification. Consider the algebraic equation

$$H\left(p, q\right) = ap^2 + 2bpq + cq^2 = 1 \qquad (a > 0). \tag{5.67}$$

in the plane p, q. If $b^2 - ac < 0$, (5.67) defines a hyperbola, if $b^2 - ac = 0$ a parabola and if $b^2 - ac < 0$ an ellipse. Accordingly, equation (5.65) is called:

   a) **hyperbolic** when $b^2 - ac < 0$,

   b) **parabolic** when $b^2 - ac = 0$,

   c) **elliptic** when $b^2 - ac > 0$.

Note that the quadratic form $H\left(p, q\right)$ is, in the three cases, *indefinite, nonnegative, positive*, respectively. In this form, the above classification extends to equations in any number of variables, as we shall see later on.

It may happen that a single equation is of different type in different subdomains. For instance, the *Tricomi* equation $xu_{tt} - u_{xx} = 0$ is hyperbolic in the half plane $x > 0$, parabolic on $x = 0$ and elliptic in the half plane $x < 0$.

Basically all the equations in two variables we have met so far are particular cases of (5.65). Specifically,

---

[18] E.g. C^2 functions.

- the *wave* equation

$$u_{tt} - c^2 u_{xx} = 0$$

is *hyperbolic:* $a(x,t) = 1$, $c(x,t) = -c^2$, and the other coefficients are zero;
- the *diffusion* equation

$$u_t - D u_{xx} = 0$$

is *parabolic:* $c(x,t) = -D$, $d(x,t) = 1$, and the other coefficients are zero;
- *Laplace* equation (using y instead of t)

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

is *elliptic:* $a(x,y) = 1$, $c(x,y) = 1$, and the other coefficients are zero.

May we reduce to a canonical form, similar to (5.64), the diffusion and the Laplace equation? Let us briefly examine why the change of variables (5.63) works for the wave equation. Decompose the wave operator as follows

$$\partial_{tt} - c^2 \partial_{xx} = (\partial_t + c\partial_x)(\partial_t - c\partial_x). \qquad (5.68)$$

If we introduce the vectors $\mathbf{v} = (c, 1)$ and $\mathbf{w} = (-c, 1)$, then (5.68) can be written in the form

$$\partial_{tt} - c^2 \partial_{xx} = \partial_{\mathbf{v}} \partial_{\mathbf{w}}.$$

On the other hand, the characteristics

$$x + ct = 0, \quad x - ct = 0$$

of the two first order equations

$$\phi_t - c\phi_x = 0 \quad \text{and} \quad \psi_t + c\psi_t = 0,$$

corresponding to the two factors in (5.68), are straight lines in the direction of $\mathbf{w}$ and $\mathbf{v}$, respectively. The change of variables

$$\xi = \phi(x,t) = x + ct \qquad \eta = \psi(x,t) = x - ct$$

maps these straight lines into $\xi = 0$ and $\eta = 0$ and

$$\partial_\xi = \frac{1}{2c}(\partial_t + c\partial_x) = \frac{1}{2c}\partial_{\mathbf{v}}, \qquad \partial_\eta = \frac{1}{2c}(\partial_t - c\partial_x) = \frac{1}{2c}\partial_{\mathbf{w}}.$$

Thus, the wave operator is converted into a multiple of its canonical form:

$$\partial_{tt} - c^2 \partial_{xx} = \partial_{\mathbf{v}} \partial_{\mathbf{w}} = 4c^2 \partial_{\xi\eta}.$$

Once the characteristics are known, the change of variables (5.63) reduces the wave equation to the form (5.64).

Proceeding in the same way, for the diffusion operator we would have

$$\partial_{xx} = \partial_x \partial_x.$$

Therefore we find only one family of characteristics, given by

$$t = \text{constant.}$$

Thus, no change of variables is necessary and the diffusion equation is already in its canonical form.

For the Laplace operator we find

$$\partial_{xx} + \partial_{yy} = (\partial_y + i\partial_x)(\partial_y - i\partial_x)$$

and there are two families of *complex* characteristics given by

$$\phi(x, y) = x + iy = \text{constant,} \qquad \psi(x, y) = x - iy = \text{constant.}$$

The change of variables

$$z = x + iy, \qquad \overline{z} = x - iy$$

leads to the equation

$$\partial_{z\overline{z}} U = 0$$

whose general solution is

$$U(z, \overline{z}) = F(z) + G(\overline{z}).$$

This formula may be considered as a characterization of the harmonic function in the complex plane.

It should be clear, however, that the characteristics for the diffusion and the Laplace equations do not play the same relevant role as they do for the wave equation.

5.5.2 Characteristics and canonical form

Let us go back to the equation in general form (5.65). Can we reduce to a canonical form its principal part? There are at least two substantial reasons to answer the question.

The first one is tied to the type of well posed problems associated with (5.65): which kind of data have to be assigned and where, in order to find a unique and stable solution? It turns out that hyperbolic, parabolic and elliptic equations share their well posed problems with their main prototypes: the wave, diffusion and Laplace equations, respectively. Also the choice of numerical methods depends very much on the type of problem to be solved.

The second reason comes from the different features the three types of equation exhibit. Hyperbolic equations model oscillatory phenomena with *finite speed of propagation of the disturbances*, while for parabolic equation, "information" travels with infinite speed. Finally, elliptic equations model stationary situations, with no evolution in time.

To obtain the canonical form of the principal part we try to apply the ideas at the end of the previous subsection. First of all, note that, if $a = c = 0$, the principal part is already in the form (5.64), so that we assume $a > 0$ (say). Now we decompose the differential operator in (5.66) into the product of two first order factors, as follows[19]:

$$a\partial_{tt} + 2b\partial_{xt} + c\partial_{xx} = a\left(\partial_t - \Lambda^+\partial_x\right)\left(\partial_t - \Lambda^-\partial x\right) \tag{5.69}$$

where

$$\Lambda^\pm = \frac{-b \pm \sqrt{b^2 - ac}}{a}.$$

Case 1: $b^2 - ac > 0$, the equation is **hyperbolic**. The two factors in (5.69) represent derivatives along the direction fields

$$\mathbf{v}\left(x, t\right) = \left(-\Lambda^+\left(x, t\right), 1\right) \quad \text{and} \quad \mathbf{w}\left(x, t\right) = \left(-\Lambda^-\left(x, t\right), 1\right)$$

respectively, so that we may write

$$a\partial_{tt} + 2b\partial_{xt} + c\partial_{xx} = a\partial_{\mathbf{v}}\partial_{\mathbf{w}}.$$

The vector fields $\mathbf{v}$ and $\mathbf{w}$ are tangent at any point to the characteristics

$$\phi\left(x, t\right) = k_1 \quad \text{and} \quad \psi\left(x, t\right) = k_2 \tag{5.70}$$

of the following *quasilinear first-order* equations

$$\phi_t - \Lambda^+\phi_x = 0 \quad \text{and} \quad \psi_t - \Lambda^-\psi_x = 0. \tag{5.71}$$

Note that we may write the two equations (5.71) in the compact form

$$av_t^2 + 2bv_xv_t + cv_x^2 = 0. \tag{5.72}$$

By analogy with the case of the wave equation, we expect that the change of variables

$$\xi = \phi\left(x, t\right), \qquad \eta = \psi\left(x, t\right) \tag{5.73}$$

should straighten the characteristics, at least locally, converting $\partial_{\mathbf{v}}\partial_{\mathbf{w}}$ into a multiple of $\partial_{\xi\eta}$.

First of all, however, we have to make sure that the transformation (5.73) is *non-degenerate*, at least locally, or, in other words, that the Jacobian of the transformation does not vanish:

$$\phi_t\psi_x - \phi_x\psi_t \neq 0. \tag{5.74}$$

---

[19] Remember that

$$ax^2 + 2bxy + cy^2 = a\left(x - x_1\right)\left(x - x_2\right)$$

where

$$x_{1,2} = \left[-b \pm \sqrt{b^2 - ac}\right]/a.$$

On the other hand, this follows from the fact that the vectors $\nabla\phi$ and $\nabla\psi$ are orthogonal to $\mathbf{v}$ and $\mathbf{w}$, respectively, and that $\mathbf{v}$, $\mathbf{w}$ are nowhere colinear (since $b^2 - ac > 0$).

Thus, at least locally, the inverse transformation

$$x = \Phi\left(\xi, \eta\right), \qquad t = \Psi\left(\xi, \eta\right)$$

exists. Let

$$U\left(\xi, \eta\right) = u\left(\Phi\left(\xi, \eta\right), \Psi\left(\xi, \eta\right)\right).$$

Then

$$u_x = U_\xi \phi_x + U_\eta \psi_x, \qquad u_t = U_\xi \phi_t + U_\eta \psi_t$$

and moreover

$$u_{tt} = \phi_t^2 U_{\xi\xi} + 2\phi_t \psi_t U_{\xi\eta} + \psi_t^2 U_{\eta\eta} + \phi_{tt} U_\xi + \psi_{tt} U_\eta$$

$$u_{xx} = \phi_x^2 U_{\xi\xi} + 2\phi_x \psi_x U_{\xi\eta} + \psi_x^2 U_{\eta\eta} + \phi_{xx} U_\xi + \psi_{xx} U_\eta$$

$$u_{xt} = \phi_t \phi_x U_{\xi\xi} + (\phi_x \psi_t + \phi_t \psi_x) U_{\xi\eta} + \psi_t \psi_x U_{\eta\eta} + \phi_{xt} U_\xi + \psi_{xt} U_\eta.$$

Then

$$a u_{tt} + 2b u_{xy} + c u_{xx} = A U_{\xi\xi} + 2B U_{\xi\eta} + C U_{\eta\eta} + D U_\xi + E U_\eta$$

where[20]

$$A = a\phi_t^2 + 2b\phi_t \phi_x + c\phi_x^2, \qquad C = a\psi_t^2 + 2b\psi_t \psi_x + c\psi_x^2$$
$$B = a\phi_t \psi_t + b(\phi_x \psi_t + \phi_t \psi_x) + c\phi_x \psi_x$$
$$D = a\phi_{tt} + 2b\phi_{xt} + c\phi_{xx}, \qquad E = a\psi_{tt} + 2b\psi_{xt} + c\psi_{xx}.$$

Now, $A = C = 0$, since ϕ and ψ both satisfy (5.72), so that

$$a u_{tt} + 2b u_{xt} + c u_{xx} = 2B U_{\xi\eta} + D U_\xi + E U_\eta.$$

We claim that $B \neq 0$; indeed, recalling that $\Lambda^+ \Lambda^- = c/a$, $\Lambda^+ + \Lambda^+ = -2b/a$ and

$$\phi_t = \Lambda^+ \phi_x, \qquad \psi_t = \Lambda^- \psi_x,$$

after elementary computations we find

$$B = \frac{2}{a}\left(ac - b^2\right)\phi_x \psi_x.$$

From (5.71) and (5.74) we deduce that $B \neq 0$. Thus, (5.65) assumes the form

$$U_{\xi\eta} = \mathcal{F}\left(\xi, \eta, U, U_\xi, U_\eta\right)$$

which is its *canonical form*.

---

[20] It is understood that all the functions are evaluated at $x = \Phi\left(\xi, \eta\right)$ and $t = \Psi\left(\xi, \eta\right)$.

The curves (5.70) are called *characteristics* for (5.65) and are the solution curves of the ordinary differential equations

$$\frac{dx}{dt} = -\Lambda^+, \ \frac{dx}{dt} = -\Lambda^-,$$ (5.75)

respectively. Note that the two equations (5.75) can be put into the compact form

$$a \left(\frac{dx}{dt}\right)^2 - 2b\frac{dx}{dt} + c = 0.$$ (5.76)

Example 5.1. Consider the equation

$$x u_{tt} - \left(1 + x^2\right) u_{xt} = 0.$$ (5.77)

Since $b^2 - ac = \left(1 + x^2\right)/4 > 0$, (5.77) is hyperbolic. Equation (5.76) is

$$x \left(\frac{dx}{dt}\right)^2 + \left(1 + x^2\right) \frac{dx}{dt} = 0$$

which yields, for $x \neq 0$,

$$\frac{dx}{dt} = -\frac{1 + x^2}{x} \quad \text{and} \quad \frac{dx}{dt} = 0 .$$

Thus, the characteristics curves are:

$$\phi(x, t) = e^{2t} \left(1 + x^2\right) = k_1 \quad \text{and} \quad \psi(x, t) = x = k_2.$$

We set

$$\xi = e^{2t} \left(1 + x^2\right) \quad \text{and} \quad \eta = x.$$

After routine calculations, we find $D = E = 0$ so that the canonical form is

$$U_{\xi\eta} = 0.$$

The general solution of (5.77) is therefore

$$u(x, t) = F \left(e^{2t} \left(1 + x^2\right)\right) + G(x)$$

with F and G arbitrary C^2 functions.

Case 2: $b^2 - ac \equiv 0$, the equation is **parabolic**. There exists **only one** family of characteristics, given by $\phi(x, t) = k$, where ϕ is a solution of the first order equation

$$a\phi_t + b\phi_x = 0,$$

since $\Lambda^+ = \Lambda^- = -b/a$. If ϕ is known, choose any smooth function ψ such that $\nabla\phi$ and $\nabla\psi$ are linearly independent and

$$a\psi_t^2 + 2b\psi_t\psi_x + c\psi_x^2 = C \neq 0.$$

Set

$$\xi = \phi\,(x,t)\,, \qquad \eta = \psi\,(x,t)$$

and

$$U\,(\xi,\eta) = u\,(\varPhi\,(\xi,\eta)\,,\varPsi\,(\xi,\eta))\,.$$

For the derivatives of U we can use the computations done in **case 1**. However, observe that, since $b^2 - ac = 0$ and $a\phi_t + b\phi_x = 0$, we have

$$B = a\phi_t\psi_t + b(\phi_t\psi_x + \phi_x\psi_t) + c\phi_x\psi_x = \psi_t(a\phi_t + b\phi_x) + \psi_x(b\phi_t + c\phi_x)$$

$$= b\psi_x\left(\phi_t + \frac{c}{b}\phi_x\right) = b\psi_x\left(\phi_t + \frac{b}{a}\phi_x\right) = \frac{b}{a}\psi_x(a\phi_t + b\phi_x) = 0.$$

Thus, the equation for U becomes

$$CU_{\eta\eta} = \mathcal{F}\,(\xi,\eta,U,U_\xi,U_\eta)$$

which is the *canonical form*.

Example 5.2. The equation

$$u_{tt} - 6u_{xt} + 9u_{xx} = u$$

is parabolic. The family of characteristics is

$$\phi\,(x,t) = 3t + x = k.$$

Choose $\psi\,(x,t) = x$ and set

$$\xi = 3t + x, \qquad \eta = x.$$

Since $\nabla\phi = (3,1)$ and $\nabla\psi = (1,0)$, the gradients are independent and we set

$$U\,(\xi,\eta) = u\left(\frac{\xi - \eta}{3},\eta\right).$$

We have, $D = E = 0$, so that the equation for U is

$$U_{\eta\eta} - U = 0$$

whose general solution is

$$U\,(\xi,\eta) = F\,(\xi)\,e^{-\eta} + G\,(\xi)\,e^{\eta}$$

with F and G arbitrary C^2 functions. Finally, we find

$$u\,(x,t) = F\,(3t + x)\,e^{-x} + G\,(3t + x)\,e^{x}.$$

Case 3: $b^2 - ac < 0$, the equation is **elliptic**. In this case there are no real characteristics. If the coefficients a, b, c are analytic functions[21] we can proceed as in case **1**, with two families of complex characteristics. This yields the canonical form

$$U_{zw} = \mathcal{G}\left(z, w, U, U_z, U_w\right) \qquad z, w \in \mathbb{C}.$$

Letting

$$z = \xi + i\eta, \; w = \xi - i\eta$$

and $\widetilde{U}\left(\xi, \eta\right) = U\left(\xi + i\eta, \xi - i\eta\right)$ we can eliminate the complex variables arriving at the real canonical form

$$\widetilde{U}_{\xi\xi} + \widetilde{U}_{\eta\eta} = \tilde{\mathcal{G}}\left(\xi, \eta, \tilde{U}, \tilde{U}_\xi, \tilde{U}_\eta\right).$$

5.6 Hyperbolic Systems with Constant Coefficients

In principle, it is always possible and often convenient, to reduce second order equations to first order systems. For instance, the change of variables

$$u_x = w_1 \;\; \text{and} \;\; u_t = w_2$$

transforms the wave equation $u_{tt} - c^2 u_{xx} = f$ into the system

$$\mathbf{w}_t + \mathbf{A}\mathbf{w}_x = \mathbf{f}, \tag{5.78}$$

where[22] $\mathbf{w} = (w_1, w_2)^\top$, $\mathbf{f} = (0, f)^\top$ and

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

Note that the matrix $\mathbf{A}$ has the two real distinct eigenvalues $\lambda_\pm = \pm c$, with eigenvectors

$$\mathbf{v}_+ = (1, -c)^\top \quad \text{and} \quad \mathbf{v}_- = (1, c)^\top$$

normal to the characteristics, reflecting the hyperbolic nature of the wave equation.
More generally, consider the linear system

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x + \mathbf{B}\mathbf{u} = \mathbf{f}\left(x, t\right) \qquad x \in \mathbb{R}, \; t > 0,$$

where $\mathbf{u}$ and $\mathbf{f}$ are column vectors in $\mathbb{R}^m$ and $\mathbf{A}, \mathbf{B}$ are constant $m \times m$ matrices, with the initial condition

$$\mathbf{u}\left(x, 0\right) = \mathbf{g}\left(x\right) \qquad x \in \mathbb{R}.$$

We say that the system is **hyperbolic** if $\mathbf{A}$ has m *real distinct* eigenvalues λ_1, $\lambda_2, ..., \lambda_m$.

---

[21] I.e. they can be locally expanded in Taylor series.
[22] The symbol $\top$ denotes *transposition*.

In this case, we can solve our initial value problem, extending the method of characteristics. Namely, there exists in $\mathbb{R}^m$ a base of m (column) eigenvectors

$$\mathbf{V}^1, \ \mathbf{V}^2, ..., \mathbf{V}^m.$$

If we introduce the non singular matrix

$$\boldsymbol{\Gamma} = \left(\mathbf{V}^1 \mid \mathbf{V}^2 \mid ... \mid \mathbf{V}^m\right)$$

then

$$\boldsymbol{\Gamma}^{-1}\mathbf{A}\boldsymbol{\Gamma} = \boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \ \lambda_2, ..., \lambda_m\right).$$

Now, letting $\mathbf{v} = \boldsymbol{\Gamma}^{-1}\mathbf{u}$, we discover that $\mathbf{v}$ solves the system

$$\mathbf{v}_t + \boldsymbol{\Lambda}\mathbf{v}_x = \mathbf{B}^*\mathbf{v} + \mathbf{f}^* \qquad x \in \mathbb{R}, \ t > 0 \tag{5.79}$$

where $\mathbf{B}^* = \boldsymbol{\Gamma}^{-1}\mathbf{B}\boldsymbol{\Gamma}$ and $\mathbf{f}^* = \boldsymbol{\Gamma}^{-1}\mathbf{f}$, with initial condition

$$\mathbf{v}\left(x, 0\right) = \mathbf{g}^*\left(x\right) = \boldsymbol{\Gamma}^{-1}\mathbf{g}\left(x\right) \qquad x \in \mathbb{R}.$$

The left hand side of system (5.79) is *uncoupled* and the equation for the component v_k of $\mathbf{v}$ takes the form:

$$\left(v_k\right)_t + \lambda_k\left(v_k\right)_x = \sum_{j=1}^m b_{kj}^* v_j + f_k^* \qquad k = 1, ..., m.$$

Note that if $b_{kj}^* = 0$ for $j \neq k$, then the right hand side is uncoupled as well. Thus the above equation becomes

$$\left(v_k\right)_t + \lambda_k\left(v_k\right)_x = b_{kk}^* v_k + f_k^* \qquad k = 1, ..., m \tag{5.80}$$

and can be solved by the method of characteristic, as described in Chapter 4.

Coherently, we call *characteristics* the straight lines γ_k

$$x - \lambda_k t = k, \qquad k = 1, ..., m.$$

In the particular case of **homogeneous systems**, that is

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = \mathbf{0} \qquad x \in \mathbb{R}, \ t > 0, \tag{5.81}$$

equation (5.80) is $\left(v_k\right)_t + \lambda_k\left(v_k\right)_x = 0$ and its general solution is the travelling wave $v_k\left(x, t\right) = w_k(x - \lambda_k t)$, with w_k arbitrary and differentiable. Then, since $\mathbf{u} = \boldsymbol{\Gamma}\mathbf{v}$, the general solution of (5.81) is given by the following linear combination of travelling waves:

$$\mathbf{u}\left(x, t\right) = \sum_{k=1}^m w_k(x - \lambda_k t)\mathbf{V}^k. \tag{5.82}$$

Choosing $w_k = g_k^*$ we find the unique solution satisfying $\mathbf{u}\left(x, 0\right) = \mathbf{g}\left(x\right)$.

• *The telegrapher's system.* Systems of first order equations arise in many areas of applied sciences. A classical example is

$$LI_t + V_x + RI = 0, \tag{5.83}$$

$$CV_t + I_x + GV = 0 \tag{5.84}$$

which describes the flow of electricity in a line, such as a coaxial cable. The variable x is a coordinate along the cable. $I = I(x, t)$ and $V(x, t)$ represent the current in the inner wire and the voltage across the cable, respectively. The electrical properties of the line are encoded by the constants C, capacitance to ground, R, resistance and G, conductance to ground, all per unit length.

We assign initial conditions

$$I(x, 0) = I_0(x), \; V(x, 0) = V_0(x).$$

Introducing the column vector $\mathbf{u} = (V, I)^\top$ and the matrices

$$\mathbf{A} = \begin{pmatrix} 0 & 1/L \\ 1/C & 0 \end{pmatrix} \qquad \mathbf{M} = \begin{pmatrix} -R/L & 0 \\ 0 & -G/L \end{pmatrix},$$

we may write the system in the form

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = \mathbf{M}\mathbf{u}. \tag{5.85}$$

Also in this case the matrix $\mathbf{A}$ has *real distinct eigenvalues* $\lambda_{1,2} = \pm 1/\sqrt{LC}$, with corresponding eigenvectors

$$\mathbf{v}_1 = \left(\sqrt{C}, \sqrt{L}\right)^\top \qquad \mathbf{v}_2 = \left(\sqrt{C}, -\sqrt{L}\right)^\top.$$

Thus, system (5.85) is *hyperbolic.* Let

$$\mathbf{\Gamma} = \begin{pmatrix} \sqrt{C} & \sqrt{C} \\ \sqrt{L} & -\sqrt{L} \end{pmatrix}$$

and

$$\mathbf{w} = \mathbf{\Gamma}^{-1}\mathbf{u} = \frac{1}{2} \begin{pmatrix} 1/\sqrt{C} & 1/\sqrt{L} \\ 1/\sqrt{C} & -1/\sqrt{L} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Then $\mathbf{w}$ solves

$$\mathbf{w}_t + \mathbf{\Lambda}\mathbf{u}_x = \mathbf{D}\mathbf{w} \tag{5.86}$$

where

$$\mathbf{\Lambda} = \begin{pmatrix} 1/\sqrt{LC} & 0 \\ 0 & -1/\sqrt{LC} \end{pmatrix}, \quad \mathbf{D} = \mathbf{\Gamma}^{-1}\mathbf{M}\mathbf{\Gamma} = -\frac{1}{2LC} \begin{pmatrix} RC + GL & RC - GL \\ RC - GL & RC + GL \end{pmatrix}.$$

The left hand side of (5.86) is uncoupled. In the special case $RC = GL$ (see Problem 5.13), the full system is uncoupled and reduces to the *equations*

$$w_t^\pm \pm \frac{1}{\sqrt{LC}} w_x^\pm = -\frac{R}{L} w^\pm \tag{5.87}$$

with initial conditions

$$w^{\pm}(x, 0) = \frac{1}{2} \left[\frac{I_0(x)}{\sqrt{C}} \pm \frac{V_0(x)}{\sqrt{L}} \right] \equiv w_0^{\pm}(x).$$

Applying to both equations (5.87) the method of characteristics, we find (Section 4.2.3.)

$$w^{\pm}(x, t) = w_0^{\pm} \left(x \pm \frac{1}{\sqrt{LC}} t \right) e^{-\frac{R}{L} t}.$$

Finally, formula (5.82) gives

$$\mathbf{u}(x, t) = \left\{ w_0^+ (x + t/\sqrt{LC}) \left(\frac{\sqrt{C}}{\sqrt{L}} \right) + w_0^- (x - t/\sqrt{LC}) \left(\frac{\sqrt{C}}{-\sqrt{L}} \right) \right\} e^{-\frac{R}{L} t}.$$

Thus, the solution is given by the superposition of two damped travelling waves. If $RC \neq GL$ there is no explicit formulas and one has to resort to numerical methods.

Remark 5.5. When the relevant domain is a quadrant, say $x > 0$, $t > 0$, or a half-strip $(a, b) \times (0 + \infty)$, some caution is necessary to get a well posed problem. For instance, consider the problem

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_x = \mathbf{0} \qquad x \in [0, R], t > 0 \tag{5.88}$$

with the initial condition

$$\mathbf{u}(x, 0) = \mathbf{g}(x) \qquad x \in [0, R].$$

Which kind of data and where should they be assigned to uniquely determine $\mathbf{u}$? Look at the $k-$th equation of the uncoupled problem

$$(v_k)_t + \lambda_k (v_k)_x = 0.$$

Suppose $\lambda_k > 0$, so that the characteristic γ_k is *inflow on* $x = 0$ and *outflow on* $x = R$. Guided by the scalar case (subsection 4.2.4), we must assign the value of v_k *only on* $x = 0$. On the contrary, if $\lambda_k < 0$, the value of v_k has to be assigned on $x = R$.

The conclusion is: suppose that r eigenvalues (say $\lambda_1, \lambda_2, ..., \lambda_r$) are positive and the other $m - r$ eigenvalues are negative. *Then the values of $v_1, ..., v_r$ have to be assigned on $x = 0$ and the values of $v_{r+1}, ..., v_m$ on $x = R$.* In terms of the original unknown $\mathbf{u}$, this amounts to assign, on $x = 0$, r independent linear combinations of the $\mathbf{u}$ components:

$$\left(\mathbf{\Gamma}^{-1} \mathbf{u} \right)_k = \sum_{j=1}^{m} c^{jk} u_j \qquad k = 1, 2, ..., r,$$

while other $m - r$ have to be assigned on $x = R$.

5.7 The Multi-dimensional Wave Equation $(n > 1)$

5.7.1 Special solutions

The wave equation

$$u_{tt} - c^2 \Delta u = f, \qquad (5.89)$$

constitutes a basic model for describing a remarkable number of oscillatory phenomena in dimension $n > 1$. Here $u = u(\mathbf{x}, t)$, $\mathbf{x} \in \mathbb{R}^n$ and, as in the one-dimensional case, c is the *speed of propagation*. If $f \equiv 0$, the equation is said *homogeneous* and the *superposition principle holds*. Let us examine some relevant solutions of (5.89).

- *Plane waves.* If $\mathbf{k} \in \mathbb{R}^n$ and $\omega^2 = c^2 |\mathbf{k}|^2$, the function

$$u(\mathbf{x}, t) = w(\mathbf{x} \cdot \mathbf{k} - \omega t)$$

is a solution of the homogeneous (5.89). Indeed,

$$u_{tt}(\mathbf{x}, t) - c^2 \Delta u(\mathbf{x}, t) = \omega^2 w''(\mathbf{x} \cdot \mathbf{n} - \omega t) - c^2 |\mathbf{k}|^2 w''(\mathbf{x} \cdot \mathbf{n} - \omega t) = 0.$$

We have already seen in subsection 5.1.1 that the planes

$$\mathbf{x} \cdot \mathbf{k} - \omega t = \text{constant}$$

constitute the wave fronts, moving at speed $c_p = \omega / |\mathbf{k}|$ in the $\mathbf{k}$ direction. The scalar $\lambda = 2\pi / |\mathbf{k}|$ is the wavelength. If $w(z) = Ae^{iz}$, the wave is said *monochromatic* or *harmonic*.

- *Cylindrical waves* $(n = 3)$ are of the form

$$u(\mathbf{x}, t) = w(r, t)$$

where $\mathbf{x} = (x_1, x_2, x_3)$, $r = \sqrt{x_1^2 + x_2^2}$. In particular, solutions like $u(\mathbf{x}, t) = e^{i\omega t} w(r)$ represent stationary cylindrical waves, that can be found solving the homogeneous version of equation (5.89) using the separation of variables, in axially symmetric domains.

If the axis of symmetry is the x_3 axis, it is appropriate to use the cylindrical coordinates $x_1 = r \cos \theta$, $x_2 = r \sin \theta$, x_3. Then, the wave equation becomes[23]

$$u_{tt} - c^2 \left(u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta} + u_{x_3 x_3} \right) = 0.$$

Looking for standing waves of the form $u(r, t) = e^{i\lambda ct} w(r)$, $\lambda \geq 0$, we find, after dividing by $c^2 e^{i\lambda ct}$,

$$w''(r) + \frac{1}{r} w' + \lambda^2 w = 0.$$

This is a Bessel equation of zero order. We know that the only solutions bounded at $r = 0$ are

$$w(r) = a J(\lambda r), \qquad a \in \mathbb{R}$$

---

[23] Appendix C.

where, we recall,

$$J_0(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k!)^2} \left(\frac{x}{2}\right)^{2k}$$

is the Bessel function of first kind of zero order. In this way we obtain waves of the form

$$u(r,t) = aJ_0(\lambda r)e^{i\lambda ct}.$$

- *Spherical waves* $(n = 3)$ are of the form

$$u(\mathbf{x},t) = w(r,t)$$

where $\mathbf{x} = (x_1, x_2, x_3)$, $r = |\mathbf{x}| = \sqrt{x_1^2 + x_2^2 + x_3^2}$. In particular $u(\mathbf{x},t) = e^{i\omega t}w(r)$ represent standing spherical waves and can be determined by solving the homogeneous version of equation (5.89) using separation of variables in spherically symmetric domains. In this case, spherical coordinates

$$x_1 = r\cos\theta\sin\psi, \ x_2 = r\sin\theta\sin\psi, \ x_3\cos\psi,$$

are appropriate and the wave equation becomes[24]

$$\frac{1}{c^2}u_{tt} - u_{rr} - \frac{2}{r}u_r - \frac{1}{r^2}\left\{\frac{1}{(\sin\psi)^2}u_{\theta\theta} + u_{\psi\psi} + \frac{\cos\psi}{\sin\psi}u_\psi\right\} = 0. \tag{5.90}$$

Let us look for solutions of the form $u(r,t) = e^{i\lambda ct}w(r)$, $\lambda \geq 0$. We find, after simplifying out $c^2 e^{i\lambda ct}$,

$$w''(r) + \frac{2}{r}w' + \lambda^2 w = 0$$

which can be written[25]

$$(rw)'' + \lambda^2 rw = 0.$$

Thus, $v = rw$ is solution of

$$v'' + \lambda^2 v = 0$$

which gives $v(r) = a\cos(\lambda r) + b\sin(\lambda r)$ and hence the attenuated spherical waves

$$w(r,t) = ae^{i\lambda ct}\frac{\cos(\lambda r)}{r}, \qquad w(r,t) = be^{i\lambda ct}\frac{\sin(\lambda r)}{r}. \tag{5.91}$$

Let us now determine the general form of a spherical wave in $\mathbb{R}^3$. Inserting $u(\mathbf{x},t) = w(r,t)$ into (5.90) we obtain

$$w_{tt} - c^2\left\{w_{rr}(r) + \frac{2}{r}w_r\right\} = 0$$

---

[24] Appendix C.

[25] Thanks to the miraculous presence of the factor 2 in the coefficient of w'!

which can be written in the form

$$(rw)_{tt} - c^2 (rw)_{rr} = 0. \tag{5.92}$$

Then, formula (5.41) gives

$$w(r, t) = \frac{F(r + ct)}{r} + \frac{G(r - ct)}{r} \equiv w_i(r, t) + w_o(r, t) \tag{5.93}$$

which represents the superposition of two attenuated progressive spherical waves. The wave fronts of u_o are the spheres $r - ct = k$, expanding as time goes on. Hence, w_o represents an *outgoing wave*. On the contrary, the wave w_i is *incoming*, since its wave fronts are the contracting spheres $r + ct = k$.

5.7.2 Well posed problems. Uniqueness

The well posed problems in dimension one, are still well posed in any number of dimensions. Let

$$Q_T = \Omega \times (0, T)$$

a *space-time cylinder*, where Ω is a bounded C^1-domain[26] in $\mathbb{R}^n$. A solution $u(\mathbf{x}, t)$ is uniquely determined by assigning initial data and appropriate boundary conditions on the boundary $\partial \Omega$ of Ω.

More specifically, we may pose the following problems: *Determine $u = u(\mathbf{x}, t)$ such that:*

$$\begin{cases} u_{tt} - c^2 \Delta u = f & \text{in } Q_T \\ u(\mathbf{x}, 0) = g(\mathbf{x}), \, u_t(\mathbf{x}, 0) = h(\mathbf{x}) & \text{in } \Omega \\ + \text{ boundary conditions} & \text{on } \partial \Omega \times [0, T] \end{cases} \tag{5.94}$$

where the boundary conditions are:

(a) $u = h$ (Dirichlet),

(b) $\partial_{\boldsymbol{\nu}} u = h$ (Neumann),

(c) $\partial_{\boldsymbol{\nu}} u + \alpha u = h$ ($\alpha > 0$, Robin),

(d) $u = h_1$ on $\partial_D \Omega$ and $\partial_\nu u = h_2$ on $\partial_N \Omega$ (mixed problem) with $\partial_N \Omega$ a relatively open subset of $\partial \Omega$ and $\partial_D \Omega = \partial \Omega \backslash \partial_N \Omega$.

The *global Cauchy problem*

$$\begin{cases} u_{tt} - c^2 \Delta u = f & \mathbf{x} \in \mathbb{R}^n, t > 0 \\ u(\mathbf{x}, 0) = g(\mathbf{x}), \, u_t(\mathbf{x}, 0) = h(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^n \end{cases} \tag{5.95}$$

is quite important also in dimension $n > 1$. We will examine it with some details later on. Particularly relevant are the different features that the solutions exhibit for $n = 2$ and $n = 3$.

---

[26] As usual we can afford corner points (e.g. a triangle or a cone) and also some edges (e.g. a cube or a hemisphere).

Under rather natural hypotheses on the data, problem (5.94) has at most one solution. To see it, we may use once again the conservation of energy, which is proportional to:

$$E\left(t\right) = \frac{1}{2} \int_{\Omega} \left\{ u_t^2 + c^2 \left|\nabla u\right|^2 \right\} d\mathbf{x}.$$

The growth rate is:

$$\dot{E}\left(t\right) = \int_{\Omega} \left\{ u_t u_{tt} + c^2 \nabla u_t \cdot \nabla u \right\} d\mathbf{x}.$$

Integrating by parts, we have

$$\int_{\Omega} c^2 \nabla u_t \cdot \nabla u \, d\mathbf{x} = c^2 \int_{\partial\Omega} u_{\boldsymbol{\nu}} u_t \, d\sigma - \int_{\Omega} c^2 u_t \Delta u \, d\mathbf{x}$$

whence, since $u_{tt} - c^2 \Delta u = f$,

$$\dot{E}\left(t\right) = \int_{\Omega} \left\{ u_{tt} - c^2 \Delta u \right\} u_t \, d\mathbf{x} + c^2 \int_{\partial\Omega} u_{\boldsymbol{\nu}} u_t \, d\sigma = \int_{\Omega} f u_t \, d\mathbf{x} + c^2 \int_{\partial\Omega} u_{\boldsymbol{\nu}} u_t \, d\sigma.$$

Now it is easy to prove the following result, where we use the symbol $C^{h,k}\left(D\right)$ to denote the set of functions h times continuously differentiable with respect to space and k times with respect to time in D.

Theorem 5.1. *Problem (5.94), coupled with one of the boundary conditions* $(a)-$ *(d) above, has at most one solution in* $C^{2,2}\left(Q_T\right) \cap C^{1,1}\left(\overline{Q}_T\right)$.

Proof. Let u_1 and u_2 be solutions of the same problem, sharing the same data. Their difference $w = u_1 - u_2$ is a solution of the homogeneous equation, with zero data. We show that $w\left(\mathbf{x},t\right) \equiv 0$.

In the case of Dirichlet, Neumann and mixed conditions, since either $w_{\boldsymbol{\nu}} = 0$ or $w_t = 0$ on $\partial\Omega \times [0,T]$, we have $\dot{E}\left(t\right) = 0$. Thus, since $E\left(0\right) = 0$, we infer:

$$E\left(t\right) = \frac{1}{2} \int_{\Omega} \left\{ w_t^2 + c^2 \left|\nabla w\right|^2 \right\} d\mathbf{x} = 0, \qquad \forall t > 0.$$

Therefore, for each $t > 0$, both w_t and $\left|\nabla w\left(\mathbf{x},t\right)\right|$ vanish so that $w\left(\mathbf{x},t\right)$ is constant. Then $w\left(\mathbf{x},t\right) \equiv 0$, since $w\left(\mathbf{x},0\right) = 0$.

For the Robin problem

$$\dot{E}\left(t\right) = -c^2 \int_{\partial\Omega} \alpha w w_t \, d\sigma = -\frac{c^2}{2} \frac{d}{dt} \int_{\partial\Omega} \alpha w^2 \, d\sigma$$

that is

$$\frac{d}{dt} \left\{ E\left(t\right) + \frac{c^2}{2} \int_{\partial\Omega} \alpha w^2 \, d\sigma \right\} = 0.$$

Hence,

$$E\left(t\right) + \frac{c^2}{2} \int_{\partial\Omega} \alpha w^2 d\sigma = \text{ constant}$$
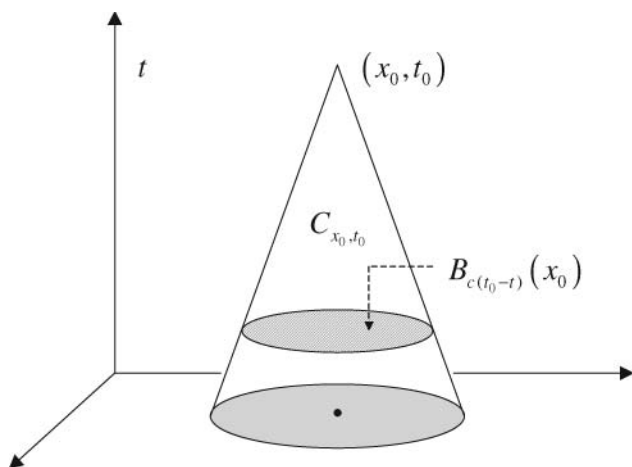
Fig. 5.9. Retrograde cone

and, being zero initially, it is zero for all $t > 0$. Since $\alpha > 0$, we again conclude that $w \equiv 0$. $\square$

Uniqueness for the global Cauchy problem follows from another energy inequality, with more interesting consequences.

First a remark. For sake of clarity, let $n = 2$. Suppose that a disturbance governed by the homogeneous wave equation $(f = 0)$ is felt at $\mathbf{x}_0$ at time t_0. Since the disturbances travel with speed c, $u(x_0, t_0)$ is, in principle, only affected by the values of the initial data in the circle $B_{ct_0}(\mathbf{x}_0)$. More generally, at time $t_0 - t$, $u(\mathbf{x}_0, t_0)$ is determined by those values in the circle $B_{c(t_0-t)}(\mathbf{x}_0)$. As t varies from 0 to t_0, the union of the circles $B_{c(t_0-t)}(\mathbf{x}_0)$ in the $\mathbf{x}, t$ space coincides with the so called *backward or retrograde cone with vertex at* $(\mathbf{x}_0, t_0)$ and opening $\theta = \tan^{-1} c$, given by (see Fig. 5.9):

$$C_{\mathbf{x}_0, t_0} = \{(\mathbf{x}, t) \colon |\mathbf{x} - \mathbf{x}_0| \le c(t_0 - t), \ 0 \le t \le t_0\}.$$

Thus, given a point $\mathbf{x}_0$, it is natural to introduce an energy associated with its backward cone by the formula

$$e(t) = \frac{1}{2} \int_{B_{c(t_0-t)}(\mathbf{x}_0)} (u_t^2 + c^2 |\nabla u|^2) d\mathbf{x}.$$

It turns out that $e(t)$ is a decreasing function. Namely:

Lemma 5.1. *Let u be a C^2-solution of the homogeneous wave equation in $\mathbb{R}^n \times [0, +\infty)$. Then*

$$\dot{e}(t) \le 0.$$

Proof. We may write

$$e(t) = \frac{1}{2} \int_0^{c(t_0-t)} dr \int_{\partial B_r(x_0)} (u_t^2 + c^2 |\nabla u|^2) d\sigma$$

so that

$$\dot{e}(t) = -\frac{c}{2}\int_{\partial B_{c(t_0-t)}(x_0)}(u_t^2 + c^2 |\nabla u|^2)d\sigma + \int_{B_{c(t_0-t)}(x_0)}\left(u_t u_{tt} + c^2 \nabla u \cdot \nabla u_t\right)d\mathbf{x}.$$

An integration by parts yields

$$\int_{B_{c(t_0-t)}(x_0)}\nabla u \cdot \nabla u_t \; d\mathbf{x} = \int_{\partial B_{c(t_0-t)}(x_0)}u_t u_{\boldsymbol{\nu}} \; d\sigma - \int_{B_{c(t_0-t)}(x_0)}u_t \Delta u \; d\mathbf{x}$$

whence

$$\dot{e}(t) = \int_{B_{c(t_0-t)}(x_0)}u_t(u_{tt} - c^2 \Delta u)d\mathbf{x} + \frac{c}{2}\int_{\partial B_{c(t_0-t)}(x_0)}(2cu_t u_{\boldsymbol{\nu}} - u_t^2 - c^2 |\nabla u|^2)d\sigma$$

$$= \frac{c}{2}\int_{\partial B_{c(t_0-t)}(x_0)}(2cu_t u_{\boldsymbol{\nu}} - u_t^2 - c^2 |\nabla u|^2)d\sigma$$

Now

$$|u_t u_{\boldsymbol{\nu}}| \le |u_t|\,|\nabla u|$$

so that

$$2cu_t u_{\boldsymbol{\nu}} - u_t^2 - c^2 |\nabla u|^2 \le 2c\,|u_t|\,|\nabla u| - u_t^2 - c^2 |\nabla u|^2 = -\left(u_t - c\,|\nabla u|\right)^2 \le 0$$

and therefore $\dot{e}(t) \le 0$. $\square$

Two almost immediate consequences are stated in the following theorem:

Theorem 5.2. *Let* $u \in C^2\left(\mathbb{R}^n \times [0, +\infty)\right)$ *be a solution of the Cauchy problem* (5.95). *Then:*

(a) *If* $g \equiv h \equiv 0$ *in* $B_{ct_0}(\mathbf{x_0})$ *and* $f \equiv 0$ *in* $C_{\mathbf{x_0}, t_0}$ *then* $u \equiv 0$ *in* $C_{\mathbf{x_0}, t_0}$.

(b) *Problem* (5.95) *has at most one solution in* $C^2\left(\mathbb{R}^n \times [0, +\infty)\right)$.

5.8 Two Classical Models

5.8.1 Small vibrations of an elastic membrane

In subsection 5.2.3 we have derived a model for the small transversal vibrations of a string. Similarly, we may derive the governing equation of the small transversal vibrations of a highly stretched membrane (think e.g. of a drum), at rest in the horizontal position. We briefly sketch the derivation leaving it to the reader to fill in the details. Assume the following hypotheses.

1. *The vibrations of the membrane are small and vertical.* This means that the changes from the plane horizontal shape are very small and horizontal displacements are negligible.

2. *The vertical displacement of a point of the membrane depends on time and on its position at rest.* Thus, if u denotes the vertical displacement of a point located at rest at (x, y), we have $u = u(x, y, t)$.

3. *The membrane is perfectly flexible and elastic.* There is no resistance to bending. In particular, the stress in the membrane can be modelled by a tangential force $\mathbf{T}$ of magnitude τ, called *tension*[27]. Perfect elasticity means that τ is a constant.

4. *Friction is negligible.*

Under the above assumptions, the equation of motion of the membrane can be derived from *conservation of mass* and *Newton's law*.

Let $\rho_0 = \rho_0(x, y)$ be the surface mass density of the membrane at rest and consider a small "rectangular" piece of membrane, with vertices at the points A, B, C, D of coordinates (x, y), $(x + \Delta x, y)$, $(x, y + \Delta y)$ and $(x + \Delta x, y + \Delta y)$, respectively. Denote by ΔS the corresponding area at time t. Then, conservation of mass yields

$$\rho_0(x, y) \Delta x \Delta y = \rho(x, y, t) \Delta S. \tag{5.96}$$

To write Newton's law of motion we have to determine the forces acting on our small piece of membrane. Since the motion is vertical, the horizontal forces have to balance.

The vertical forces are given by body forces (e.g. gravity and external loads) and the vertical component of the tension.

Denote by $f(x, y, t)\,\mathbf{k}$ the resultant of the body forces per unit mass. Then, using (5.96), the body forces acting on the membrane element are well approximated by:

$$\rho(x, y, t) f(x, y, t) \Delta S\, \mathbf{k} = \rho_0(x, y) f(x, y, t) \Delta x \Delta y\, \mathbf{k}.$$

Along the edges AB and CD, the tension is perpendicular to the $x-$axis and almost parallel to the $y-$axis. Its (scalar) vertical components are respectively given by

$$\tau_{vert}(x, y, t) \simeq \tau u_y(x, y, t) \Delta x, \qquad \tau_{vert}(x, y + \Delta y, t) \simeq \tau u_y(x, y + \Delta y, t) \Delta x.$$

Similarly, along the edge AC, the tension is perpendicular to the $y-$axis and almost parallel to the $x-$axis. Its (scalar) vertical components are respectively given by

$$\tau_{vert}(x, y, t) \simeq \tau u_x(x, y, t) \Delta y, \qquad \tau_{vert}(x + \Delta x, y, t) \simeq \tau u_x(x + \Delta x, y, t) \Delta y.$$

---

[27] The tension $\mathbf{T}$ has the following meaning. Consider a small region on the membrane, delimited by a closed curve γ. The material on one side of γ exerts on the material on the other side a *force per unit length* $\mathbf{T}$ *(pulling)* along γ. A constitutive law for $\mathbf{T}$ is

$$\mathbf{T}(x, y, t) = \tau(x, y, t)\, \mathbf{N}(x, y, t) \qquad (x, y) \in \gamma$$

where $\mathbf{N}$ is the outward unit normal vector to γ, tangent to the membrane.

Again, the tangentiality of the tension force is due to the absence of distributed moments over the membrane.

Thus, using (5.96) again and observing that u_{tt} is the (scalar) vertical acceleration, Newton's law gives:

$$\rho_0 \left(x, y \right) \Delta x \Delta y \; u_{tt} =$$

$$= \tau[u_y \left(x, y + \Delta y, t \right) - u_y \left(x, y, t \right)]\Delta x + \tau[u_x \left(x + \Delta x, y, t \right) - u_x \left(x, y, t \right)]\Delta y +$$
$$+ \rho_0 \left(x, y \right) f \left(x, y, t \right) \Delta x \Delta y.$$

Dividing for $\Delta x \Delta y$ and letting $\Delta x, \Delta y \to 0$, we obtain the equation

$$u_{tt} - c^2(u_{yy} + u_{xx}) = f \left(x, y, t \right) \tag{5.97}$$

where $c^2 \left(x, y, t \right) = \tau/\rho_0 \left(x, y \right)$.

• *Square Membrane.* Consider a membrane occupying at rest a square of side a, pinned at the boundary. We want to study its vibrations when the membrane is initially horizontal, with speed $h = h \left(x, y \right)$. If there is no external load and the weight of the membrane is negligible, the vibrations are governed by the following initial-boundary value problem:

$$\begin{cases} u_{tt} - c^2 \Delta u = 0 & 0 < x < a, \, 0 < y < a, \, t > 0 \\ u \left(x, y, 0 \right) = 0, \, u_t \left(x, y, 0 \right) = h \left(x, y \right) & 0 < x < a, \, 0 < y < a \\ u \left(0, y, t \right) = u \left(a, y, t \right) = 0 & 0 \leq y \leq a, \, t \geq 0 \\ u \left(x, 0, t \right) = u \left(x, a, t \right) = 0 & 0 \leq x \leq a, \, t \geq 0. \end{cases}$$

The square shape of the membrane and the homogeneous boundary conditions suggest the use of separation of variables. Let us look for solution of the form

$$u \left(x, y, t \right) = v \left(x, y \right) q \left(t \right)$$

with $v = 0$ at the boundary. Substituting into the wave equation we find

$$q'' \left(t \right) v \left(x, y \right) - c^2 q \left(t \right) \Delta v \left(x, y \right) = 0$$

and, separating the variables,

$$\frac{q'' \left(t \right)}{c^2 q \left(t \right)} = \frac{\Delta v \left(x, y \right)}{v \left(x, y \right)} = -\lambda^2$$

whence[28] the equation

$$q'' \left(t \right) + c^2 \lambda^2 q \left(t \right) = 0. \tag{5.98}$$

and the *eigenvalue problem*

$$\Delta v + \lambda^2 v = 0 \tag{5.99}$$

$$v \left(0, y \right) = v \left(a, y \right) = v \left(x, 0 \right) = v \left(x, a \right) = 0, \qquad 0 \leq x, y \leq a.$$

---

[28] The two ratios must be equal to the same constant. The choice of $-\lambda^2$ is guided by our former experience....

We first solve the eigenvalue problem, using once more separation of variables and setting $v(x, y) = X(x) Y(y)$, with the conditions

$$X(0) = X(a) = 0, \qquad Y(0) = Y(a) = 0.$$

Substituting into (5.99), we obtain

$$\frac{Y''(y)}{Y(y)} + \lambda^2 = -\frac{X''(x)}{X(x)} = \mu^2$$

where μ is a new constant.

Letting $\nu^2 = \lambda^2 - \mu^2$, we have to solve the following two one-dimensional eigenvalue problems, in $0 < x < a$ and $0 < y < a$, respectively:

$$\begin{cases} X''(x) + \mu^2 X(x) = 0 \\ X(0) = X(a) = 0 \end{cases} \qquad \begin{cases} Y''(y) + \nu^2 Y(y) = 0 \\ Y(0) = Y(a) = 0. \end{cases}$$

The solutions are:

$$X(x) = A_m \sin \mu_m x, \qquad \mu_m = \frac{m\pi}{a}$$

$$Y(y) = B_n \sin \nu_n y, \qquad \nu_n = \frac{n\pi}{a}$$

where $m, n = 1, 2, \dots$. Since $\lambda^2 = \nu^2 + \mu^2$, we have

$$\lambda_{mn}^2 = \frac{\pi^2}{a^2} \left(m^2 + n^2 \right), \qquad m, n = 1, 2, \dots \tag{5.100}$$

corresponding to the eigenfunctions

$$v_{mn}(x, y) = C_{mn} \sin \mu_m x \sin \nu_n y.$$

For $\lambda = \lambda_{mn}$, the general integral of (5.98) is

$$q_{mn}(t) = a_{mn} \cos c\lambda_{mn} t + b_m \sin c\lambda_{mn} t.$$

Thus we have found infinitely many special solutions to the wave equations, of the form,

$$u_{mn} = (a_{mn} \cos c\lambda_{mn} t + b_{mn} \sin c\lambda_{mn} t) \sin \mu_m x \sin \nu_n y.$$

which, moreover, vanish on the boundary.

Every u_{mn} is a standing wave and corresponds to a particular mode of vibration of the membrane. The *fundamental frequency* is $f_{11} = c\sqrt{2}/2a$, while the other frequencies are $f_{mn} = c\sqrt{m^2 + n^2}/2a$, which are **not** integer multiple of the fundamental one (as they do for the vibrating string).

Going back to our problem, to find a solution which satisfies the initial conditions, we superpose the modes u_{mn} defining

$$u(x, y, t) = \sum_{m, n = 1}^{\infty} (a_{mn} \cos c\lambda_{mn} t + b_{mn} \sin c\lambda_{mn} t) \sin \mu_m x \sin \nu_n y.$$

Since $u(x, y, 0) = 0$ we choose $a_{mn} = 0$ for every $m, n \geq 1$. From $u_t(x, y, 0) = h(x, y)$ we find the condition

$$\sum_{m,n=1}^{\infty} cb_{mn}\lambda_{mn} \sin \mu_m x \sin \nu_n x = h(x, y). \tag{5.101}$$

Therefore, we assume that h can be expanded in a double Fourier sine series as follows:

$$h(x, y) = \sum_{m,n=1}^{\infty} h_{mn} \sin \mu_m x \sin \nu_n y,$$

where the coefficients h_{mn} are given by

$$h_{mn} = \frac{4}{a^2} \int_Q h(x, y) \sin \frac{m\pi}{a} x \sin \frac{n\pi}{a} y \, dxdy.$$

Then, if we choose $b_{mm} = h_{mm}/c\lambda_{mn}$, (5.101) is satisfied. Thus, we have constructed the *formal* solution

$$u(x, y, t) = \sum_{m,n=1}^{\infty} \frac{h_{mn}}{c\lambda_{mn}} \sin c\lambda_{mn} t \sin \mu_m x \sin \nu_n y. \tag{5.102}$$

If the coefficients $h_{mm}/c\lambda_{mn}$ vanish fast enough as $m, n \to +\infty$, it can be shown that (5.102) gives the unique solution[29].

5.8.2 Small amplitude sound waves

Sound waves are small disturbances in the density and pressure of a compressible gas. In an isotropic gas, their propagation can be described in terms of a single scalar quantity. Moreover, due to the small amplitudes involved, it is possible to *linearize* the equations of motion, within a reasonable range of validity. Three are the relevant equations: two of them are *conservation of mass* and *balance of linear momentum*, the other one is a *constitutive* relation between density and pressure.

Conservation of mass expresses the relation between the gas density $\rho = \rho(\mathbf{x}, t)$ and its velocity $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$:

$$\rho_t + \operatorname{div}(\rho\mathbf{v}) = 0. \tag{5.103}$$

The balance of linear momentum describes how the volume of gas occupying a region V reacts to the pressure exerted by the rest of the gas. Assuming that the viscosity of the gas is negligible, this force is given by the *normal pressure* $-p\boldsymbol{\nu}$ on the boundary of V ($\boldsymbol{\nu}$ is the exterior normal to ∂V).

---

[29] We leave it the reader to find appropriate smoothness hypotheses on h, in order to assure that (5.102) is the unique solution.

Thus, if there are no significant external forces, the linear momentum equation is

$$\frac{D\mathbf{v}}{Dt} \equiv \mathbf{v}_t + (\mathbf{v}{\cdot}\nabla)\,\mathbf{v} = -\frac{1}{\rho}\nabla p. \tag{5.104}$$

The last equation is an empirical relation between p and ρ. Since the pressure fluctuations are very rapid, the compressions/expansions of the gas are *adiabatic*, *without any loss of heat* .

In these conditions, if $\gamma = c_p/c_v$ is the ratio of the specific heats of the gas ($\gamma \approx 1.4$ in air) then p/ρ^γ is constant, so that we may write

$$p = f(\rho) = C\rho^\gamma \tag{5.105}$$

with C constant.

The system of equations (5.103), (5.104), (5.105) is quite complicated and it would be extremely difficult to solve it in its general form. Here, the fact that sound waves are only small perturbation of normal atmospheric conditions allows a major simplification. Consider a static atmosphere, where ρ_0 and p_0 are constant density and pressure, with zero velocity field. We may write

$$\rho = (1 + s)\,\rho_0 \approx \rho_0$$

where s is a small dimensionless quantity, called *condensation* and representing the fractional variation of the density from equilibrium. Then, from (5.105), we have

$$p - p_0 \approx f'(\rho_0)\,(\rho - \rho_0) = s\rho_0 f'(\rho_0) \tag{5.106}$$

and

$$\nabla p \approx \rho_0 f'(\rho_0)\,\nabla s.$$

Now, if $\mathbf{v}$ is also small, we may keep only first order terms in s and $\mathbf{v}$. Thus, we may neglect the convective acceleration $(\mathbf{v}{\cdot}\nabla)\,\mathbf{v}$ and approximate (5.104) and (5.103) by the linear equations

$$\mathbf{v}_t = -c_0^2 \nabla s \tag{5.107}$$

and

$$s_t + \operatorname{div}\mathbf{v} = 0 \tag{5.108}$$

where we have set $c_0^2 = f'(\rho_0) = C\gamma\rho_0^{\gamma-1}$.

Let us pause for a moment to examine which implications the above linearization has. Suppose that V and S are average values of $|\mathbf{v}|$ and s, respectively. Moreover, let L and T typical order of magnitude for space and time in the wave propagation, such as wavelength and period. Rescale $\mathbf{v}$, s, $\mathbf{x}$ and t as follows:

$$\boldsymbol{\xi} = \frac{\mathbf{x}}{L}, \quad \tau = \frac{t}{T}, \quad \mathbf{U}(\boldsymbol{\xi}, \tau) = \frac{\mathbf{v}(L\boldsymbol{\xi}, T\tau)}{V}, \quad \sigma(\boldsymbol{\xi}, \tau) = \frac{s(L\boldsymbol{\xi}, T\tau)}{S}. \tag{5.109}$$

Substituting (5.109) into (5.107) and (5.108) we obtain

$$\frac{V}{T}\mathbf{U}_\tau + \frac{c_0^2 S}{L}\nabla\sigma = \mathbf{0} \quad \text{and} \quad \frac{S}{T}\sigma_\tau + \frac{V}{L}\operatorname{div}\mathbf{U} = 0.$$

In this equations the coefficients must be of the same order of magnitude, therefore

$$\frac{V}{T} \approx \frac{c_0^2 S}{L} \quad \text{and} \quad \frac{S}{T} \approx \frac{V}{L}$$

which implies

$$\frac{L}{T} \approx c_0.$$

As we see, c_0 is a typical propagation speed, namely it is **the sound speed**. Now, the convective acceleration is negligible with respect to (say) $\mathbf{v}_t$, if

$$\frac{V^2}{L}\mathbf{U}\cdot\nabla\mathbf{U} \ll \frac{V}{T}\mathbf{U}_\tau$$

or $V \ll c_0$.

Thus if the gas speed is much smaller than the sound speed, our linearization makes sense. The ratio $M = V/c_0$ is called **Mach number**.

We want to derive from (5.107) and (5.108) the following theorem in which we assume that both s and $\mathbf{v}$ are smooth functions.

Theorem 5.3. a*) The condensation s is a solution of the wave equation*

$$s_{tt} - c_0^2 \Delta s = 0 \tag{5.110}$$

where $c_0 = \sqrt{f'(\rho_0)} = \sqrt{\gamma p_0/\rho_0}$ is the speed of sound.

b) If $\mathbf{v}(\mathbf{x},0) = \mathbf{0}$, there exists an acoustic potential ϕ such that $\mathbf{v} = \nabla\phi$. Moreover ϕ satisfies (5.110) as well.

Proof. a) Taking the divergence on both sides of (5.107) and the $t-$derivative on both sides of (5.108) we get, respectively:

$$\text{div } \mathbf{v}_t = -c_0^2 \Delta s$$

and

$$s_{tt} = -(\text{div } \mathbf{v})_t.$$

Since $(\text{div } \mathbf{v})_t = \text{div } \mathbf{v}_t$, equation (5.110) follows.

b) From (5.107) we have

$$\mathbf{v}_t = -c_0^2 \nabla s.$$

Let

$$\phi(\mathbf{x},t) = -c_0^2 \int_0^t s(\mathbf{x},z)\, dz.$$

Then

$$\phi_t = -c_0^2 s$$

and we may write (5.107) in the form

$$\frac{\partial}{\partial t}\left[\mathbf{v} - \nabla\phi\right] = \mathbf{0}.$$

Hence, since $\phi\left(\mathbf{x},0\right) = 0$, $\mathbf{v}\left(\mathbf{x},0\right) = \mathbf{0}$, we infer

$$\mathbf{v}\left(\mathbf{x},t\right) - \nabla\phi\left(\mathbf{x},t\right) = \mathbf{v}\left(\mathbf{x},0\right) - \nabla\phi\left(\mathbf{x},0\right) = \mathbf{0}.$$

Thus $\mathbf{v} = \nabla\phi$. Finally, from (5.108),

$$\phi_{tt} = -c_0^2 s_t = c_0^2 \mathrm{div}\ \mathbf{v} = c_0^2 \Delta\phi$$

which is (5.110). $\square$

Once the potential ϕ is known, the velocity field $\mathbf{v}$, the condensation s and the pressure fluctuation $p - p_0$ can be computed from the following formulas:

$$\mathbf{v} = \nabla\phi, \qquad s = -\frac{1}{c_0^2}\phi_t, \qquad p - p_0 = -\rho_0\phi_t.$$

Consider, for instance, a plane wave represented by the following potential:

$$\phi\left(\mathbf{x},t\right) = w\left(\mathbf{x}\cdot\mathbf{k} - \omega t\right).$$

We know that if $c_0^2\left|\mathbf{k}\right|^2 = \omega^2$, ϕ is a solution of (5.110). In this case, we have:

$$\mathbf{v} = w'\mathbf{k}, \qquad s = \frac{\omega}{c_0^2}w', \qquad p - p_0 = \rho_0\omega w'.$$

Example 5.3. Motion of a gas in a tube. Consider a straight cylindrical tube with axis along the x_1-axis, filled with gas in the region $x_1 > 0$. A flat piston, whose face moves according to $x_1 = h\left(t\right)$, sets the gas into motion. We assume that $\left|h\left(t\right)\right| \ll 1$ and $\left|h'\left(t\right)\right| \ll c_0$. Under these conditions, the motion of the piston generates sound waves of small amplitude and the acoustic potential ϕ is a solution of the homogeneous wave equation. To compute ϕ we need boundary conditions. The continuity of the normal velocity of the gas at the contact surface with the piston gives

$$\phi_{x_1}\left(h\left(t\right), x_2, x_3, t\right) = h'\left(t\right).$$

Since $h\left(t\right) \sim 0$, we may approximate this condition by

$$\phi_{x_1}\left(0, x_2, x_3, t\right) = h'\left(t\right). \tag{5.111}$$

At the tube walls the normal velocity of the gas is zero, so that, if $\boldsymbol{\nu}$ denotes the outward unit normal vector at the tube wall, we have

$$\nabla\phi\cdot\boldsymbol{\nu} = 0. \tag{5.112}$$

Finally since the waves are generated by the piston movement, we may look for *outgoing plane waves*[30] solution of the form:

$$\phi\left(\mathbf{x},t\right) = w\left(\mathbf{x}\cdot\mathbf{n} - ct\right)$$

---

[30] We do not expect *incoming* waves, which should be generated by sources placed far from the piston.

where $\mathbf{n}$ is a unit vctor. From (5.112) we have

$$\nabla \phi \cdot \boldsymbol{\nu} = w' \left(\mathbf{x} \cdot \mathbf{n} - ct \right) \mathbf{n} \cdot \boldsymbol{\nu} = 0$$

whence $\mathbf{n} \cdot \boldsymbol{\nu} = 0$ for every $\boldsymbol{\nu}$ orthogonal to the wall tube. Thus, we infer $\mathbf{n} = (1, 0, 0)$ and, as a conseguence,

$$\phi \left(\mathbf{x}, t \right) = w \left(x_1 - ct \right).$$

From (5.111) we get

$$w' \left(-ct \right) = h' \left(t \right)$$

so that (assuming $h \left(0 \right) = 0$),

$$w \left(s \right) = -ch \left(-\frac{s}{c} \right).$$

Hence, the acoustic potential is given by

$$\phi \left(\mathbf{x}, t \right) = -ch \left(t - \frac{x_1}{c} \right)$$

which represents a *progressive wave* propagating along the tube. In this case:

$$\mathbf{v} = c\mathbf{i}, \quad s = \frac{1}{c} h' \left(t - \frac{x_1}{c} \right), \quad p = c\rho_0 h' \left(t - \frac{x_1}{c} \right) + p_0.$$

5.9 The Cauchy Problem

5.9.1 Fundamental solution ($n = 3$) and strong Huygens' principle

In this section we consider the global Cauchy problem for the three-dimensional homogeneous wave equation:

$$\begin{cases} u_{tt} - c^2 \Delta u = 0 & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ u \left(\mathbf{x}, 0 \right) = g \left(\mathbf{x} \right), \quad u_t \left(\mathbf{x}, 0 \right) = h \left(\mathbf{x} \right) & \mathbf{x} \in \mathbb{R}^3. \end{cases} \tag{5.113}$$

We know from Theorem 5.2 that problem (5.113) has at most one solution $u \in C^2 \left(\mathbb{R}^3 \times [0, +\infty) \right)$. Our purpose here is to show that the solution u exists and to find an explicit formula for it, in terms of the data g and h. Our derivation is rather heuristic so that, for the time being, we do not worry too much about the correct hypotheses on h and g, which we assume as smooth as we need to carry out the calculations.

First we need a lemma that reduces the problem to the case $g = 0$ (and which actually holds in any dimension). Denote by w_h the solution of the problem

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ w \left(\mathbf{x}, 0 \right) = 0, \quad w_t \left(\mathbf{x}, 0 \right) = h \left(\mathbf{x} \right) & \mathbf{x} \in \mathbb{R}^3. \end{cases} \tag{5.114}$$

Lemma 5.2. *If w_g has continuous third-order partials, then $v = \partial_t w_g$ solves the problem*

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ w(\mathbf{x}, 0) = g(\mathbf{x}), \quad w_t(\mathbf{x},0) = 0 & \mathbf{x} \in \mathbb{R}^3. \end{cases} \tag{5.115}$$

Therefore the solution of (5.113) is given by

$$u = \partial_t w_g + w_h. \tag{5.116}$$

Proof. Let $v = \partial_t w_g$. Differentiating the wave equation with respect to t we have

$$0 = \partial_t(\partial_{tt} w_g - c^2 \Delta w_g) = (\partial_{tt} - c^2 \Delta)\partial_t w_g = v_{tt} - c^2 \Delta v.$$

Moreover,

$$v(\mathbf{x},0) = \partial_t w_g(\mathbf{x},0) = g(\mathbf{x}), \qquad v_t(\mathbf{x},0) = \partial_{tt} w_g(\mathbf{x},0) = c^2 \Delta w_g(\mathbf{x},0) = 0.$$

Thus, v is a solution of (5.115) and $u = v + w_h$ is the solution of (5.113). $\square$

The lemma shows that, once the solution of (5.114) is determined, the solution of the complete problem (5.113) is given by (5.116).

Therefore, we focus on the solution of (5.114), first with a special h, given by the three-dimensional Dirac measure at $\mathbf{y}$, $\delta(\mathbf{x} - \mathbf{y})$. For example, in the case of sound waves, this initial data models a sudden change of the air density, concentrated at a point $\mathbf{y}$. If w represents the density variation with respect to a static atmosphere, then w solves the problem

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ w(\mathbf{x}, 0) = 0, \quad w_t(\mathbf{x},0) = \delta(\mathbf{x} - \mathbf{y}) & \mathbf{x} \in \mathbb{R}^3. \end{cases} \tag{5.117}$$

The solution of (5.117), which we denote by $K(\mathbf{x}, \mathbf{y},t)$, is called **fundamental solution** of the three-dimensional wave equation. To solve (5.117) we use ... *the heat equation* (!), approximating the Dirac measure with the *fundamental* solution of the three-dimensional diffusion equation. Indeed, from section 2.3.4, (choosing $t = \varepsilon$, $D = 1$, $n = 3$) we know that

$$\Gamma(\mathbf{x} - \mathbf{y},\varepsilon) = \frac{1}{(4\pi\varepsilon)^{3/2}} \exp\left\{-\frac{|\mathbf{x} - \mathbf{y}|^2}{4\varepsilon}\right\} \to \delta(\mathbf{x} - \mathbf{y})$$

as $\varepsilon \to 0$. Denote by w_ε the solution of (5.117) with $\delta(\mathbf{x} - \mathbf{y})$ replaced by $\Gamma(\mathbf{x} - \mathbf{y},\varepsilon)$. Since $\Gamma(\mathbf{x} - \mathbf{y},\varepsilon)$ is radially symmetric with pole at $\mathbf{y}$, we expect that w_ε shares the same type of symmetry and is a spherical wave of the form $w_\varepsilon = w_\varepsilon(r, t)$, $r = |\mathbf{x} - \mathbf{y}|$. Thus, from (5.93) we may write

$$w_\varepsilon(r,t) = \frac{F(r + ct)}{r} + \frac{G(r - ct)}{r}. \tag{5.118}$$

The initial conditions require

$$F(r) + G(r) = 0 \qquad \text{and} \qquad c(F'(r) - G'(r)) = r\Gamma(r,\varepsilon)$$

or

$$F = -G \qquad \text{and} \qquad G'(r) = -r\Gamma(r,\varepsilon)/2c.$$

Integrating the second relation yields

$$G(r) = -\frac{1}{2c(4\pi\varepsilon)^{3/2}} \int_0^r s \exp\left\{-\frac{s^2}{4\varepsilon}\right\} ds = \frac{1}{4\pi c}\frac{1}{\sqrt{4\pi\varepsilon}}\left(\exp\left\{-\frac{r^2}{4\varepsilon}\right\} - 1\right)$$

and finally

$$w_\varepsilon(r,t) = \frac{1}{4\pi cr}\left\{\frac{1}{\sqrt{4\pi\varepsilon}}\exp\left\{-\frac{(r-ct)^2}{4\varepsilon}\right\} - \frac{1}{\sqrt{4\pi\varepsilon}}\exp\left\{-\frac{(r+ct)^2}{4\varepsilon}\right\}\right\}.$$

Now observe that the function

$$\widetilde{\Gamma}(r,\varepsilon) = \frac{1}{\sqrt{4\pi\varepsilon}}\exp\left\{-\frac{r^2}{4\varepsilon}\right\}$$

is the fundamental solution of the one-dimensional diffusion equation with $x = r$ and $t = \varepsilon$. Letting $\varepsilon \to 0$ we find[31]

$$w_\varepsilon(r,t) \to \frac{1}{4\pi cr}\left\{\delta(r-ct) - \delta(r+ct)\right\}.$$

Since $r + ct > 0$ for every $t > 0$, we deduce that $\delta(r + ct) = 0$ and therefore we conclude that

$$K(\mathbf{x},\mathbf{y},t) = \frac{\delta(r-ct)}{4\pi cr} \qquad r = |\mathbf{x} - \mathbf{y}|. \qquad (5.119)$$

Thus, the fundamental solution is an *outgoing travelling wave*, initially concentrated at $\mathbf{y}$ and thereafter on

$$\partial B_{ct}(\mathbf{y}) = \{\mathbf{x} : |\mathbf{x} - \mathbf{y}| = ct\}.$$

The union of the surfaces $\partial B_{ct}(\mathbf{y})$ is called the **support** of K and coincides with **the boundary** of *the forward space-time cone, with vertex at* $(\mathbf{y}, 0)$ *and opening* $\theta = \tan^{-1}c$, given by

$$C_{\mathbf{y},0}^* = \{(\mathbf{x},t): |\mathbf{x} - \mathbf{y}| \le ct,\, t > 0\}.$$

In the terminology of Section 4, $\partial C_{\mathbf{y},0}^*$ constitutes the **range of influence of the point y**.

The fact that the range of influence of the point $\mathbf{y}$ is only the *boundary* of the forward cone and *not the full* cone has important consequences on the nature of the disturbances governed by the three-dimensional wave equation. The most striking phenomenon is that a perturbation generated at time $t = 0$ by a point source placed at $\mathbf{y}$ is felt at the point $\mathbf{x}_0$ **only at time** $t_0 = |\mathbf{x}_0 - \mathbf{y}|/c$ (Fig. 5.10). This is known as *strong Huygens' principle* and explains why *sharp signals* are propagated from a point source.

We will shortly see that this is not the case in two dimensions.

---

[31] Here δ is one-dimensional.
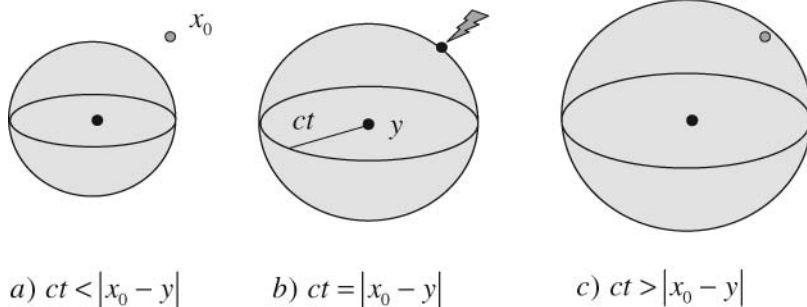
Fig. 5.10. Huygens principle

5.9.2 The Kirchhoff formula

Using the fundamental solution as in subsection 5.4.3, we may derive a formula for the solution of (5.114) with a general h. Since

$$h\left(\mathbf{x}\right) = \int_{\mathbb{R}^3} \delta\left(\mathbf{x} - \mathbf{y}\right) h\left(\mathbf{y}\right) d\mathbf{y},$$

we may see h as a superposition of impulses $\delta\left(\mathbf{x} - \mathbf{y}\right) h\left(\mathbf{y}\right)$ localized at $\mathbf{y}$, of strength $h\left(\mathbf{y}\right)$. Accordingly, the solution of (5.114) is given by the superposition of the corresponding solutions $K\left(\mathbf{x}, \mathbf{y}, t\right) h\left(\mathbf{y}\right)$, that is

$$w_h\left(\mathbf{x}, t\right) = \int_{\mathbb{R}^3} K\left(\mathbf{x}, \mathbf{y}, t\right) h\left(\mathbf{y}\right) d\mathbf{y} = \int_{\mathbb{R}^3} \frac{\delta(|\mathbf{x} - \mathbf{y}| - ct)}{4\pi c\left|\mathbf{x} - \mathbf{y}\right|} h\left(\mathbf{y}\right) d\mathbf{y} =$$

$$= \int_0^\infty \frac{\delta(r - ct)}{4\pi cr} dr \int_{\partial B_r(\mathbf{x})} h\left(\boldsymbol{\sigma}\right) d\sigma = \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} h\left(\boldsymbol{\sigma}\right) d\sigma.$$

where we have used the formula

$$\int_0^\infty \delta\left(r - ct\right) f\left(r\right) dr = f\left(ct\right).$$

Lemma 5.2 and the above intuitive argument lead to the following theorem:

Theorem 5.4. (Kirchhoff's formula). *Let* $g \in C^3\left(\mathbb{R}^3\right)$ *and* $h \in C^2\left(\mathbb{R}^3\right)$. *Then,*

$$u\left(\mathbf{x}, t\right) = \frac{\partial}{\partial t}\left[\frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} g\left(\boldsymbol{\sigma}\right) d\sigma\right] + \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} h\left(\boldsymbol{\sigma}\right) d\sigma \qquad (5.120)$$

is the unique solution $u \in C^2\left(\mathbb{R}^3 \times [0, +\infty)\right)$ *of problem (5.113)*

Proof. Letting $\boldsymbol{\sigma} = \mathbf{x} + ct\boldsymbol{\omega}$, where $\boldsymbol{\omega} \in \partial B_1\left(\mathbf{0}\right)$, we have $d\sigma = c^2 t^2 d\omega$ and we may write

$$w_g\left(\mathbf{x}, t\right) = \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} g\left(\boldsymbol{\sigma}\right) d\sigma = \frac{t}{4\pi} \int_{\partial B_1(\mathbf{0})} g\left(\mathbf{x} + ct\boldsymbol{\omega}\right) d\omega.$$

Since $g \in C^3 \left(\mathbb{R}^3 \right)$, this formula shows that w_g satisfies the hypotheses of Lemma 5.2. Therefore it is enough to check that

$$w_h \left(\mathbf{x},t \right) = \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} h \left(\boldsymbol{\sigma} \right) d\sigma = \frac{t}{4\pi} \int_{\partial B_1(\mathbf{0})} h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) d\omega$$

solves problem (5.114). We have:

$$\partial_t w_h \left(\mathbf{x},t \right) = \frac{1}{4\pi} \int_{\partial B_1(\mathbf{0})} h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) d\omega + \frac{ct}{4\pi} \int_{\partial B_1(\mathbf{0})} \nabla h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) \cdot \boldsymbol{\omega} \, d\omega. \quad (5.121)$$

Thus,

$$w_h \left(\mathbf{x},0 \right) = 0 \quad \text{and} \quad \partial_t w_h \left(\mathbf{x},0 \right) = h \left(\mathbf{x} \right).$$

Moreover, by Gauss' formula, we may write

$$\frac{ct}{4\pi} \int_{\partial B_1(\mathbf{0})} \nabla h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) \cdot \boldsymbol{\omega} \, d\omega = \frac{1}{4\pi ct} \int_{\partial B_{ct}(\mathbf{x})} \partial_\nu h \left(\boldsymbol{\sigma} \right) d\sigma$$

$$= \frac{1}{4\pi ct} \int_{B_{ct}(\mathbf{x})} \Delta h \left(\mathbf{y} \right) d\mathbf{y}$$

$$= \frac{1}{4\pi ct} \int_0^{ct} dr \int_{\partial B_r(\mathbf{x})} \Delta h \left(\boldsymbol{\sigma} \right) d\sigma$$

whence, from (5.121),

$$\partial_{tt} w_h \left(\mathbf{x},t \right) = \frac{c}{4\pi} \int_{\partial B_1(\mathbf{0})} \nabla h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) \cdot \boldsymbol{\omega} \, d\omega - \frac{1}{4\pi c t^2} \int_{B_{ct}(\mathbf{x})} \Delta h \left(\mathbf{y} \right) d\mathbf{y}$$

$$+ \frac{1}{4\pi t} \int_{\partial B_{ct}(\mathbf{x})} \Delta h \left(\boldsymbol{\sigma} \right) d\sigma$$

$$= \frac{1}{4\pi t} \int_{\partial B_{ct}(\mathbf{x})} \Delta h \left(\boldsymbol{\sigma} \right) d\sigma.$$

On the other hand,

$$\Delta w_h \left(\mathbf{x},t \right) = \frac{t}{4\pi} \int_{\partial B_1(\mathbf{0})} \Delta h \left(\mathbf{x}+ct\boldsymbol{\omega} \right) d\omega = \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x})} \Delta h \left(\boldsymbol{\sigma} \right) d\sigma$$

and therefore

$$\partial_{tt} w_h - c^2 \Delta w_h = 0.$$

□

Using the calculations in the proof of the above theorem, we may write the Kirchhoff formula in the following form:

$$u \left(\mathbf{x},t \right) = \frac{1}{4\pi c^2 t^2} \int_{\partial B_{ct}(\mathbf{x})} \left\{ g \left(\boldsymbol{\sigma} \right) + \nabla g \left(\boldsymbol{\sigma} \right) \cdot \left(\boldsymbol{\sigma} - \mathbf{x} \right) + t h \left(\boldsymbol{\sigma} \right) \right\} d\sigma. \quad (5.122)$$

The presence of the gradient of g in (5.122) suggests that, unlike the one-dimensional case, the solution u *may be more irregular than the data.* Indeed, if $g \in C^k (\mathbb{R}^3)$ and $h \in C^{k-1} (\mathbb{R}^3)$, $k \geq 2$, then we can only guarantee that u is C^{k-1} and u_t is C^{k-2} at a later time.

Formula (5.122) makes perfect sense also for $g \in C^1 (\mathbb{R}^3)$ and h bounded. Clearly, under these weaker hypotheses, (5.122) satisfies the wave equation in an appropriate generalized sense, as in subsection 5.4.2, for instance.

In this case, scattered singularities in the initial data h may concentrate at later time on smaller sets, giving rise to stronger singularities (*focussing effect,* see problem 5.17).

According to (5.122), $u (\mathbf{x}, t)$ depends upon the data g and h only on the surface $\partial B_{ct} (\mathbf{x})$, which therefore coincides with **the domain of dependence for *(**x**,t)*.

Assume that the support of g and h is the compact set D. Then $u (\mathbf{x}, t)$ is different from zero only for $t_{\min} < t < t_{\max}$ where $t_{\min}$ and $t_{\max}$ are the *first* and the *last* time t such that $D \cap \partial B_{ct} (\mathbf{x}) \neq \emptyset$. In other words, a disturbance, initially localized inside D, starts affecting the point $\mathbf{x}$ at time $t_{\min}$ and ceases to affect it after time $t_{\max}$. This is another way to express the *strong Huygens' principle.*

Fix t and consider the union of all the spheres $\partial B_{ct} (\boldsymbol{\xi})$ as $\boldsymbol{\xi}$ varies on ∂D. The envelope of these surfaces constitutes the *wave front* and bounds the support of u, which spreads at speed c (see Problem 5.16).

5.9.3 Cauchy problem in dimension 2

The solution of the Cauchy problem in two dimensions can be obtained from Kirchhoff's formula, using the so called *Hadamard's method of descent.* Consider first the problem

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^2, \ t > 0 \\ w (\mathbf{x}, 0) = 0, \quad w_t (\mathbf{x}, 0) = h (\mathbf{x}) & \mathbf{x} \in \mathbb{R}^2. \end{cases} \tag{5.123}$$

The key idea is to "immerse" the two-dimensional problem (5.123) in a three-dimensional setting. More precisely, write points in $\mathbb{R}^3$ as $(\mathbf{x}, x_3)$ and set $h (\mathbf{x}, x_3) = h (\mathbf{x})$. The solution U of the three-dimensional problem is given by Kirchhoff formula:

$$U (\mathbf{x}, x_3, t) = \frac{1}{4\pi c^2 t} \int_{\partial B_{ct}(\mathbf{x}, x_3)} h \, d\sigma. \tag{5.124}$$

We claim that, since h does not depend on x_3, U is independent of x_3 as well, and therefore the solution of (5.123) is given by (5.124) with, say, $x_3 = 0$.

To prove the claim, note that the spherical surface $\partial B_{ct} (\mathbf{x}, x_3)$ is a union of the two hemispheres whose equation are

$$y_3 = F_\pm (y_1, y_2) = x_3 \pm \sqrt{c^2 t^2 - r^2},$$

where $r^2 = (y_1 - x_1)^2 + (y_2 - x_2)^2$. On both hemispheres we have:

$$d\sigma = \sqrt{1 + |\nabla F_\pm|^2} \, dy_1 dy_2$$

$$= \sqrt{1 + \frac{r^2}{c^2 t^2 - r^2}} \, dy_1 dy_2 = \frac{ct}{\sqrt{c^2 t^2 - r^2}} \, dy_1 dy_2$$

so that we may write $(d\mathbf{y} = dy_1 dy_2)$

$$U(\mathbf{x}, x_3, t) = \frac{1}{2\pi c} \int_{B_{ct}(\mathbf{x})} \frac{h(\mathbf{y})}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} d\mathbf{y}$$

and U is independent of x_3 as claimed. From the above calculations and recalling Lemma 5.2 we deduce the following theorem.

Theorem 5.5. (Poisson's formula). *Let* $g \in C^3(\mathbb{R}^2)$ *and* $h \in C^2(\mathbb{R}^2)$. *Then,*

$$u(\mathbf{x}, t) = \frac{1}{2\pi c} \left\{ \frac{\partial}{\partial t} \int_{B_{ct}(\mathbf{x})} \frac{g(\mathbf{y}) \, d\mathbf{y}}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} + \int_{B_{ct}(\mathbf{x})} \frac{h(\mathbf{y}) \, d\mathbf{y}}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} \right\}.$$

is the unique solution $u \in C^2(\mathbb{R}^2 \times [0, +\infty))$ *of the problem*

$$\begin{cases} u_{tt} - c^2 \Delta u = 0 & \mathbf{x} \in \mathbb{R}^2, t > 0 \\ u(\mathbf{x}, 0) = g(\mathbf{x}), \quad u_t(\mathbf{x}, 0) = h(\mathbf{x}) & \mathbf{x} \in \mathbb{R}^2. \end{cases}$$

Also Poisson's formula can be written in a somewhat more explicit form. Indeed, letting $\mathbf{y} - \mathbf{x} = ct\mathbf{z}$, we have

$$d\mathbf{y} = c^2 t^2 d\mathbf{z}, \quad |\mathbf{x} - \mathbf{y}|^2 = c^2 t^2 |\mathbf{z}|^2$$

whence

$$\int_{B_{ct}(\mathbf{x})} \frac{g(\mathbf{y})}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} d\mathbf{y} = ct \int_{B_1(\mathbf{0})} \frac{g(\mathbf{x} + ct\mathbf{z})}{\sqrt{1 - |\mathbf{z}|^2}} d\mathbf{z}.$$

Then

$$\frac{\partial}{\partial t} \int_{B_{ct}(\mathbf{x})} \frac{g(\mathbf{y})}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} d\mathbf{y}$$

$$= c \int_{B_1(\mathbf{0})} \frac{g(\mathbf{x} + ct\mathbf{z})}{\sqrt{1 - |\mathbf{z}|^2}} d\mathbf{z} + c^2 t \int_{B_1(\mathbf{0})} \frac{\nabla g(\mathbf{x} + ct\mathbf{z}) \cdot \mathbf{z}}{\sqrt{1 - |\mathbf{z}|^2}} d\mathbf{z}$$

and, going back to the original variables, we obtain

$$u(\mathbf{x}, t) = \frac{1}{2\pi ct} \int_{B_{ct}(\mathbf{x})} \frac{g(\mathbf{y}) + \nabla g(\mathbf{y}) \cdot (\mathbf{y} - \mathbf{x}) + th(\mathbf{y})}{\sqrt{c^2 t^2 - |\mathbf{x} - \mathbf{y}|^2}} d\mathbf{y}. \tag{5.125}$$

Poisson's formula displays an important difference with respect to its three-dimensional analogue, Kirkhhoff's formula. In fact *the domain of dependence* for the point $(\mathbf{x},t)$ is given by the **full circle** $B_{ct}(\mathbf{x}) = \{\mathbf{y}: |\mathbf{x} - \mathbf{y}| < ct\}$. This entails that a disturbance, initially localized at $\boldsymbol{\xi}$, starts affecting the point $\mathbf{x}$ at time $t_{\min} = |\mathbf{x} - \boldsymbol{\xi}|/c$. However, this effect does not vanish for $t > t_{\min}$, since $\boldsymbol{\xi}$ still belongs to the circle $B_{ct}(\mathbf{x})$ after $t_{\min}$.

It is the phenomenon one may observe by placing a cork on still water and dropping a stone not too far away. The cork remains undisturbed until it is reached by the wave front but its oscillations persist thereafter.

Thus, sharp signals do not exist in dimension two and *the strong Huygens principle does not hold.*

Remark 5.6. An examination of Poisson's formula reveals that the fundamental solution for the two dimensional wave equation is given by

$$K(\mathbf{x}, \mathbf{y}, t) = \frac{1}{2\pi c} \frac{\mathcal{H}(ct - r)}{\sqrt{c^2 t^2 - r^2}}$$

where $r^2 = |\mathbf{x} - \mathbf{y}|$ and $\mathcal{H}$ is the Heaviside function. For $\mathbf{y}$ fixed, its support is the **full** forward space-time cone, with vertex at $(\mathbf{y}, 0)$ and opening $\theta = \tan^{-1} c$, given by

$$C^*_{\mathbf{y},0} = \{(\mathbf{x}, t): |\mathbf{x} - \mathbf{y}| \leq ct, t > 0\}.$$

5.9.4 Non homogeneous equation. Retarded potentials

The solution of the non-homogeneous Cauchy problem can be obtained via Duhamel's method. We give the details for $n = 3$ only (for $n = 2$ see Problem 5.18). By linearity it is enough to derive a formula for the solution of the problem with zero initial data:

$$\begin{cases} u_{tt} - c^2 \Delta u = f(\mathbf{x},t) & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ u(\mathbf{x}, 0) = 0, \quad u_t(\mathbf{x},0) = 0 & \mathbf{x} \in \mathbb{R}^3. \end{cases} \tag{5.126}$$

Assume that $f \in C^2(\mathbb{R}^3 \times [0, +\infty))$. For $s \geq 0$ fixed, let $w = w(\mathbf{x},t; s)$ be the solution of the problem

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^3, t \geq s \\ w(\mathbf{x}, s; s) = 0, \quad w_t(\mathbf{x}, s; s) = f(\mathbf{x}, s) & \mathbf{x} \in \mathbb{R}^3. \end{cases}$$

Since the wave equation is invariant under time translations, w is given by Kirkhhoff's formula with t replaced by $t - s$:

$$w(\mathbf{x}, t; s) = \frac{1}{4\pi c^2 (t - s)} \int_{\partial B_{c(t-s)}(\mathbf{x})} f(\boldsymbol{\sigma}, s) \, d\sigma.$$

Then,

$$u\left(\mathbf{x}, t\right) = \int_0^t w\left(\mathbf{x}, t; s\right) \, ds = \frac{1}{4\pi c^2} \int_0^t \frac{ds}{(t-s)} \int_{\partial B_{c(t-s)}(\mathbf{x})} f\left(\boldsymbol{\sigma}, s\right) d\sigma \qquad (5.127)$$

is the unique solution $u \in C^2\left(\mathbb{R}^3 \times [0, +\infty)\right)$ of (5.126)[32].

Formula (5.126) shows that $u\left(\mathbf{x}, t\right)$ depends on the values of f in the full **backward** cone

$$C_{\mathbf{x}, t} = \{(\mathbf{z}, s): |\mathbf{z} - \mathbf{x}| \leq c(t-s), \, 0 \leq s \leq t\}.$$

Note that (5.126) may be written in the form

$$u\left(\mathbf{x}, t\right) = \frac{1}{4\pi} \int_{B_{ct}(\mathbf{x})} \frac{1}{|\mathbf{x} - \mathbf{y}|} f\left(\mathbf{y}, t - \frac{|\mathbf{x} - \mathbf{y}|}{c}\right) d\mathbf{y} \qquad (5.128)$$

which is a so called *retarded potential*. Indeed, $u\left(\mathbf{x}, t\right)$ depends on the values of the source f at the earlier times

$$t' = t - \frac{|\mathbf{x} - \mathbf{y}|}{c}.$$

5.10 Linear Water Waves

A great variety of interesting phenomena occurs in the analysis of water waves. Here we briefly analyze *surface water waves*, that is disturbances of the free surface of an incompressible fluid, resulting from the balance between a restoring force, due to gravity and/or surface tension, and fluid inertia due to an external action (such as wind, passage of a ship, sub-sea earthquakes). We will focus on the special case of *linear waves*, whose amplitude is small compared to wavelength, analyzing the dispersive relations in the approximation of deep water.

5.10.1 A model for surface waves

We start deriving a basic model for surface water waves, assuming the following hypotheses:

1. The fluid has *constant density* ρ and *negligible viscosity*. In particular, the force exerted on a control fluid volume V by the rest of the fluid is given by the normal pressure[33] $-p\boldsymbol{\nu}$ on ∂V.
2. The motion is *laminar* (no breaking waves or turbulence) and *two dimensional*. This means that in a suitable coordinate system x, z, where the coordinate x measures horizontal distance and z is a vertical coordinate, we can describe the free surface by a function $z = h\left(x, t\right)$, while the velocity vector has the form $\mathbf{w} = u\left(x, z, t\right)\mathbf{i} + v\left(x, z, t\right)\mathbf{k}$.

---

[32] Check it, mimicking the proof in dimension one (subsection 5.4.4).
[33] $\boldsymbol{\nu}$ is the exterior normal unit vector to ∂V.

3. The motion is **irrotational**, so that *there exists a (smooth) scalar potential* $\phi = \phi(x, z, t)$ such that:

$$\mathbf{w} = \nabla\phi = \phi_x\mathbf{i} + \phi_z\mathbf{k}.$$

We need equations for the unknowns h and ϕ, together with *initial conditions* and
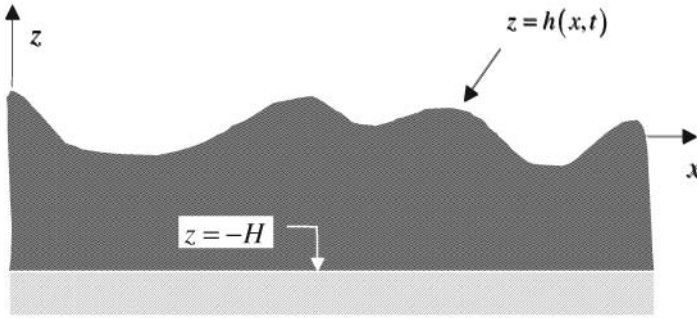
Fig. 5.11. Vertical section of the fluid region

suitable *conditions at the boundary* of our relevant domain, composed by the free surface, the lower boundary and the lateral sides.

We assume that the side boundaries are so far apart that their influence can be neglected. Therefore x varies all along the real axis.

Furthermore, we assume, for simplicity, that the lower boundary is flat, at the level $z = -H$.

Two equations for h and ϕ come from conservation of mass and balance of linear momentum, taking into account hypotheses 2 and 3 above.

Mass conservation gives:

$$\text{div } \mathbf{w} = \Delta\phi = 0 \qquad x \in \mathbb{R}, \; -H < z < h(x, t). \tag{5.129}$$

Thus, ϕ is a harmonic function.

Balance of linear momentum yields:

$$\mathbf{w}_t + (\mathbf{w}\cdot\nabla)\mathbf{w} = \mathbf{g} - \frac{1}{\rho}\nabla p \tag{5.130}$$

where $\mathbf{g}$ is the gravitational acceleration.

Let us rewrite (5.130) in terms of the potential ϕ. From the identity

$$\mathbf{w} \times \text{curl } \mathbf{w} = \frac{1}{2}\nabla(|\mathbf{w}|^2) - (\mathbf{w}\cdot\nabla)\mathbf{w}$$

we get, being curl $\mathbf{w} = \mathbf{0}$,

$$(\mathbf{w}\cdot\nabla)\mathbf{w} = \frac{1}{2}\nabla(|\nabla\phi|^2).$$

Moreover, writing $\mathbf{g} = \nabla(-gz)$, (5.130) becomes

$$\frac{\partial}{\partial t}(\nabla\phi) + \frac{1}{2}\nabla(|\nabla\phi|^2) = -\frac{1}{\rho}\nabla p + \nabla(-gz)$$

or

$$\nabla\left\{\phi_t + \frac{1}{2}|\nabla\phi|^2 + \frac{p}{\rho} + gz\right\} = 0.$$

As a consequence

$$\phi_t + \frac{1}{2}|\nabla\phi|^2 + \frac{p}{\rho} + gz = C(t)$$

with $C = C(t)$ is an arbitrary function. Since ϕ is uniquely defined up to an additive function of time, we can choose $C(t) = 0$ by adding to ϕ the function $\int_0^t C(s)\,ds$.

In this case, we obtain **Bernoulli's equation**

$$\phi_t + \frac{1}{2}|\nabla\phi|^2 + \frac{p}{\rho} + gz = 0. \tag{5.131}$$

We consider now the boundary conditions. On the bottom, we impose the so called **bed condition**, according to which the normal component of the velocity vanishes there; therefore

$$\phi_z(x, -H, t) = 0, \qquad x \in \mathbb{R}. \tag{5.132}$$

More delicate is the condition on the free surface $z = h(x, t)$; in fact, since this surface is itself an unknown of the problem, we actually need *two conditions* on it.

The first one comes from Bernoulli's equation. Namely, the total pressure on the free surface is given by

$$p = p_{at} - \sigma h_{xx}\left\{1 + h_x^2\right\}^{-3/2}. \tag{5.133}$$

In (5.133) the term p_{at} is the atmospheric pressure, that we can take equal to zero, while the second term is due to the *surface tension,* as we will shortly see below.

Thus, inserting $z = h(x, t)$ and (5.133) into (5.131), we obtain the following **dynamic condition at the free surface**:

$$\phi_t + \frac{1}{2}|\nabla\phi|^2 - \frac{\sigma h_{xx}}{\rho\left\{1 + h_x^2\right\}^{3/2}} + gh = 0, \qquad x \in \mathbb{R}, z = h(x, t). \tag{5.134}$$

A second condition follows imposing that fluid particles on the free surface always remain there. If the particle path is described by the equations $x = x(t)$, $z = z(t)$, this amounts to requiring that

$$z(t) - h(x(t), t) \equiv 0.$$

Differentiating yields

$$\dot{z}(t) - h_x(x(t), t)\dot{x}(t) - h_t(x(t), t) = 0$$

that is, since $\dot{x}(t) = \phi_x(x(t), z(t), t)$ and $\dot{z} = \phi_z(x(t), z(t), t)$,

$$\phi_z - h_t - \phi_x h_x = 0, \qquad x \in \mathbb{R}, \ z = h(x,t). \tag{5.135}$$

which is known as the **kinematic condition at the free surface**.

Finally, we require a reasonable behavior of ϕ and h as $x \to \pm\infty$, for instance

$$\int_{\mathbb{R}} |\phi| < \infty, \ \int_{\mathbb{R}} |h| < \infty \quad \text{and} \quad \phi, \, h \to 0 \ \text{ as } x \to \pm\infty. \tag{5.136}$$

Equation (5.129) and the boundary conditions (5.132), (5.134), (5.135) constitute our model for water waves. After a brief justification of formula (5.133), in the next subsection we go back to the above model, deriving a dimensionless formulation and a linearized version of it.

• *Effect of surface tension.* In a water molecule the two hydrogen atoms take an asymmetric position with respect to the oxygen atom. This asymmetric structure generates an electric dipole moment. Inside a bulk of water these moments balance, but on the surface they tend to be parallel and create a macroscopic inter-molecular force per unit length, confined to the surface, called *surface tension.*

The way this force manifests itself is similar to the action exerted on a small portion of an elastic material by the surrounding material and described by a *stress vector*, which is a force per unit area, on the boundary of the portion. Analogously, consider a small region on the water surface, delimited by a closed curve γ. The surface water on one side of γ exerts on the water on the other side a *force per unit length* $\mathbf{f}$ (*pulling*) along γ.

Let $\mathbf{n}$ be a unit vector normal to the water surface and $\boldsymbol{\tau}$ a unit tangent vector to γ (Fig. 5.12a) so chosen that $\mathbf{N} = \boldsymbol{\tau} \times \mathbf{n}$ points outwards the region bounded by γ. A simple constitutive law for $\mathbf{f}$ is

$$\mathbf{f}(\mathbf{x},t) = \sigma(\mathbf{x},t)\,\mathbf{N}(\mathbf{x},t) \qquad \mathbf{x} \in \gamma.$$

Thus, $\mathbf{f}$ acts in the direction of $\mathbf{N}$; its magnitude σ, independent of $\mathbf{N}$, is called **surface tension**.

Formula (5.133) is obtained by balancing the net vertical component of the force produced by surface tension with the difference of the pressure force across the surface.

Consider the section ds of a small surface element shown in figure 5.12b. A surface tension of magnitude σ acts tangentially at both ends. Up to higher order terms, the downward vertical component is given by $2\sigma \sin(\alpha/2)$. On the other hand, this force is equal to $(p_{at} - p)ds$ where p is the fluid pressure beneath the surface. Thus,

$$(p_{at} - p)ds = 2\sigma \sin(\alpha/2).$$

Since for small α we have $ds \approx R d\alpha$ and $2\sin(\alpha/2) \approx \alpha$, we may write

$$p_{at} - p = \frac{\sigma}{R} = \sigma\kappa \tag{5.137}$$
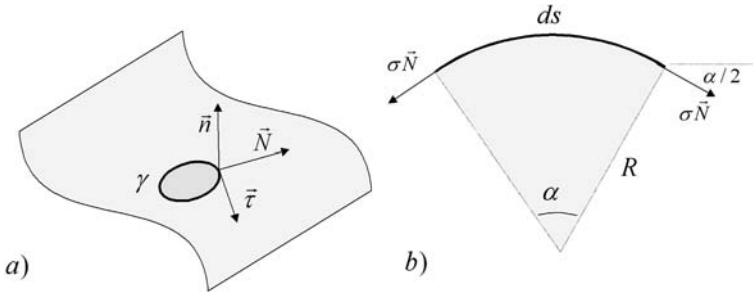
Fig. 5.12. Surface tension $\sigma\mathbf{N}$

where $\kappa = R^{-1}$ is the **curvature of the surface**. If the atmospheric pressure prevails, the curvature is positive and the surface is convex, otherwise the curvature is negative and the surface is concave, as in figure 5.12b. If the surface is described by $z = h(x,t)$, we have

$$\kappa = \frac{h_{xx}}{\{1 + h_x^2\}^{3/2}}$$

which inserted into (5.137) gives (5.133).

5.10.2 Dimensionless formulation and linearization

The nonlinearities in (5.134), (5.135) and the fact that the free surface is an unknown make the above model quite difficult to analyze by elementary means. However, if we restrict our considerations to waves whose amplitude is much smaller than their wavelength, then both difficulties disappear. In spite of this simplification, the resulting theory has a rather wide range of applications, since it is not rare to observe waves with amplitude from 1 to 2 meters and a wavelength of up to a kilometer or more.

To perform a correct linearization procedure, we first introduce *dimensionless* variables. Denote by L, A and T, an average *wavelength, amplitude* and *period*[34], respectively. Set

$$\tau = \frac{t}{T}, \qquad \xi = \frac{x}{L}, \qquad \eta = \frac{z}{L}.$$

Since the dimensions of h and ϕ are, respectively $[length]$ and $[length]^2 \times [time]^{-1}$, we may rescale ϕ and h by setting:

$$\Phi(\xi, \eta, \tau) = \frac{T}{LA}\phi(L\xi, L\eta, T\tau), \qquad \Gamma(\xi, \tau) = \frac{1}{A}h(L\xi, L\eta, T\tau).$$

---

[34] The time a crest takes to travel a distance of order L.

In terms of these dimensionless variables, our model becomes, after elementary calculations:

$$\Delta\Phi = 0, \qquad\qquad\qquad\qquad\qquad - H_0 < \eta < \varepsilon\Gamma(\xi,\tau),\ \xi \in \mathbb{R}$$

$$\Phi_\tau + \tfrac{\varepsilon}{2}|\nabla\Phi|^2 + \mathcal{F}\left\{\Gamma - \mathcal{B}\Gamma_{\xi\xi}\left\{1 + \varepsilon^2\Gamma_\xi^2\right\}^{3/2}\right\} = 0, \quad \eta = \varepsilon\Gamma(\xi,\tau), \xi \in \mathbb{R}$$

$$\Phi_\eta - \Gamma_\tau - \varepsilon\Phi_\xi\Gamma_\xi = 0, \qquad\qquad\qquad \eta = \varepsilon\Gamma(\xi,\tau),\ \xi \in \mathbb{R}$$

$$\Phi_\eta(\xi, -H_0, \tau) = 0, \qquad\qquad\qquad\qquad \xi \in \mathbb{R},$$

where we have emphasized the four dimensionless combinations[35]

$$\varepsilon = \frac{A}{L}, \qquad H_0 = \frac{H}{L}, \qquad \mathcal{F} = \frac{gT^2}{L}, \qquad \mathcal{B} = \frac{\sigma}{\rho g L^2}. \qquad (5.138)$$

The parameter $\mathcal{B}$, called *Bond number,* measures the importance of surface tension while $\mathcal{F}$, the *Froude number,* measures the importance of gravity.

At this point, the assumption of *small amplitude compared to the wavelength,* translates simply into

$$\varepsilon = \frac{A}{L} \ll 1$$

and the linearization of the above system is achieved by letting $\varepsilon = 0$:

$$\Delta\Phi = 0, \qquad\qquad - H_0 < \eta < 0,\ \xi \in \mathbb{R}$$

$$\Phi_\tau + \mathcal{F}\{\Gamma - \mathcal{B}\Gamma_{\xi\xi}\} = 0, \qquad \eta = 0, \xi \in \mathbb{R}$$

$$\Phi_\eta - \Gamma_\tau = 0, \qquad\qquad \eta = 0,\ \xi \in \mathbb{R}$$

$$\Phi_\eta(\xi, -H_0, \tau) = 0, \qquad\qquad \xi \in \mathbb{R}.$$

Going back to the original variables, we finally obtain the linearized system

$$\begin{cases} \Delta\phi = 0, & - H < z < 0,\ x \in \mathbb{R} & \text{(Laplace)} \\ \phi_t + gh - \frac{\sigma}{\rho}h_{xx} = 0, & z = 0,\ x \in \mathbb{R} & \text{(Bernoulli)} \\ \phi_z - h_t = 0, & z = 0,\ x \in \mathbb{R} & \text{(kinematic)} \\ \phi_z(x, -H, t) = 0, & x \in \mathbb{R} & \text{(bed condition)} \end{cases} \qquad (5.139)$$

It is possible to obtain an equation for ϕ only. Differentiate twice with respect to x the kinematic equation and use $\phi_{xx} = -\phi_{zz}$; this yields

$$h_{txx} = \phi_{zxx} = -\phi_{zzz}. \qquad (5.140)$$

Differentiate Bernoulli's equation with respect to t, then use $h_t = \phi_z$ and (5.140). The result is:

$$\phi_{tt} + g\phi_z + \frac{\sigma}{\rho}\phi_{zzz} = 0, \qquad z = 0,\ x \in \mathbb{R}. \qquad (5.141)$$

---

[35] Note the reduction of the number of relevant parameters from seven $(A, L, T, H, g, \sigma, \rho)$ to four.

5.10.3 Deep water waves

We solve now system (5.139) with the following initial conditions:

$$\phi(x, z, 0) = 0, \quad h(x, 0) = h_0(x), \quad h_t(x, 0) = 0. \tag{5.142}$$

Thus, initially ($t = 0$) the fluid velocity is zero and the free surface has been perturbed into a non horizontal profile h_0, that we assume (for simplicity) *smooth, even* (i.e. $h_0(-x) = h_0(x)$) and *compactly supported*. In addition we consider the case of *deep water* ($H \gg 1$) so that the bed condition can be replaced by[36]

$$\phi_z(x, z, t) \to 0 \quad \text{as } z \to -\infty. \tag{5.143}$$

The resulting initial-boundary value problem is not of the type we considered so far, but we are reasonably confident that it is well posed. Since x varies over all the real axis, we may use the Fourier transform with respect to x, setting

$$\widehat{\phi}(k, z, t) = \int_{\mathbb{R}} e^{-ikx} \phi(x, z, t)\, dx, \qquad \widehat{h}(k, t) = \int_{\mathbb{R}} e^{-ikx} h(x, t)\, dx.$$

Note that, the assumptions on h_0 implies that $\widehat{h}_0(k) = \widehat{h}_0(k, 0)$ rapidly vanishes as $|k| \to \infty$ and $\widehat{h}_0(-k) = \widehat{h}_0(k)$. Moreover, since $\widehat{\phi}_{xx} = -k^2 \widehat{\phi}$, the Laplace equation transforms into the ordinary differential equation

$$\widehat{\phi}_{zz} - k^2 \widehat{\phi} = 0$$

whose general solution is

$$\widehat{\phi}(k, z, t) = A(k, t)\, e^{|k|z} + B(k, t)\, e^{-|k|z}.$$

From (5.143) we deduce $B(k, t) = 0$, so that

$$\widehat{\phi}(k, z, t) = A(k, t)\, e^{|k|z}. \tag{5.144}$$

Trasforming (5.141) we get

$$\widehat{\phi}_{tt} + g\widehat{\phi}_z + \frac{\sigma}{\rho}\widehat{\phi}_{zzz} = 0, \qquad z = 0,\, k \in \mathbb{R}$$

and (5.144) yields for A the equation

$$A_{tt} + \left(g\,|k| + \frac{\sigma}{\rho}\,|k|^3 \right) A = 0.$$

Thus, we obtain

$$A(k, t) = a(k)\, e^{i\omega t} + b(k)\, e^{-i\omega t}$$

---

[36] For the case of finite depth see Problem 5.19.

where (**dispersion relation**)

$$\omega\left(k\right) = \sqrt{g\left|k\right| + \frac{\sigma}{\rho}\left|k\right|^3},$$

and

$$\widehat{\phi}\left(k, z, t\right) = \left\{a\left(k\right)e^{i\omega\left(k\right)t} + b\left(k\right)e^{-i\omega\left(k\right)t}\right\}e^{\left|k\right|z}.$$

To determine $a\left(k\right)$ e $b\left(k\right)$, observe that the Bernoulli condition gives

$$\widehat{\phi}_t\left(k, 0, t\right) + \left\{g + \frac{\sigma}{\rho}k^2\right\}\widehat{h}\left(k, t\right) = 0, \qquad k \in \mathbb{R} \qquad (5.145)$$

from which

$$i\omega\left(k\right)\left\{a\left(k\right)e^{i\omega\left(k\right)t} - b\left(k\right)e^{-i\omega\left(k\right)t}\right\} + \left(g + \frac{\sigma}{\rho}k^2\right)\widehat{h}\left(k, t\right) = 0, \qquad k \in \mathbb{R}$$

and for $t = 0$

$$i\omega\left(k\right)\left\{a\left(k\right) - b\left(k\right)\right\} + \left(g + \frac{\sigma}{\rho}k^2\right)\widehat{h}_0\left(k\right) = 0. \qquad (5.146)$$

Similarly, the kinematic condition gives

$$\widehat{\phi}_z\left(k, 0, t\right) + \widehat{h}_t\left(k, t\right) = 0, \qquad k \in \mathbb{R}. \qquad (5.147)$$

We have

$$\widehat{\phi}_z\left(k, 0, t\right) = \left|k\right|\left\{a\left(k\right)e^{i\omega\left(k\right)t} + b\left(k\right)e^{-i\omega\left(k\right)t}\right\}$$

and since $\widehat{h}_t\left(k, 0\right) = 0$, we get, from (5.147) for $t = 0$ and $k \neq 0$,

$$a\left(k\right) + b\left(k\right) = 0. \qquad (5.148)$$

From (5.146) and (5.148) we have ($k \neq 0$)

$$a\left(k\right) = -b\left(k\right) = \frac{i\left(g + \frac{\sigma}{\rho}k^2\right)}{2\omega\left(k\right)}\widehat{h}_0\left(k\right)$$

and therefore

$$\widehat{\phi}\left(k, y, t\right) = \frac{i\left(g + \frac{\sigma}{\rho}k^2\right)}{2\omega\left(k\right)}\left\{e^{i\omega\left(k\right)t} - e^{-i\omega\left(k\right)t}\right\}e^{\left|k\right|z}\widehat{h}_0\left(k\right).$$

From (5.145) we deduce:

$$\widehat{h}\left(k, t\right) = \left(g + \frac{\sigma}{\rho}k^2\right)^{-1}\widehat{\phi}_t\left(k, 0, t\right) = \frac{1}{2}\left\{e^{i\omega\left(k\right)t} + e^{-i\omega\left(k\right)t}\right\}\widehat{h}_0\left(k\right)$$

and finally, transforming back[37]

$$h\left(x,t\right)=\frac{1}{4\pi}\int_{\mathbb{R}}\left\{e^{i(kx-\omega(k)t)}+e^{i(kx+\omega(k)t)}\right\}\widehat{h}_{0}\left(k\right)\ dk.\qquad(5.149)$$

5.10.4 Interpretation of the solution

The surface displacement appears in **wave packet** form. The dispersion relation

$$\omega\left(k\right)=\sqrt{g\left|k\right|+\frac{\sigma}{\rho}\left|k\right|^{3}}$$

shows that each Fourier component of the initial free surface propagates both in the positive and negative $x-$directions. The phase and group velocities are (considering only $k>0$, for simplicity)

$$c_{p}=\frac{\omega}{k}=\sqrt{\frac{g}{k}+\frac{\sigma k}{\rho}}$$

and

$$c_{g}=\frac{g+3\sigma k^{2}/\rho}{2\sqrt{gk+\sigma k^{3}/\rho}}.$$

Thus, we see that the speed of a wave of wavelength $\lambda=2\pi/k$ **depends on its wavelength**. The fundamental parameter is

$$B^{*}=4\pi^{2}\mathcal{B}{=}\frac{\sigma k^{2}}{\rho g}$$

where $\mathcal{B}$ is the Bond number. For water, under "normal" conditions,

$$\rho=1\ \mathrm{gr/cm^{3}},\ \sigma=72\ \mathrm{gr/sec^{2}},\ g=980\ \mathrm{cm/sec^{2}}\qquad(5.150)$$

so that $B^{*}=1$ for wavelengths $\lambda\simeq1.7$ cm. When $\lambda\gg1.7$ cm, then $B^{*}<1$, $k=\frac{2\pi}{\lambda}\ll1$ and surface tension becomes negligible. This is the case of **gravity waves** (generated e.g. by dropping a stone into a pond) whose phase speed is well approximated by

$$c_{p}=\sqrt{\frac{g}{k}}=\sqrt{\frac{g\lambda}{2\pi}}$$

while their group velocity is

$$c_{g}=\frac{1}{2}\sqrt{\frac{g}{k}}=\frac{1}{2}c_{p}.$$

---

[37] Note that, since also $\omega\left(k\right)$ is even, we may write

$$h\left(x,t\right)=\frac{1}{2\pi}\int_{\mathbb{R}}\cos\left[kx-\omega\left(k\right)t\right]\hat{h}_{0}\left(k\right)dk.$$

Thus, **longer waves move faster** and **energy is slower than the crests**.

On the other hand, if $\lambda \ll 1.7$ cm, then $B^* > 1$, $k = \frac{2\pi}{\lambda} \gg 1$ and this time surface tension prevails over gravity. In fact, short wavelengths are associated with relative high curvature of the free surface and high curvature is concomitant with large surface tension effects. This is the case of **capillarity waves** (generated e.g. by raindrops in a pond) and their speed is well approximated by

$$c_p = \sqrt{\frac{\sigma k}{\rho}} = \sqrt{\frac{2\pi\sigma}{\lambda\rho}}$$

while the group velocity is

$$c_g = \frac{3}{2}\sqrt{\frac{\sigma k}{\rho}} = \frac{3}{2}c_p.$$

Thus **shorter waves move faster** and **energy is faster than the crests**.

When both gravity and surface tension are relevant, figure 5.13 shows the graph of c_p^2 versus λ, for water, with the values (5.150):

$$c_p^2 = 156.\,97\,\lambda + \frac{452.\,39}{\lambda}.$$

The main feature of this graph is the presence of the minimum

$$c_{\min} = 23 \text{ cm/sec}$$

corresponding just to the value $\lambda = 1.7$ cm. The consequence is curious: linear gravity and capillarity deep water waves can appear simultaneously only when the speed is greater than 23 cm/sec. A typical situation occurs when a small obstacle (e.g. a twig) moves at speed v in still water. The motion of the twig results in the formation of a wave system that moves along with it, with gravity waves behind and capillarity waves ahead. In fact, the result above shows that this wave system can actually appear only if $v > 23$ cm/sec.
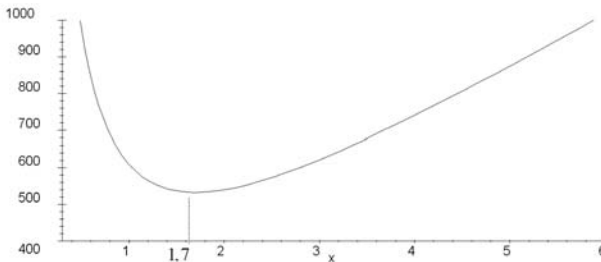
Fig. 5.13. c_p^2 versus λ

5.10.5 Asymptotic behavior

As we have already observed, the behavior of a wave packet is dominated for short times by the initial conditions and only after a relatively long time it is possible to observe the intrinsic features of the perturbation. For this reason, information about the asymptotic behavior of the packet as $t \to +\infty$ are important. Thus, we need a good asymptotic formula for the integral in (5.149) when $t \gg 1$.

For simplicity, consider gravity waves only, for which

$$\omega(k) = \sqrt{g\,|k|}.$$

Let us follow a particle $x = x(t)$ moving along the positive $x-$direction with constant speed $v > 0$, so that $x = vt$. Inserting $x = vt$ into (5.149) we find

$$h(vt, t) = \frac{1}{4\pi} \int_{\mathbb{R}} e^{it(kv-\omega(k))} \widehat{h}_0(k) \; dk + \frac{1}{4\pi} \int_{\mathbb{R}} e^{it(kv+\omega(k))} \widehat{h}_0(k) \; dk$$
$$\equiv h_1(vt, t) + h_2(vt, t).$$

According to Theorem 5.6 in the next subsection (see also Remark 5.10), with

$$\varphi(k) = kv - \omega(k),$$

if there exists exactly one stationary point for φ, i.e. only one point k_0 such that

$$\omega'(k_0) = v \quad \text{and} \quad \varphi''(k_0) = -\omega''(k_0) \neq 0,$$

we may estimate h_1 for $t \gg 1$ by the following formula:

$$h_1(vt, t) = \frac{A(k_0)}{t} \exp\{it[k_0 v - \omega(k_0)]\} + O(t^{-1}) \tag{5.151}$$

where

$$A(k_0) = \widehat{h}_0(k_0) \sqrt{\frac{1}{8\pi\,|\omega''(k_0)|}} \exp i\left\{-\frac{\pi}{4}\text{sign}\,\omega''(k_0)\right\}.$$

We have

$$\omega'(k) = \frac{1}{2}\sqrt{g}\,|k|^{-1/2}\,\text{sign}(k)$$

and

$$\omega''(k) = -\frac{\sqrt{g}}{4}\,|k|^{-3/2}.$$

Since $v > 0$, equation $\omega'(k_0) = v$ gives the unique *point of stationary phase*

$$k_0 = \frac{g}{4v^2} = \frac{gt^2}{4x^2}.$$

Moreover,

$$k_0 v - \omega(k_0) = -\frac{g}{4v} = -\frac{gt}{4x}$$

and

$$\omega''(k_0) = -\frac{2v^3}{g} = -\frac{2x^3}{gt^3} < 0$$

so that from (5.151) we find

$$h_1(vt,t) = \frac{1}{4}\widehat{h}_0\left(\frac{g}{4v^2}\right)\sqrt{\frac{g}{\pi tv^3}}\exp i\left\{-\frac{gt}{4v} + \frac{\pi}{4}\right\} + O\left(t^{-1}\right)$$

Similarly, since

$$\widehat{h}_0(k_0) = \widehat{h}_0(-k_0),$$

we find

$$h_2(vt,t) = \frac{1}{4}\widehat{h}_0\left(\frac{g}{4v^2}\right)\sqrt{\frac{g}{\pi tv^3}}\exp i\left\{\frac{gt}{4v} - \frac{\pi}{4}\right\} + O\left(t^{-1}\right).$$

Finally,

$$h(vt,t) = h_1(vt,t) + h_2(vt,t)$$
$$= \widehat{h}_0\left(\frac{g}{4v^2}\right)\sqrt{\frac{g}{4\pi v^3 t}}\cos\left\{\frac{gt}{4v} - \frac{\pi}{4}\right\} + O\left(t^{-1}\right).$$

This formula shows that, for large x and t, with $x/t = v$, constant, the wave packet is locally sinusoidal with wave number

$$k(x,t) = \frac{gt}{4vx} = \frac{gt^2}{4x^2}.$$

In other words, an observer moving at the constant speed $v = x/t$ sees a dominant wavelength $2\pi/k_0$, where k_0 is the solution of $\omega'(k_0) = x/t$. The amplitude decreases as $t^{-1/2}$. This is due to the dispersion of the various Fourier components of the initial configuration, after a sufficiently long time.

5.10.6 The method of stationary phase

The *method of stationary phase*, essentially due to Laplace, gives an asymptotic formula for integrals of the form

$$I(t) = \int_a^b f(k)\,e^{it\varphi(k)}dk \qquad (-\infty \le a < b \le \infty)$$

as $t \to +\infty$. Actually, only the real part of $I(t)$, in which the factor $\cos[t\varphi(k)]$ appears, is of interest. Now, as t increases and $\varphi(k)$ varies, $\cos[t\varphi(k)]$ oscillates more and more and eventually much more than f. For this reason, the contributions of the intervals where $\cos[t\varphi(k)] > 0$ will balance those in which $\cos[t\varphi(k)] < 0$, so that we expect that $I(t) \to 0$ as $t \to +\infty$, just as the Fourier coefficients of an integrable function tend to zero as the frequency goes to infinity.

To obtain information on the vanishing speed, assume φ is constant on a certain interval J. On this interval $\cos[t\varphi(k)]$ is constant as well and hence there are neither oscillations nor cancellations. Thus, it is reasonable that, for $t \gg 1$, the relevant contributions to $I(t)$ come from intervals where φ is constant or at least almost constant. The same argument suggests that eventually, a however small interval, containing a stationary point k_0 for φ, will contribute to the integral much more than any other interval without stationary points.

The method of stationary phase makes the above argument precise through the following theorem.

Theorem 5.6. *Let f and φ belong to $C^2([a, b])$. Assume that*

$$\varphi'(k_0) = 0, \varphi''(k_0) \neq 0 \qquad and \qquad \varphi'(k) \neq 0 \text{ for } k \neq k_0.$$

Then, as $t \to +\infty$

$$\int_a^b f(k)\, e^{it\varphi(k)} dk = \sqrt{\frac{2\pi}{|\varphi''(k_0)|}} \frac{f(k_0)}{\sqrt{t}} \exp\left\{i\left[t\varphi(k_0) + \frac{\pi}{4}\operatorname{sign}\varphi''(k_0)\right]\right\} + O\left(t^{-1}\right)$$

First a lemma.

Lemma 5.3. *Let f, φ as in Theorem 5.6. Let $[c, d] \subseteq [a, b]$ and assume that $|\varphi'(k)| \geq C > 0$ in (c, d). Then*

$$\int_c^d f(k)\, e^{it\varphi(k)} dk = O\left(t^{-1}\right) \qquad t \to +\infty. \tag{5.152}$$

Proof. Integrating by parts we get (multiplying and dividing by φ'):

$$\int_c^d \frac{f}{\varphi'}\varphi' e^{it\varphi} dk = \frac{1}{it}\left\{\frac{f(d)\,e^{it\varphi(d)}}{\varphi'(d)} - \frac{f(c)\,e^{it\varphi(c)}}{\varphi'(c)} - \int_c^d \frac{f'\varphi' - f\varphi''}{(\varphi')^2} e^{it\varphi} dk\right\}.$$

Thus, from $\left|e^{it\varphi(k)}\right| \leq 1$ and our hypotheses, we have

$$\left|\int_c^d f\, e^{it\varphi} dk\right| \leq \frac{1}{Ct}\left\{|f(d)| + |f(c)| + \frac{1}{C}\int_c^d |f'\varphi' - f\varphi''|\, dk\right\}$$

$$\leq \frac{K}{t}$$

which gives (5.152). $\square$

Proof of Theorem 5.6. Without loss of generality, we may assume $k_0 = 0$, so that $\varphi'(0) = 0$, $\varphi''(0) \neq 0$. From Lemma 5.3, it is enough to consider the integral

$$\int_{-\varepsilon}^\varepsilon f(k)\, e^{it\varphi(k)} dk$$

where $\varepsilon > 0$ is as small as we wish. We distinguish two cases.

Case 1: φ is a quadratic polynomial, that is

$$\varphi(k) = \varphi(0) + Ak^2, \qquad A = \frac{1}{2}\varphi''(0).$$

Then, write

$$f(k) = f(0) + \frac{f(k) - f(0)}{k}k \equiv f(0) + q(k)k,$$

and observe that, since $f \in C^2([-\varepsilon, \varepsilon])$, $q'(k)$ is bounded in $[-\varepsilon, \varepsilon]$. Then, we have:

$$\int_{-\varepsilon}^{\varepsilon} f(k)e^{it\varphi(k)}dk = 2f(0)e^{it\varphi(0)}\int_0^{\varepsilon} e^{itAk^2}dk + e^{it\varphi(0)}\int_{-\varepsilon}^{\varepsilon} q(k)ke^{itAk^2}dk.$$

Now, an integration by parts shows that the second integral is $O(1/t)$ as $t \to \infty$ (the reader should check the details).

In the first integral, if $A > 0$, let

$$tAk^2 = y^2.$$

Then

$$\int_0^{\varepsilon} e^{itAk^2}dk = \frac{1}{\sqrt{tA}}\int_0^{\varepsilon\sqrt{tA}} e^{iy^2}dy.$$

Since[38]

$$\int_0^{\varepsilon\sqrt{tA}} e^{iy^2}dy = \frac{\sqrt{\pi}}{2}e^{i\frac{\pi}{4}} + O\left(\frac{1}{\varepsilon\sqrt{tA}}\right),$$

we get

$$\int_0^{\varepsilon} f(k)e^{it\varphi(k)}dk = \sqrt{\frac{2\pi}{|\varphi''(0)|}}\frac{f(0)}{\sqrt{t}}\exp\left\{i\left[\varphi(0)t + \frac{\pi}{4}\right]\right\} + O\left(\frac{1}{t}\right),$$

which proves the theorem when $A > 0$. The proof is similar if $A < 0$.

Case 2. General φ. By a suitable change of variable we reduce case 2 to case 1. First we write

$$\varphi(k) = \varphi(0) + \frac{1}{2}a(k)k^2 \qquad (5.153)$$

where

$$a(k) = 2\int_0^1 (1-r)\varphi''(rk)dr.$$

---

[38] Recall that $e^{i\pi/4} = \left(\sqrt{2} + i\sqrt{2}\right)/2$. Moreover, the following formulas hold:

$$\left|\frac{\sqrt{\pi}}{2\sqrt{2}} - \int_0^{\lambda} \cos(y^2)dy\right| \le \frac{\sqrt{\pi}}{\lambda}$$

$$\left|\frac{\sqrt{\pi}}{2\sqrt{2}} - \int_0^{\lambda} \sin(y^2)dy\right| \le \frac{\sqrt{\pi}}{\lambda}.$$

Equation (5.153) follows by applying to $\psi(s) = \varphi(sk)$ the following Taylor formula:

$$\psi(1) = \psi(0) + \psi'(0)s + \frac{1}{2}\int_0^1 (1-r)\psi''(r)\,dr.$$

Note that $a(0) = \varphi''(0)$. Consider the function

$$p(k) = k\sqrt{a(k)/\varphi''(0)}.$$

We have $p(0) = 0$ and $p'(0) = 1$. Therefore, p is invertible near zero. Let

$$k = p^{-1}(y).$$

Then, since

$$\varphi(k) = \varphi(0) + \frac{\varphi''(0)}{2}[p(k)]^2,$$

we have,

$$\tilde{\varphi}(y) \equiv \varphi\left(p^{-1}(y)\right)$$
$$= \varphi(0) + \frac{\varphi''(0)}{2}\left[p\left(p^{-1}(y)\right)\right]^2$$
$$= \varphi(0) + \frac{\varphi''(0)}{2}y^2$$

and

$$\int_{-\varepsilon}^{\varepsilon} f(k)e^{it\varphi(k)}dk = \int_{p^{-1}(-\varepsilon)}^{p^{-1}(\varepsilon)} F(y)e^{it\tilde{\varphi}(y)}dy$$

where

$$F(y) = \frac{f\left(p^{-1}(y)\right)}{p'\left(p^{-1}(y)\right)}.$$

Since $F(0) = f(0)$ and $\tilde{\varphi}$ is a quadratic polynomial with $\tilde{\varphi}(0) = \varphi(0)$, $\tilde{\varphi}''(0) = \varphi''(0)$, case 2 follows from case 1. $\square$

Remark 5.7. Theorem 5.6 holds for integrals extended over the whole real axis as well (actually this is the most interesting case) as long as, in addition, f is bounded, $|\varphi'(\pm\infty)| \geq C > 0$, and $\int_{\mathbb{R}} |f'\varphi' - f\varphi''|\,(\varphi')^{-2}\,dk < \infty$. Indeed, it is easy to check that Lemma 5.3 is true under these hypotheses and then the proof of Theorem 5.6 is exactly the same.

Problems

5.1. The chord of a guitar of length L is plucked at its middle point and then released. Write the mathematical model which governs the vibrations and solve it. Compute the energy $E(t)$.

5.2. Solve the problem

$$\begin{cases} u_{tt} - u_{xx} = 0 & 0 < x < 1,\, t > 0 \\ u\,(x,0) = u_t\,(x,0) = 0 & 0 \le x \le 1 \\ u_x\,(0,t) = 1,\, u\,(1,t) = 0 & t \ge 0. \end{cases}$$

5.3. *Forced vibrations.* Solve the problem.

$$\begin{cases} u_{tt} - u_{xx} = g\,(t)\sin x & 0 < x < \pi,\, t > 0 \\ u\,(x,0) = u_t\,(x,0) = 0 & 0 \le x \le \pi \\ u\,(0,t) = u\,(\pi,t) = 0 & t \ge 0 \end{cases}$$

[*Answer.* $u\,(x,t) = \sin x \int_0^t g\,(t-\tau)\sin\tau\;d\tau$].

5.4. *Equipartition of energy.* Let $u = u\,(x,t)$ be the solution of the global Cauchy problem for the equation $u_{tt} - cu_{xx} = 0$, with initial data $u\,(x,0) = g\,(x)$, $u_t\,(x,0) = h\,(x)$. Assume that g and h are smooth functions with compact support contained in the interval (a,b). Show that there exists T such that, for $t \ge T$,

$$E_{cin}\,(t) = E_{pot}\,(t)\,.$$

5.5. Solve the global Cauchy problem for the equation $u_{tt} - cu_{xx} = 0$, with the following initial data:

a) $u\,(x,0) = 1$ if $|x| < a$, $u\,(x,0) = 0$ if $|x| > a$; $u_t\,(x,0) = 0$

b) $u\,(x,0) = 0$; $u_t\,(x,0) = 1$ if $|x| < a$, $u_t\,(x,0) = 0$ if $|x| > a$.

5.6. Check that formula (5.42) may be written in the following form:

$$u\,(x + c\xi - c\eta,\, t + \xi + \eta) - u\,(x + c\xi,\, t + \xi) - u\,(x - c\eta,\, t + \eta) + u\,(x,t) = 0. \tag{5.154}$$

Show that if u is a C^2 function and satisfies (5.154), then

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

Thus, (5.154) can be considered as a weak formulation of the wave equation.

5.7. The small longitudinal free vibrations of an elastic bar are governed by the following equation

$$\rho\,(x)\,\sigma\,(x)\,\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left[E\,(x)\,\sigma\,(x)\,\frac{\partial u}{\partial x}\right] \tag{5.155}$$

where u is the longitudinal displacement, ρ is the linear density of the material, σ is the cross section of the bar and E is its *Young's modulus*[39].

---

[39] E is the proportionality factor in the *strain-stress* relation given by Hooke's law: T (strain) $= E\,\varepsilon$ (stress). Here $\varepsilon \simeq u_x$. For steel, $E = 2 \times 10^{11}$ dine/cm^2, for alluminium, $E = 7 \times 10^{12}$ dine/cm^2.

Assume the bar has constant cross section but it is constructed by welding together two bars, of different (constant) Young's modulus E_1, E_2 and density ρ_1, ρ_2, respectively.

Since the two bars are welded together, the displacement u is continuous across the junction, which we locate at $x = 0$. In this case:

(a) Give a weak formulation of the global initial value problem for equation (5.155).

(b) Deduce that the following jump condition must hold at $x = 0$:

$$E_1 u(0-, t) = E_2 u(0+, t) \qquad\qquad t > 0. \qquad\qquad (5.156)$$

(c) Let $c_j = E_j/\rho_j$, $j = 1, 2$. A left incoming wave $u_{inc}(x, t) = \exp\left[i\left(x - c_1 t\right)\right]$ produces at the junction a reflected wave $u_{ref}(x, t) = a \exp\left[i\left(x + c_1 t\right)\right]$ and a transmitted wave $u_{tr}(x, t) = b \exp\left[i\left(x - c_2 t\right)\right]$. Determine a, b and interpret the result.

[*Hint.* (c) Look for a solution of the form

$$u = u_{inc} + u_{ref}$$

for $x < 0$ and $u = u_{tr}$ for $x > 0$. Use the continuity of u and the jump condition (5.156)].

5.8. Determine the characteristics of Tricomi equation

$$u_{tt} - t u_{xx} = 0.$$

[*Answer:* $3x \pm 2t^{3/2} = k$, for $t > 0$].

5.9. Classify the equation

$$t^2 u_{tt} + 2t u_{xt} + u_{xx} - u_x = 0$$

and find the characteristics. After a reduction to canonical form, find the general solution.

[*Answer:*

$$u(x, t) = F\left(te^{-x}\right) + G\left(te^{-x}\right)e^x,$$

with F, G arbitrary].

5.10. Consider the following *characteristic Cauchy problem*[40] for the wave equation in the half-plane $x > t$:

$$\begin{cases} u_{tt} - u_{xx} = 0 & x > t \\ u(x, x) = f(x) & x \in \mathbb{R} \\ u_\nu(x, x) = g(x) & x \in \mathbb{R} \end{cases}$$

where $\nu = (1, -1)/\sqrt{2}$. Establish whether or not this problem is well posed.

---

[40] Note that the data are the values of u and of the normal derivative on the characteristic $y = x$.

5.11. Consider the following so called *Goursat problem*[41] for the wave equation in the sector $-t < x < t$:

$$\begin{cases} u_{tt} - u_{xx} = 0 & -t < x < t \\ u(x, x) = f(x), \ u(x, -x) = g(x) & x > 0 \\ f(0) = g(0). \end{cases}$$

Establish whether or not this problem is well posed.

5.12. *Ill posed non-characteristic Cauchy problem for the heat equation.* Check that for every integer k, the function

$$u_k(x, t) = \frac{1}{k}[\cosh kx \cos kx \cos 2k^2 t - \sinh kx \sin kx \sin 2k^2 t]$$

solves $u_t = u_{xx}$ and the (non characteristic) initial conditions:

$$u(0, t) = \frac{1}{k}\cos 2k^2 t, \ u_x(0, t) = 0.$$

Deduce that the corresponding Cauchy problem in the half-plane $x > 0$ is **ill posed**.

5.13. Consider the telegrapher's system (5.83), (5.84).

(a) By elementary manipulations derive the following second order equation for the inner current I:

$$I_{tt} - \frac{1}{LC}I_{xx} + \frac{RC + GL}{LC}I_t + \frac{RG}{LC}I = 0.$$

(b) Let

$$I = e^{-kt}v$$

and choose k in order for v to satisfy an equation of the form

$$v_{tt} - \frac{1}{LC}v_{xx} + hv = 0.$$

Check that the condition $RC = GL$ is necessary to have non dispersive waves (*distorsionless transmission* line).

5.14. *Circular membrane.* A perfectly flexible and elastic membrane at rest has the shape of the circle $B_1 = \{(x, y) : x^2 + y^2 \le 1\}$. If the boundary is fixed and there are no external loads, the vibrations of the membrane are governed by the following system:

$$\begin{cases} u_{tt} - c^2\left(u_{rr} + \frac{1}{r}u_r + \frac{1}{r^2}u_{\theta\theta}\right) = 0 & 0 < r < 1, \ 0 \le \theta \le 2\pi, \ t > 0 \\ u(r, \theta, 0) = g(r, \theta), \ u_t(r, \theta, 0) = h(r, \theta) & 0 < r < 1, \ 0 \le \theta \le 2\pi \\ u(1, \theta, t) = 0 & 0 \le \theta \le 2\pi, \ t \ge 0. \end{cases}$$

---

[41] Note that the data are the values of u on the characteristics $y = x$ and $y = -x$, for $x > 0$.

In the case $h = 0$ e $g = g(r)$, use the method of separation of variables to find the solution

$$u(r, t) = \sum_{n=1}^{\infty} a_n J_0(\lambda_n r) \cos \lambda_n t$$

where J_0 is the Bessel function of order zero, $\lambda_1, \lambda_2, ...$ are the zeros of J_0 and the coefficients a_n are given by

$$a_n = \frac{2}{c_n^2} \int_0^1 sg(s) J_0(\lambda_n s)\, ds$$

where

$$c_n = \sum_{n=1}^{\infty} \frac{(-1)^k}{k!(k+1)!} \left(\frac{\lambda_n}{2}\right)^{2k+1}$$

(see Remark 2.2.5).

5.15. *Circular waveguide.* Consider the equation $u_{tt} - c^2 \Delta u = 0$ in the cylinder

$$C_R = \{(r, \theta, z) : 0 \le r \le R, \, 0 \le \theta \le 2\pi, \, -\infty < z < +\infty\}.$$

Determine the axially symmetric solutions of the form

$$u(r, z, t) = v(r) w(z) h(t)$$

satisfying the Neumann condition $u_r = 0$ on $r = R$.

[*Answer.*

$$u_n(r, z, t) = \exp\{-i(\omega t - kz)\} J_0(\mu_n r/R), \qquad n \in \mathbb{N}.$$

where J_0 is the Bessel function, μ_n are its stationary points $(J_0'(\mu_n) = 0)$ and

$$\frac{\omega^2}{c^2} = k^2 + \frac{\mu_n^2}{R^2}\,].$$

5.16. Let u be the solution of $u_{tt} - c^2 \Delta u = 0$ in $\mathbb{R}^3 \times (0, +\infty)$ with data

$$u(\mathbf{x}, 0) = g(\mathbf{x}) \quad \text{and} \quad u_t(\mathbf{x}, \mathbf{0}) = h(\mathbf{x}),$$

both supported in the sphere $\bar{B}_\rho(\mathbf{0})$. Describe the support of u for $t > 0$.
[*Answer*: The spherical shell $\bar{B}_{\varrho+ct}(\mathbf{0}) \setminus B_{\varrho-ct}(\mathbf{0})$, of width 2ρ, which expands at speed c].

5.17. *Focussing effect.* Solve the problem

$$\begin{cases} w_{tt} - c^2 \Delta w = 0 & \mathbf{x} \in \mathbb{R}^3, t > 0 \\ w(\mathbf{x}, 0) = 0, \quad w_t(\mathbf{x}, 0) = h(|\mathbf{x}|) & \mathbf{x} \in \mathbb{R}^3 \end{cases}$$

where $(r = |\mathbf{x}|)$

$$h(r) = \begin{cases} 1 & 0 \le r \le 1 \\ 0 & r > 1. \end{cases}$$

Check that $w(r, t)$ displays a discontinuity at the origin at time $t = 1/c$.

5.18. Show that the solution of the two-dimensional non-homogeneous Cauchy problem with zero initial data is given by

$$u(\mathbf{x}, \mathbf{t}) = \frac{1}{2\pi c} \int_0^t \int_{B_{c(t-s)}(\mathbf{x})} \frac{1}{\sqrt{c^2 (t-s)^2 + |\mathbf{x} - \mathbf{y}|^2}} f(\mathbf{y}, s) \, d\mathbf{y} ds.$$

5.19. For *linear gravity waves* ($\sigma = 0$), examine the case of uniform finite depth, replacing condition (5.143) by

$$\phi_z(x, -H, t) = 0$$

under the initial conditions (5.142).

(*a*) Write the dispersion relation.

Deduce that:

(*b*) The phase and group velocity have a finite upper bound.

(*c*) The square of the phase velocity in deep water $(H \gg \lambda)$ is proportional to the wavelength.

(*d*) Linear shallow water waves $(H \ll \lambda)$ are not dispersive.

[*Answer:* (*a*) $\omega^2 = gk \tanh(kH)$,

(*b*) $c_{p\,\max} = \sqrt{kH}$,

(*c*) $c_p^2 \sim g\lambda/2\pi$,

(*d*) $c_p^2 \sim gH$].

5.20. Determine the travelling wave solutions of the linearized system (5.139) of the form

$$\phi(x, z, t) = F(x - ct) G(z).$$

Rediscover the dispersion relation found in Problem 5.19 (*a*).
[*Answer:*

$$\phi(x, z, t) = \cosh k (z + H) \{A \cos k (x - ct) + B \sin k (x - ct)\},$$

A, B arbitrary constants and $c^2 = g \tanh(kH)/k$].

6

Elements of Functional Analysis

Motivations – Norms and Banach Spaces – Hilbert Spaces – Projections and Bases – Linear Operators and Duality – Abstract Variational Problems – Compactness and Weak Convergence – The Fredholm Alternative – Spectral Theory for Symmetric Bilinear Forms

6.1 Motivations

The main purpose in the previous chapters has been to introduce part of the basic and classical theory of some important equations of mathematical physics. The emphasis on phenomenological aspects and the connection with a probabilistic point of view should have conveyed to the reader some intuition and feeling about the interpretation and the limits of those models.

The few rigorous theorems and proofs we have presented had the role of bringing to light the main results on the qualitative properties of the solutions and justifying, partially at least, the well-posedness of the relevant boundary and initial/boundary value problems we have considered.

However, these purposes are somehow in competition with one of the most important role of modern mathematics, which is to reach a unifying vision of large classes of problems under a common structure, capable not only of increasing theoretical understanding, but also of providing the necessary flexibility to guide the numerical methods which will be used to compute approximate solutions.

This conceptual jump requires a change of perspective, based on the introduction of abstract methods, historically originating from the vain attempts to solve basic problems (e.g. in electrostatics) at the end of the 19th century. It turns out that the new level of knowledge opens the door to the solution of complex problems in modern technology.

These abstract methods, in which analytical and geometrical aspects fuse, are the core of the branch of Mathematics, called Functional Analysis.

It could be useful for understanding the subsequent development of the theory, to examine in an informal way how the main ideas come out, working on a couple of specific examples.

Let us go back to the derivation of the diffusion equation, in subsection 2.1.2. If the body is heterogeneous or anisotropic, may be with discontinuities in its thermal parameters (e.g. due to the mixture of two different materials), the Fourier law of heat conduction gives for the flux function $\mathbf{q}$ the form

$$\mathbf{q} = -\mathbf{A}(\mathbf{x})\nabla u,$$

where the matrix $\mathbf{A}$ satisfies the condition

$$\mathbf{q}\cdot\nabla u = -\mathbf{A}(\mathbf{x})\nabla u \cdot \nabla u \leq 0 \qquad (ellipticity\ condition),$$

reflecting the tendency of heat to flow from hotter to cooler regions. If $\rho = \rho(\mathbf{x})$ and $c_v = c_v(\mathbf{x})$ are the density and the specific heat of the material, and $f = f(\mathbf{x})$ is the rate of external heat supply per unit volume, we are led to the diffusion equation

$$\rho c_v u_t - \operatorname{div}(\mathbf{A}(\mathbf{x})\nabla u) = f.$$

In stationary conditions, $u(\mathbf{x},t) = u(\mathbf{x})$, and we are reduced to

$$-\operatorname{div}(\mathbf{A}(\mathbf{x})\nabla u) = f. \tag{6.1}$$

Since the matrix $\mathbf{A}$ encodes the conductivity properties of the medium, we expect a low degree of regularity of $\mathbf{A}$, but then a natural question arises: what is the meaning of equation (6.1) if we cannot compute the divergence of $\mathbf{A}$?

We have already faced similar situations in subsections 4.4.2, where we have introduced discontinuous solutions of a conservation law, and in subsection 5.4.2, where we have considered solutions of the wave equation with irregular initial data. Let us follow the same ideas.

Suppose we want to solve equation (6.1) in a bounded domain Ω, with zero boundary data (Dirichlet problem). Formally, we multiply the differential equation by a smooth test function *vanishing on* $\partial\Omega$, and we integrate over Ω:

$$\int_\Omega -\operatorname{div}(\mathbf{A}(\mathbf{x})\nabla u)\, v\ d\mathbf{x} = \int_\Omega fv\ d\mathbf{x}.$$

Since $v = 0$ on $\partial\Omega$, using Gauss' formula we obtain

$$\int_\Omega \mathbf{A}(\mathbf{x})\nabla u \cdot \nabla v\ d\mathbf{x} = \int_\Omega fv\ d\mathbf{x} \tag{6.2}$$

which is called *weak* or *variational* formulation of our Dirichlet problem.

Equation (6.2) makes perfect sense for $\mathbf{A}$ and f bounded (possibly discontinuous) and $u, v \in \mathring{C}^1(\overline{\Omega})$, the set of of functions in $C^1(\overline{\Omega})$, vanishing on $\partial\Omega$. Then, we may say that $u \in \mathring{C}^1(\overline{\Omega})$ is a *weak* solution of our Dirichlet problem if (6.2)

holds *for every* $v \in \overset{\circ}{C}^1\left(\overline{\Omega}\right)$. Fine, but now we have to prove the well-posedness of the problem so formulated!

Things are not so straightforward, as we have experienced in section 4.4.3 and, actually, it turns out that $\overset{\circ}{C}^1\left(\overline{\Omega}\right)$ is not the proper choice, although it seems to be the natural one. To see why, let us consider another example, somewhat more revealing.

Consider the equilibrium position of a stretched membrane having the shape of a square Ω, subject to an external load f (force per unit mass) and kept at level zero on $\partial\Omega$.

Since there is no time evolution, the position of the membrane may be described by a function $u = u\left(\mathbf{x}\right)$, solution of the Dirichlet problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ \quad\; u = 0 & \text{on } \partial\Omega. \end{cases} \tag{6.3}$$

For problem (6.3), equation (6.2) becomes

$$\int_\Omega \nabla u \cdot \nabla v \, d\mathbf{x} = \int_\Omega fv \, d\mathbf{x} \qquad \forall v \in \overset{\circ}{C}^1\left(\overline{\Omega}\right). \tag{6.4}$$

Now, this equation has an interesting physical interpretation. The integral in the left hand side represents the work done by the internal elastic forces, due to a *virtual displacement* v. On the other hand $\int_\Omega fv$ expresses the work done by the external forces.

Thus, the weak formulation (6.4) states that these two works balance, which constitutes a version of the *principle of virtual work*.

There is more, if we bring into play the energy. In fact, the *total potential energy* is proportional to

$$E\left(v\right) = \underbrace{\int_\Omega \left|\nabla v\right|^2 dx}_{\text{internal elastic energy}} - \underbrace{\int_\Omega fv \, d\mathbf{x}}_{\text{external potential energy}}. \tag{6.5}$$

Since nature likes to save energy, the equilibrium position u corresponds to the minimizer of (6.5) among all the *admissible* configurations v. This fact is closely connected with the principle of virtual work and, actually, it is equivalent to it (see subsection 8.4.1).

Thus, changing point of view, instead of looking for a weak solution of (6.4) we may, equivalently, look for a minimizer of (6.5).

However there is a drawback. It turns out that the minimum problem *does not have a solution*, except for some trivial cases. The reason is that we are looking in the wrong set of admissible functions.

Why $\overset{\circ}{C}^1\left(\overline{\Omega}\right)$ is a wrong choice? To be minimalist, it is like looking for the minimizer of the function

$$f\left(x\right) = \left(x - \pi\right)^2$$

among the rational numbers!

Anyway, the answer is simple: $\mathring{C}^1\left(\overline{\Omega}\right)$ is not naturally tied to the physical meaning of $E\left(v\right)$, which is an energy and only requires *the gradient of u to be square integrable,* that is $|\nabla u| \in L^2\left(\Omega\right)$. There is no need of *a priori* continuity of the derivatives, actually neither of u. The space $\mathring{C}^1\left(\overline{\Omega}\right)$ is too narrow to have any hope of finding the minimizer there. Thus, we are forced to enlarge the set of admissible functions and the correct one turns out to be the so called *Sobolev space* $H_0^1\left(\Omega\right)$, whose elements are exactly the functions belonging to $L^2\left(\Omega\right)$, together with their first derivatives, vanishing on $\partial\Omega$. We could call them functions of finite energy!

Although we feel we are on the right track, there is a price to pay, to put everything in a rigorous perspective and avoid risks of contradiction or non-senses. In fact many questions arise immediately.

For instance, what do we mean by the *gradient* of a function which is only in $L^2(\Omega)$, maybe with a lot of discontinuities? More: a function in $L^2(\Omega)$ is, in principle, well defined except on sets of measure zero. But, then, what does it mean "vanishing on $\partial\Omega$", which is precisely a set of measure zero?

We shall answer these questions in Chapter 7. We may anticipate that, for the first one, the idea is the same we used to define the *Dirac delta* as a derivative of the Heaviside function, resorting to a weaker notion of derivative (we shall say *in the sense of distributions*), based on the *miraculous* formula of Gauss and the introduction of a suitable set of test function.

For the second question, there is a way to introduce in a suitable coherent way a so called *trace operator* which associates to a function $u \in L^2\left(\Omega\right)$, with gradient in $L^2\left(\Omega\right)$, a function $u_{|\partial\Omega}$ representing its values on $\partial\Omega$ (see subsection 6.6.1). The elements of $H_0^1\left(\Omega\right)$ vanish on $\partial\Omega$ in the sense that they have zero trace.

Another question is what makes the space $H_0^1\left(\Omega\right)$ so special. Here the conjunction between geometrical and analytical aspects comes into play. First of all, although it is an infinite-dimensional vector space, we may endow $H_0^1\left(\Omega\right)$ with a structure which reflects as much as possible the structure of a *finite* dimensional vector space like $\mathbb{R}^n$, where life is obviously easier.

Indeed, in this vector space (thinking of $\mathbb{R}$ as the scalar field) we may introduce an *inner product* given by

$$(u, v)_1 = \int_{\Omega} \nabla u \cdot \nabla v$$

with the same properties of an inner product in $\mathbb{R}^n$. Then, it makes sense to talk about *orthogonality between two functions u and v in* $H_0^1\left(\Omega\right)$, expressed by the vanishing of their inner product:

$$(u, v)_1 = 0.$$

Having defined the inner product $(\cdot, \cdot)_1$, we may define the *size (norm)* of u by

$$\|u\|_1 = \sqrt{(u, u)_1}$$

and the distance between u and v by

$$\text{dist}\,(u, v) = \|u - v\|_1\,.$$

Thus, we may say that a sequence $\{u_n\} \subset H_0^1\,(\Omega)$ converges to u in $H_0^1\,(\Omega)$ if

$$\text{dist}\,(u_n, u) \to 0 \quad \text{as } n \to \infty.$$

It may be observed that all of this can be done, even more comfortably, in the space $\mathring{C}^1\,(\overline{\Omega})$. This is true, but with a key difference.

Let us use an analogy with an elementary fact. The minimizer of the function

$$f\,(x) = (x - \pi)^2$$

does not exist among the rational numbers $\mathbb{Q}$, although it can be approximated as much as one likes by these numbers. If from a very practical point of view, rational numbers could be considered satisfactory enough, certainly it is not so from the point of view of the development of science and technology, since, for instance, no one could even conceive the achievements of *Calculus* without the real number system.

As $\mathbb{R}$ is the *completion of* $\mathbb{Q}$, in the sense that $\mathbb{R}$ contains all the limits of sequences in $\mathbb{Q}$ that converge somewhere, the same is true for $H_0^1\,(\Omega)$ with respect to $\mathring{C}^1\,(\overline{\Omega})$. This makes $H_0^1\,(\Omega)$ a so called *Hilbert space* and gives it a big advantage with respect to $\mathring{C}^1\,(\overline{\Omega})$, which we illustrate going back to our membrane problem and precisely to equation (6.4). This time we use a geometrical interpretation.

In fact, (6.4) means that we are searching for an element u, whose inner product with any element v of $H_0^1\,(\Omega)$ reproduces "the action of f on v", given by the linear map

$$v \longmapsto \int_\Omega fv.$$

This is a familiar situation in Linear Algebra. Any function $F : \mathbb{R}^n \to \mathbb{R}$, which is *linear*, that is such that

$$F\,(a\mathbf{x} + b\mathbf{y}) = aF\,(\mathbf{x}) + bF\,(\mathbf{y}) \qquad \forall a, b \in \mathbb{R},\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

can be *expressed as the inner product with a unique representative vector* $\mathbf{z}_F \in \mathbb{R}^n$ (Representation Theorem). This amounts to saying that there is exactly one solution $\mathbf{z}_F$ of the equation

$$\mathbf{z} \cdot \mathbf{y} = \mathbf{F}\,(\mathbf{y}) \qquad \text{for every } \mathbf{y} \in \mathbb{R}^n. \tag{6.6}$$

The structure of the two equations (6.4), (6.6) is the same: on the left hand side there is an *inner product* and on the other one a *linear map*.

Another natural question arises: is there any analogue of the Representation Theorem in $H_0^1\,(\Omega)$?

The answer is yes (see Riesz's Theorem 6.3), with a little effort due to the infinite dimension of $H_0^1\,(\Omega)$. The Hilbert space structure of $H_0^1\,(\Omega)$ plays a key

role. This requires the study of *linear functionals* and the related concept of *dual space*. Then, an abstract result of geometric nature, implies the well-posedness of a concrete boundary value problem.

What about equation (6.2)? Well, if the matrix **A** is symmetric and strictly positive, the left hand side of (6.2) *still defines an inner product* in $H_0^1(\Omega)$ and again Riesz's Theorem yields the well-posedness of the Dirichlet problem.

If **A** is not symmetric, things change only a little. Various generalizations of Riesz's Theorem (e.g. the Lax-Milgram Theorem 6.4) allow the unified treatment of more general problems, through their *weak or variational formulation.* Actually, as we have experienced with equation (6.2), the variational formulation is often the only way of formulating and solving a problem, without losing its original features.

The above arguments should have convinced the reader of the existence of a general Hilbert space structure underlying a large class of problems, arising in the applications. In this chapter we develop the tools of Functional Analysis, essential for a correct variational formulation of a wide variety of boundary value problems. The results we present constitute the theoretical basis for numerical methods such as *finite elements* or more generally, *Galerkin's methods*, and this makes the theory even more attractive and important.

More advanced results, related to general solvability questions and the spectral properties of elliptic operators are included at the end of this chapter.

A final comment is in order. Look again at the minimization problem above. We have enlarged the class of admissible configurations from a class of quite smooth functions to a rather wide class of functions. What kind of solutions are we finding with these abstract methods? If the data (e.g. Ω and f, for the membrane) are regular, could the corresponding solutions be irregular? If yes, this does not sound too good! In fact, although we are working in a setting of possibly irregular configurations, it turns out that the solution actually possesses its natural degree of regularity, once more confirming the intrinsic coherence of the method.

It also turns out that the knowledge of the optimal regularity of the solution plays an important role in the error control for numerical methods. However, this part of the theory is rather technical and we do not have much space to treat it in detail. We shall only state some of the most common results.

The power of abstract methods is not restricted to stationary problems. As we shall see, Sobolev spaces depending on time can be introduced for the treatment of evolution problems, both of diffusive or wave propagation type (see Chapter 7).

Also, in this introductory book, the emphasis is mainly to *linear* problems.

6.2 Norms and Banach Spaces

It may be useful for future developments, to introduce *norm* and *distance* independently of an *inner product*, to emphasize better their axiomatic properties.

Let X be a linear space over the scalar field $\mathbb{R}$ or $\mathbb{C}$. A *norm* in X, is a real function

$$\|\cdot\| : X \to \mathbb{R} \tag{6.7}$$

such that, for each scalar λ and every $x,y \in X$, the following properties hold:

1. $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$ (*positivity*)
2. $\|\lambda x\| = |\lambda| \, \|x\|$ (*homogeneity*)
3. $\|x + y\| \leq \|x\| + \|y\|$ (*triangular inequality*).

A norm is introduced to measure the size (or the "length") of each vector $x \in X$, so that properties 1, 2, 3 should appear as natural requirements.

A *normed space* is a linear space X endowed with a norm $\|\cdot\|$. With a norm is associated the *distance* between two vectors given by

$$d(x, y) = \|x - y\|$$

which makes X a *metric space* and allows to define a *topology in* X and a notion of convergence in a very simple way.

We say that a sequence $\{x_n\} \subset X$ *converges* to x in X, and we write $x_m \to x$ in X, if

$$d(x_m, x) = \|x_m - x\| \to 0 \qquad \text{as } m \to \infty.$$

An important distinction is between convergent and *Cauchy* sequences. A sequence $\{x_m\} \subset X$ is a *Cauchy* sequence if

$$d(x_m, x_k) = \|x_m - x_k\| \to 0 \qquad \text{as } m, k \to \infty.$$

If $x_m \to x$ in X, from the triangular inequality, we may write

$$\|x_m - x_m\| \leq \|x_m - x\| + \|x_k - x\| \to 0 \qquad \text{as } m, k \to \infty$$

and therefore

$$\{x_m\} \text{ convergent } \textbf{implies} \text{ that } \{x_m\} \text{ is a Cauchy sequence.} \qquad (6.8)$$

The converse in not true, in general. Take $X = \mathbb{Q}$, with the usual norm given by $|x|$. The sequence of rational numbers

$$x_m = \left(1 + \frac{1}{m}\right)^m$$

is a Cauchy sequence but it is *not* convergent in $\mathbb{Q}$, since its limit is the irrational number e.

A normed space in which every Cauchy sequence converges is called **complete** and deserves a special name.

Definition 6.1. *A complete, normed linear space is called* **Banach space**.

The notion of convergence (or of limit) can be extended to functions from a normed space into another, always reducing it to the convergence of distances, that are real functions.

Let X, Y linear spaces, endowed with the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, and let $F : X \to Y$. We say that F is continuous at $x \in X$ if

$$\|F(y) - F(x)\|_Y \to 0 \qquad \text{when } \|y - x\|_X \to 0$$

or, equivalently, if, for every sequence $\{x_m\} \subset X$,

$$\|x_m - x\|_X \to 0 \quad \text{implies} \quad \|F(x_m) - F(x)\|_Y \to 0.$$

F is *continuous in* X if it is *continuous at every* $x \in X$. In particular:

Proposition 6.1. *Every norm in a linear space X is continuous in X.*

Proof. Let $\|\cdot\|$ be a norm in X. From the triangular inequality, we may write

$$\|y\| \le \|y - x\| + \|x\| \quad \text{and} \quad \|x\| \le \|y - x\| + \|y\|$$

whence

$$|\|y\| - \|x\|| \le \|y - x\|.$$

Thus, if $\|y - x\| \to 0$ then $|\|y\| - \|x\|| \to 0$, which is the continuity of the norm. $\square$

Some examples are in order.

Spaces of continuous functions. Let $X = C(A)$ be the set of (real or complex) continuous functions on A, where A is a compact subset of $\mathbb{R}^n$, endowed with the norm (called *maximum norm*)

$$\|f\|_{C(A)} = \max_A |f|.$$

A sequence $\{f_m\}$ converges to f in $C(A)$ if

$$\max_A |f_m - f| \to 0,$$

that is, if f_m *converges uniformly to* f in A. Since a uniform limit of continuous functions is continuous, $C(A)$ is a Banach space.

Note that other norms may be introduced in $C(A)$, for instance the *least squares* or $L^2(A)$ *norm*

$$\|f\|_{L^2(A)} = \left(\int_A |f|^2 \right)^{1/2}.$$

Equipped with this norm $C(A)$ is *not complete*. Let, for example $A = [-1, 1] \subset \mathbb{R}$. The sequence

$$f_m(t) = \begin{cases} 0 & t \le 0 \\ mt & 0 < t \le \frac{1}{m} \qquad (m \ge 1), \\ 1 & t > \frac{1}{m} \end{cases}$$

contained in $C\left([-1,1]\right)$, is a Cauchy sequence with respect to the L^2 norm. In fact (letting $m > k$),

$$\|f_m - f_k\|_{L^2(A)}^2 = \int_{-1}^{1} |f_m(t) - f_k(t)|^2 \, dt = (m-k)^2 \int_{0}^{1/m} t^2 dt + \int_{0}^{1/k} (1-kt)^2 \, dt$$

$$= \frac{(m-k)^2}{3m^3} + \frac{1}{3k} < \frac{1}{3}\left(\frac{1}{m} + \frac{1}{k}\right) \to 0 \qquad \text{as } m, k \to \infty.$$

However, f_n converges in $L^2\left(-1,1\right)-$norm (and pointwise) to the Heaviside function

$$\mathcal{H}(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0, \end{cases}$$

which is discontinuous at $t = 0$ and therefore does not belong to $C\left([-1,1]\right)$.

More generally, let $X = C^k\left(A\right)$, $k \geq 0$ integer, the set of functions continuously differentiable in A up to order k, included.

To denote a derivative of order m, it is convenient to introduce an $n - uple$ of nonnegative integers, $\alpha = (\alpha_1, ..., \alpha_n)$, called *multi-index*, of *length*

$$|\alpha| = \alpha_1 + ... + \alpha_n = m,$$

and set

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}.$$

We endow $C^k\left(A\right)$ with the norm (*maximum norm of order k*)

$$\|f\|_{C^k(A)} = \|f\|_{C(A)} + \sum_{|\alpha|=1}^{k} \|D^\alpha f\|_{C(A)}.$$

If $\{f_n\}$ is a Cauchy sequence in $C^k\left(A\right)$, all the sequences $\{D^\alpha f_n\}$ with $0 \leq |\alpha| \leq k$ are Cauchy sequences in $C\left(A\right)$. From the theorems on term by term differentiation of sequences, it follows that the resulting space is a Banach space.

Remark 6.1. With the introduction of *function spaces* we are actually making a step towards abstraction, regarding a function from a different perspective. In calculus we see it as a point map while here we have to consider it as *a single element* (or a point or a vector) *of a vector space.*

Summable and bounded functions. Let Ω be an *open set* in $\mathbb{R}^n$ and $p \geq 1$ a real number. Let $X = L^p\left(\Omega\right)$ be the set of functions f such that $|f|^p$ is Lebesgue integrable in Ω. Identifying two functions f and g when they are *equal a.e.*[1] in Ω,

---

[1] A property is valid *almost everywhere* in a set Ω, *a.e.* in short, if it is true at all points in Ω, but for a subset of measure zero (Appendix B).

$L^p(\Omega)$ becomes a Banach space[2] when equipped with the norm (*integral norm of order p*)

$$\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f|^p\right)^{1/p}.$$

The identification of two functions equal a.e. amounts to saying that an element of $L^p(\Omega)$ is not a single function but, actually, an equivalence class of functions, different from one another only on subsets of measure zero. At first glance, this fact could be annoying, but after all, the situation is perfectly analogous to considering a *rational number* as an equivalent class of fractions (2/3, 4/6, 8/12 .... represent the *same* number). For practical purposes one may always refer to the more convenient representative of the class.

Let $X = L^\infty(\Omega)$ the set of *essentially bounded* functions in Ω. Recall[3] that $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) is *essentially bounded* if there exists M such that

$$|f(x)| \le M \qquad \text{a.e. in } \Omega. \tag{6.9}$$

The infimum of all numbers M with the property (6.9) is called *essential supremum of f*, and denoted by

$$\|f\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_\Omega |f|.$$

If we identify two functions when they are equal a.e., $\|f\|_{L^\infty(\Omega)}$ is a norm in $L^\infty(\Omega)$, and $L^\infty(\Omega)$ becomes a Banach space.

Hölder inequality (1.9) mentioned in chapter 1, may be now rewritten in terms of norms as follows:

$$\left|\int_\Omega fg\right| \le \|f\|_{L^p(\Omega)} \|g\|_{L^q(\Omega)}, \tag{6.10}$$

where $q = p/(p-1)$ is the *conjugate exponent* of p, allowing also the case $p = 1$, $q = \infty$.

Note that, if Ω has *finite measure* and $1 \le p_1 < p_2 \le \infty$, from (6.10) we have, choosing $g \equiv 1$, $p = p_2/p_1$ and $q = p_2/(p_2 - p_1)$:

$$\left|\int_\Omega |f|^{p_1}\right| \le |\Omega|^{1/q} \|f\|_{L^{p_2}(\Omega)}^{p_1}$$

and therefore $L^{p_2}(\Omega) \subset L^{p_1}(\Omega)$. If the measure of Ω is *infinite*, this inclusion is not true, in general; for instance, $f \equiv 1$ belongs to $L^\infty(\mathbb{R})$ but is not in $L^p(\mathbb{R})$ for $1 \le p < \infty$.

6.3 Hilbert Spaces

Let X be a linear space over $\mathbb{R}$. An *inner or scalar product* in X is a function

$$(\cdot, \cdot) : X \times X \to \mathbb{R}$$

---

[2] See e.g. *Yoshida*, 1965.
[3] Appendix B.

with the following three properties. For every x, y, $z \in X$ and scalars λ, $\mu \in \mathbb{R}$:

1. $(x, x) \geq 0$ and $(x, x) = 0$ if and only if $x = 0$ (*positivity*)
2. $(x, y) = (y, x)$ (*symmetry*)
3. $(\mu x + \lambda y, z) = \mu (x, z) + \lambda (y, z)$ (*bilinearity*).

A linear space endowed with an inner product is called an *inner product space.* Property 3 shows that the inner product is linear with respect to its first argument. From 2, the same is true for the second argument as well. Then, we say that $(\cdot, \cdot)$ constitutes a *symmetric bilinear* form in X. When different inner product spaces are involved it may be necessary the use of notations like $(\cdot, \cdot)_X$, to avoid confusion.

Remark 6.2. If the scalar field is $\mathbb{C}$, then

$$(\cdot, \cdot) : X \times X \to \mathbb{C}$$

and property 2 has to be replaced by

2_{bis}. $(x, y) = \overline{(y, x)}$ where the bar denotes complex conjugation. As a consequence, we have

$$(z, \mu x + \lambda y) = \overline{\mu} (z, x) + \overline{\lambda} (z, y)$$

and we say that $(\cdot, \cdot)$ is *antilinear* with respect to its second argument or that it is a *sesquilinear form* in X.

An inner product *induces* a norm, given by

$$\|x\| = \sqrt{(x, x)} \tag{6.11}$$

In fact, properties 1 and 2 in the definition of norm are immediate, while the triangular inequality is a consequence of the following quite important theorem.

Theorem 6.1. *Let $x, y \in X$. Then:*
   (1) **Schwarz's inequality:**

$$|(x, y)| \leq \|x\| \, \|y\| \,. \tag{6.12}$$

Moreover equality holds in (6.12) if and only if x and y are linearly dependent.
   (2) **Parallelogram law:**

$$\|x + y\|^2 + \|x - y\|^2 = 2 \, \|x\|^2 + 2 \, \|y\|^2 \,.$$

The parallelogram law generalizes an elementary result in euclidean plane geometry: *in a parallelogram, the sum of the squares of the sides length equals the sum of the squares of the diagonals length.* The Schwarz inequality implies that the inner product is continuous; in fact, writing

$$(w, z) - (x, y) = (w - x, z) + (x, z - y)$$

we have
$$|(w, z) - (x, y)| \leq \|w - x\| \|z\| + \|x\| \|z - y\|$$
so that, if $w \to x$ and $z \to y$, then $(w, z) \to (x, y)$.

Proof. (1) We mimic the finite dimensional proof. Let $t \in \mathbb{R}$ and x, $y \in X$. Using the properties of the inner product and (6.11), we may write:

$$0 \leq (tx + y, tx + y) = t^2 \|x\|^2 + 2t(x, y) + \|y\|^2 \equiv P(t).$$

Thus, the second degree polynomial $P(t)$ is always nonnegative, whence

$$(x, y)^2 - \|x\|^2 \|y\|^2 \leq 0$$

which is the Schwarz inequality. Equality is possible only if $tx + y = 0$, i.e. if x and y are linearly dependent.

(2) Just observe that

$$\|x \pm y\|^2 = (x \pm y, y \pm y) = \|x\|^2 \pm 2(x, y) + \|y\|^2. \tag{6.13}$$

$\square$

Definition 6.2. *Let H be an inner product space. We say that H is a **Hilbert space** if it is complete with respect to the norm (6.11), induced by the inner product.*

Two Hilbert spaces H_1 and H_2 are *isomorphic* if *there exists a linear map* $L : H_1 \to H_2$ *which preserves the inner product,* i.e.:

$$(x, y)_{H_1} = (Lx, Ly)_{H_2} \qquad \forall x, y \in H_1.$$

In particular
$$\|x\|_{H_1} = \|Lx\|_{H_2}$$

Example 6.1. $\mathbb{R}^n$ is a Hilbert space with respect to the usual inner product

$$(\mathbf{x}, \mathbf{y})_{\mathbb{R}^n} = \mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{n} x_j y_j, \qquad \mathbf{x} = (x_1, ..., x_n), \mathbf{y} = (y_1, ..., y_n).$$

The induced norm is

$$|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sum_{j=1}^{n} x_j^2.$$

More generally, if $\mathbf{A} = (a_{ij})_{i,j=1,...,n}$ is a square matrix of order n, *symmetric* and *positive*,

$$(\mathbf{x}, \mathbf{y})_{\mathbf{A}} = \mathbf{x} \cdot \mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} a_{ij} x_i y_j \tag{6.14}$$

defines another scalar product in $\mathbb{R}^n$. Actually, *every* inner product in $\mathbb{R}^n$ may be written in the form (6.14), with a suitable matrix $\mathbf{A}$.

$\mathbb{C}^n$ is a Hilbert space with respect to the inner product

$$\mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{n} x_j \overline{y}_j \qquad \mathbf{x} = (x_1, ..., x_n), \mathbf{y} = (y_1, ..., y_n).$$

It is easy to show that every real (resp. complex) linear space of dimension n is isomorphic to $\mathbb{R}^n$ (resp. $\mathbb{C}^n$).

Example 6.2. $L^2(\Omega)$ is a Hilbert space (perhaps the most important one) with respect to the inner product

$$(u, v)_{L^2(\Omega)} = \int_\Omega uv.$$

If Ω is fixed, **we will simply use the notations** $(u, v)_0$ instead of $(u, v)_{L^2(\Omega)}$ and $\|u\|_0$ instead of $\|u\|_{L^2(\Omega)}$.

Example 6.3. Let $l_{\mathbb{C}}^2$ be the set of complex sequences $\mathbf{x} = \{x_m\}$ such that

$$\sum_{i=1}^{\infty} |x_m|^2 < \infty.$$

For $\mathbf{x} = \{x_m\}$ and $\mathbf{y} = \{y_m\}$, define

$$(\mathbf{x}, \mathbf{y})_{l_{\mathbb{C}}^2} = \sum_{i=1}^{\infty} x_i \overline{y}_j, \qquad \mathbf{x} = \{x_n\}, \mathbf{y} = \{y_n\}.$$

Then $(\mathbf{x}, \mathbf{y})_{l_{\mathbb{C}}^2}$ is an inner product which makes $l_{\mathbb{C}}^2$ a Hilbert space over $\mathbb{C}$ (see Problem 6.3). This space constitutes the discrete analogue of $L^2(0, 2\pi)$. Indeed, each $u \in L^2(0, 2\pi)$ has an expansion in Fourier series (Appendix A)

$$u(x) = \sum_{m \in \mathbb{Z}} \widehat{u}_m e^{imx},$$

where

$$\widehat{u}_m = \frac{1}{2\pi} \int_0^{2\pi} u(x) e^{-imx} dx.$$

Note that $\overline{\widehat{u}_m} = \widehat{u}_{-m}$, since u is a real function. From Parseval's identity, we have

$$(u, v)_0 = \int_0^{2\pi} uv = 2\pi \sum_{m \in \mathbb{Z}} \widehat{u}_m \widehat{v}_{-m}$$

and (Bessel's equation)

$$\|u\|_0^2 = \int_0^{2\pi} u^2 = 2\pi \sum_{m \in \mathbb{Z}} |\widehat{u}_m|^2.$$

Example 6.4. A *Sobolev space.* It is possible to use the frequency space introduced in the previous example to define the derivatives of a function in $L^2(0, 2\pi)$ in a weak or generalized sense. Let $u \in C^1(\mathbb{R})$, $2\pi-$periodic. The Fourier coefficients of u' are given by

$$\widehat{u'}_m = im\widehat{u}_m$$

and we may write

$$\|u'\|_0^2 = \int_0^{2\pi} (u')^2 = 2\pi \sum_{m \in \mathbb{Z}} m^2 |\widehat{u}_m|^2. \qquad (6.15)$$

Thus, both sequences $\{\widehat{u}_m\}$ and $\{m\widehat{u}_m\}$ belong to $l_{\mathbb{C}}^2$. But the right hand side in (6.15) does not involve u' directly, so that it makes perfect sense to define

$$H_{per}^1(0, 2\pi) = \left\{u \in L^2(0, 2\pi) : \{\widehat{u}_m\}, \{m\widehat{u}_m\} \in l^2\right\}$$

and introduce the inner product

$$(u, v)_{1,2} = (2\pi) \sum_{m \in \mathbb{Z}} \left[1 + m^2\right] \widehat{u}_m \widehat{v}_{-m}$$

which makes $H_{per}^1(0, 2\pi)$ into a Hilbert space. Since

$$\{m\widehat{u}_m\} \in l_{\mathbb{C}}^2,$$

with each $u \in H_{per}^1(0, 2\pi)$ is associated the function $v \in L^2(0, 2\pi)$ given by

$$v(x) = \sum_{m \in \mathbb{Z}} im\widehat{u}_m e^{imx}.$$

We see that v may be considered as *a generalized derivative of* u and $H_{per}^1(0, 2\pi)$ as the space of functions in $L^2(0, 2\pi)$, together with their first derivatives. Let $u \in H_{per}^1(0, 2\pi)$ and

$$u(x) = \sum_{m \in \mathbb{Z}} \widehat{u}_m e^{imx}.$$

Since

$$\left|\widehat{u}_m e^{imx}\right| = \frac{1}{m} m |\widehat{u}_m| \leq \frac{1}{2}\left(\frac{1}{m^2} + m^2 |\widehat{u}_m|^2\right)$$

the Weierstrass test entails that the Fourier series of u converges uniformly in $\mathbb{R}$. Thus u has a continuous, $2\pi-$periodic extension to all $\mathbb{R}$. Finally observe that, if we use the symbol u' also for the generalized derivative of u, the inner product in $H_{per}^1(0, 1)$ can be written in the form

$$(u, v)_{1,2} = \int_0^1 (u'v' + uv).$$

6.4 Projections and Bases

6.4.1 Projections

Hilbert spaces are the ideal setting to solve problems in infinitely many dimensions. They unify through the inner product and the induced norm, both an analytical and a geometric structure. As we shall shortly see, we may coherently introduce the concepts of orthogonality, projection and basis, prove a infinite-dimensional Pythagoras' Theorem (an example is just Bessel's equation) and introduce other operations, extremely useful from both a theoretical and practical point of view.

As in finite-dimensional linear spaces, two elements x, y belonging to an inner product space are called **orthogonal or normal** if $(x, y) = 0$, and we write $x \perp y$.

Now, if we consider a subspace V of $\mathbb{R}^n$, e.g. a hyperplane through the origin, every $\mathbf{x} \in \mathbb{R}^n$ has a unique orthogonal projection on V. In fact, if $\dim V = k$ and the unit vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$ constitute an *orthonormal basis* in V, we may always find an orthonormal basis in $\mathbb{R}^n$, given by

$$\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k, \mathbf{w}_{k+1}, ..., \mathbf{w}_n,$$

where $\mathbf{w}_{k+1}, ..., \mathbf{w}_n$ are suitable unit vectors. Thus, if

$$\mathbf{x} = \sum_{j=1}^{k} x_j \mathbf{v}_j + \sum_{j=k+1}^{n} x_j \mathbf{w}_j,$$

the projection of $\mathbf{x}$ on V is given by

$$P_V \mathbf{x} = \sum_{j=1}^{k} x_j \mathbf{v}_j.$$

On the other hand, the projection $P_V \mathbf{x}$ can be characterized through the following property, which does not involve a basis in $\mathbb{R}^n$: $P_V \mathbf{x}$ is *the point in V that minimizes the distance from* $\mathbf{x}$, that is

$$|P_V \mathbf{x} - \mathbf{x}| = \inf_{\mathbf{y} \in V} |\mathbf{y} - \mathbf{x}| . \tag{6.16}$$

In fact, if $\mathbf{y} = \sum_{j=1}^{k} y_j \mathbf{v}_j$, we have

$$|\mathbf{y} - \mathbf{x}|^2 = \sum_{j=1}^{k} (y_j - x_j)^2 + \sum_{j=k+1}^{n} x_j^2 \geq \sum_{j=k+1}^{N} x_j^2 = |P_V \mathbf{x} - \mathbf{x}|^2 .$$

In this case, the "infimum" in (6.16) is actually a "minimum".

The uniqueness of $P_V \mathbf{x}$ follows from the fact that, if $\mathbf{y}^* \in V$ and

$$|\mathbf{y}^* - \mathbf{x}| = |P_V \mathbf{x} - \mathbf{x}| ,$$

then we must have

$$\sum_{j=1}^{k}(y_j^* - x_j)^2 = 0,$$

whence $y_j^* = x_j$ for $j = 1, ..., k$, and therefore $\mathbf{y}^* = P_V\mathbf{x}$. Since

$$(\mathbf{x} - P_V\mathbf{x})\perp\mathbf{v}, \qquad \forall\mathbf{v} \in V$$

every $\mathbf{x} \in \mathbb{R}^n$ may be written in a unique way in the form

$$\mathbf{x} = \mathbf{y} + \mathbf{z}$$

with $\mathbf{y} \in V$ and $\mathbf{z} \in V^{\perp}$, where $V^{\perp}$ denotes the subspace of the vectors orthogonal to V.

Then, we say that $\mathbb{R}^n$ is *direct sum* of the subspaces V and $V^{\perp}$ and we write

$$\mathbb{R}^n = V \oplus V^{\perp}.$$

Finally,

$$|\mathbf{x}|^2 = |\mathbf{y}|^2 + |\mathbf{z}|^2$$

which is the Pythagoras' Theorem in $\mathbb{R}^n$.
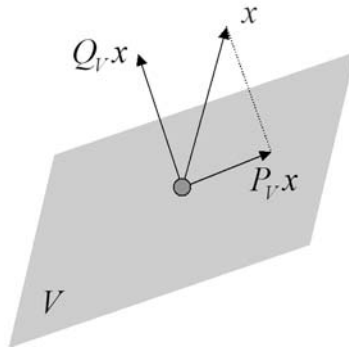
Fig. 6.1. Projection Theorem

We may extend all the above consideration to infinite-dimensional Hilbert spaces H, if we consider **closed subspaces** V of H. Here *closed* means with respect to the convergence induced by the norm. More precisely, a subset $U \subset H$ is closed in H if it contains all the limit points of sequences in U. Observe that if V has *finite dimension* k, it is automatically closed, since it is isomorphic to $\mathbb{R}^k$ (or $\mathbb{C}^k$). Also, a closed subspace of a Hilbert space is a Hilbert space as well, with respect to the inner product in H.

Unless stated explicitly, **from now on we consider Hilbert spaces over $\mathbb{R}$** (real Hilbert spaces), endowed with inner product $(\cdot, \cdot)$ and induced norm $\|\cdot\|$.

Theorem 6.2. (Projection Theorem). *Let V be a closed subspace of a Hilbert space H. Then, for every $x \in H$, there exists a unique element $P_V x \in V$ such that*

$$\|P_V x - x\| = \inf_{v \in V} \|v - x\|. \tag{6.17}$$

Moreover, the following properties hold:

1. *$P_V x = x$ if and only if $x \in V$.*
2. *Let $Q_V x = x - P_V x$. Then $Q_V x \in V^\perp$ and*

$$\|x\|^2 = \|P_V x\|^2 + \|Q_V x\|^2.$$

Proof. Let

$$d = \inf_{v \in V} \|v - x\|.$$

By the definition of least upper bound, we may select a sequence $\{v_m\} \subset V$, such that $\|v_m - x\| \to d$ as $m \to \infty$. In fact, for every integer $m \geq 1$ there exists $v_m \in V$ such that

$$d \leq \|v_m - x\| < d + \frac{1}{m}. \tag{6.18}$$

Letting $m \to \infty$ in (6.18), we get $\|v_m - x\| \to d$.

We now show that $\{v_m\}$ is a Cauchy sequence. In fact, using the parallelogram law for the vectors $v_k - x$ and $v_m - x$, we obtain

$$\|v_k + v_m - 2x\|^2 + \|v_k - v_m\|^2 = 2\|v_k - x\|^2 + 2\|v_m - x\|^2. \tag{6.19}$$

Since $\frac{v_k + v_m}{2} \in V$, we may write

$$\|v_k + v_m - 2x\|^2 = 4\left\|\frac{v_k + v_m}{2} - x\right\|^2 \geq 4d^2$$

whence, from (6.19):

$$\|v_k - v_m\|^2 = 2\|v_k - x\|^2 + 2\|v_m - x\|^2 - \|v_k + v_m - 2x\|^2$$
$$\leq 2\|v_k - x\|^2 + 2\|v_m - x\|^2 - 4d^2.$$

Letting $k, m \to \infty$, the right hand side goes to zero and therefore

$$\|v_k - v_m\| \to 0$$

as well. This proves that $\{v_m\}$ is a Cauchy sequence.

Since H is complete, v_m converges to an element $w \in H$ which belongs to V, because V *is closed*. Using the norm continuity (Proposition 6.1) we deduce

$$\|v_m - x\| \to \|w - x\| = d$$

so that w realizes the minimum distance from x among the elements in V.

We have to prove the uniqueness of w. Suppose $\bar{w} \in V$ is another element such that $\|\bar{w} - x\| = d$. The parallelogram law, applied to the vectors $w - x$ and $\bar{w} - x$, yields

$$\|w - \bar{w}\|^2 = 2 \|w - x\|^2 + 2 \|\bar{w} - x\|^2 - 4 \left\| \frac{w + \bar{w}}{2} - x \right\|^2$$

$$\leq 2d^2 + 2d^2 - 4d^2 = 0$$

whence $w = \bar{w}$.

We have proved that there exists a unique element $w = P_V x \in V$ such that

$$\|x - P_V x\| = d.$$

To prove 1, observe that, since V is closed, $x \in V$ if and only if $d = 0$, which means $x = P_V x$.

To show 2, let $Q_V x = x - P_V x$, $v \in V$ e $t \in \mathbb{R}$. Since $P_V x + tv \in V$ for every t, we have:

$$d^2 \leq \|x - (P_V x + tv)\|^2 = \|Q_V x - tv\|^2$$
$$= \|Q_V x\|^2 - 2t \, (Q_V x, v) + t^2 \|v\|^2$$
$$= d^2 - 2t \, (Q_V x, v) + t^2 \|v\|^2 \,.$$

Erasing d^2 and dividing by $t > 0$, we get

$$(Q_V x, v) \leq \frac{t}{2} \|v\|^2$$

which forces $(Q_V x, v) \leq 0$; dividing by $t < 0$ we get

$$(Q_V x, v) \geq \frac{t}{2} \|v\|^2$$

which forces $(Q_V x, v) \geq 0$. Thus $(Q_V x, v) = 0$ which means $Q_V x \in V^\perp$ and implies that

$$\|x\|^2 = \|P_V x + Q_V x\|^2 = \|P_V x\|^2 + \|Q_V x\|^2 \,,$$

concluding the proof. $\square$

The elements $P_V x$, $Q_V x$ are called **orthogonal projections** of x **on** V and $V^\perp$, respectively. The least upper bound in (6.17) is actually a minimum. Moreover thanks to properties 1, 2, we say that H *is direct sum of* V *and* $V^\perp$:

$$H = V \oplus V^\perp.$$

Note that

$$V^\perp = \{0\} \qquad \text{if and only if} \qquad V = H.$$

Remark 6.3. Another characterization of $P_V x$ is the following (see Problem 6.4): $u = P_V x$ *if and only if*

$$\begin{cases} \textbf{1.}\ u \in V \\ \textbf{2.}\ (x - u, v) = 0,\ \forall v \in V. \end{cases}$$

Remark 6.4. It is useful to point out that, even if V is *not a closed* subspace of H, the subspace $V^\perp$ is *always closed*. In fact, if $y_n \to y$ and $\{y_n\} \subset V^\perp$, we have, for every $x \in V$,

$$(y, x) = \lim (y_n, x) = 0$$

whence $y \in V^\perp$.

Example 6.5. Let $\Omega \subset \mathbb{R}^n$ be a set of finite measure. Consider in $L^2(\Omega)$ the $1-$dimensional subspace V of constant functions (a basis is given by $f \equiv 1$, for instance). Since it is finite-dimensional, V is closed in $L^2(\Omega)$. Given $f \in L^2(\Omega)$, to find the projection $P_V f$, we solve the minimization problem

$$\min_{\lambda \in \mathbb{R}} \int_\Omega (f - \lambda)^2.$$

Since

$$\int_\Omega (f - \lambda)^2 = \int_\Omega f^2 - 2\lambda \int_\Omega f + \lambda^2 |\Omega|,$$

we see that the minimizer is

$$\lambda = \frac{1}{|\Omega|} \int_\Omega f.$$

Therefore

$$P_V f = \frac{1}{|\Omega|} \int_\Omega f \quad \text{and} \quad Q_V f = f - \frac{1}{|\Omega|} \int_\Omega f.$$

Thus, the subspace $V^\perp$ is given by the functions $g \in L^2(\Omega)$ with *zero mean value*. In fact these functions are orthogonal to $f \equiv 1$:

$$(g, 1)_0 = \int_\Omega g = 0.$$

6.4.2 Bases

A Hilbert space is said to be **separable** when there exists a *countable dense* subset of H. An *orthonormal basis* in a separable *Hilbert space H* is sequence $\{w_k\}_{k \geq 1} \subset H$ such that[4]

$$\begin{cases} (w_k, w_j) = \delta_{kj} & k, j \geq 1, ... \\ \|w_k\| = 1 & k \geq 1 \end{cases}$$

---

[4] δ_{jk} is the Kronecker symbol.

and *every $x \in H$ may be expanded in the form*

$$x = \sum_{k=1}^{\infty} (x, w_k) \, w_k. \tag{6.20}$$

The series (6.20) is called **generalized Fourier series** and the numbers $c_k = (x, w_k)$ are the *Fourier coefficients* of x with respect to the basis $\{w_k\}$. Moreover (Pythagoras again!):

$$\|x\|^2 = \sum_{k=1}^{\infty} (x, w_k)^2 .$$

Given an orthonormal basis $\{w_k\}_{k \geq 1}$, the projection of $x \in H$ on the subspace V spanned by, say, $w_1, ..., w_N$ is given by

$$P_V x = \sum_{k=1}^{N} (x, w_k) \, w_k.$$

An example of separable Hilbert space is $L^2 (\Omega)$, $\Omega \subseteq \mathbb{R}^n$. In particular, the set of functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, ..., \frac{\cos mx}{\sqrt{\pi}}, \frac{\sin mx}{\sqrt{\pi}}, ...$$

constitutes an orthonormal basis in $L^2 (0, 2\pi)$ (see Appendix A).

It turns out that:

Proposition 6.2. *Every separable Hilbert space H admits an orthonormal basis.*

Proof (sketch). Let $\{z_k\}_{k \geq 1}$ be dense in H. Disregarding, if necessary, those elements which are spanned by other elements in the sequence, we may assume that $\{z_k\}_{k \geq 1}$ constitutes an *independent set*, i.e. every finite subset of $\{z_k\}_{k \geq 1}$ is composed by independent elements.

Then, an orthonormal basis $\{w_k\}_{k \geq 1}$ is obtained by applying to $\{z_k\}_{k \geq 1}$ the following so called *Gram-Schmidt process*. First, construct by induction a sequence $\{\tilde{w}\}_{k \geq 1}$ as follows. Let $\tilde{w}_1 = z_1$. Once $\tilde{w}_{k-1}$ is known, we construct $\tilde{w}_k$ by subtracting from z_k its components with respect to $\tilde{w}_1, ..., \tilde{w}_{k-1}$:

$$\tilde{w}_k = z_k - \frac{(z_k, \tilde{w}_{k-1})}{\|\tilde{w}_{k-1}\|^2} \tilde{w}_{k-1} - \cdots - \frac{(z_k, \tilde{w}_1)}{\|\tilde{w}_1\|^2} \tilde{w}_1.$$

In this way, $\tilde{w}_k$ is orthogonal to $\tilde{w}_1, ..., \tilde{w}_{k-1}$. Finally, set $w_k = \tilde{w}_k / \|\tilde{w}_{k-1}\|$. Since $\{z_k\}_{k \geq 1}$ is dense in H, then $\{w_k\}_{k \geq 1}$ is dense in H as well. Thus $\{w_k\}_{k \geq 1}$ is an orthonormal basis. $\square$

In the applications, orthonormal bases arise from solving particular boundary value problems, often in relation to the separation of variables method. Typical examples come form the vibrations of a non homogeneous string or from diffusion

in a rod with non constant thermal properties c_v, ρ, κ. The first example leads to the wave equation

$$\rho(x)u_{tt} - \tau u_{xx} = 0.$$

Separating variables $(u(x,t) = v(x)z(t))$, we find for the spatial factor the equation

$$\tau v'' + \lambda \rho v = 0.$$

In the second example we are led to

$$(\kappa v')' + \lambda c_v \rho v = 0.$$

These equations are particular cases of a general class of ordinary differential equations of the form

$$(pu')' + qu + \lambda w u = 0 \tag{6.21}$$

called *Sturm-Liouville* equations. Usually one looks for solutions of (6.21) in an interval (a, b), $-\infty \le a < b \le +\infty$, satisfying suitable conditions at the end points. The natural assumptions on p and q are $p \ne 0$ in (a, b) and p, q, p^{-1} locally integrable in (a, b). The function w plays the role of a *weight function*, continuous in $[a, b]$ and positive in (a, b).

In general, the resulting boundary value problem has non trivial solutions only for particular values of λ, called *eigenvalues*. The corresponding solutions are called *eigenfunctions* and it turns out that, when suitably normalized, they constitute an orthonormal basis in the Hilbert space $L^2_w(a, b)$, the set of Lebesgue measurable functions in (a, b) such that

$$\|u\|^2_{L^2_w} = \int_a^b u^2(x)w(x)\,dx < \infty,$$

endowed with the inner product

$$(u, v)_{L^2_w} = \int_a^b u(x)v(x)w(x)\,dx.$$

We list below some examples[5].

- Consider the problem

$$\begin{cases} (1 - x^2)u'' - xu' + \lambda u = 0 & \text{in } (-1, 1) \\ u(-1) < \infty, \quad u(1) < \infty. \end{cases}$$

The differential equation is known as *Chebyshev's* equation and may be written in the form (6.21):

$$((1 - x^2)^{1/2}u')' + \lambda(1 - x^2)^{-1/2}u = 0$$

---

[5] For the proofs, see *Courant-Hilbert*, vol. I, 1953.

which shows the proper weight function $w(x) = \left(1 - x^2\right)^{-1/2}$. The eigenvalues are $\lambda_n = n^2$, $n = 0, 1, 2, ...$ The corresponding eigenfunctions are the *Chebyshev polynomials* T_n, recursively defined by $T_0(x) = 1$, $T_1(x) = x$ and

$$T_{n+1} = 2xT_n - T_{n-1} \quad (n > 1).$$

For instance:

$$T_2(x) = 2x^2 - 1, \ \ T_3(x) = 4x^3 - 3x, \ \ T_4(x) = 8x^4 - 8x^2 - 1.$$

The normalized polynomials $\sqrt{1/\pi}T_0$, $\sqrt{2/\pi}T_1$, ..., $\sqrt{2/\pi}T_n$, ... constitute an orthonormal basis in $L_w^2(-1, 1)$.

- Consider the problem[6]

$$\left(\left(1 - x^2\right)u'\right)' + \lambda u = 0 \qquad \text{in } (-1, 1)$$

with weighted Neumann conditions

$$\left(1 - x^2\right)u'(x) \to 0 \qquad \text{as } x \to \pm 1.$$

The differential equation is known as *Legendre's* equation. The eigenvalues are $\lambda_n = n(n+1)$, $n = 0, 1, 2, ...$ The corresponding eigenfunctions are the *Legendre polynomials*, defined by $L_0(x) = 1, L_1(x) = x$,

$$(n+1)L_{n+1} = (2n+1)xL_n - nL_{n-1} \quad (n > 1)$$

or by *Rodrigues' formula*

$$L_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left(x^2 - 1\right)^n \qquad (n \geq 0).$$

For instance, $L_2(x) = (3x^2 - 1)/2$, $L_3(x) = (5x^3 - 3x)/2$. The normalized polynomials

$$\sqrt{\frac{2n+1}{2}}L_n$$

constitute an orthonormal basis in $L^2(-1, 1)$ (here $w(x) \equiv 1$). Every function $f \in L^2(-1, 1)$ has an expansion

$$f(x) = \sum_{n=0}^{\infty} f_n L_n(x)$$

where $f_n = \frac{2n+1}{2}\int_{-1}^{1} f(x)L_n(x)\,dx$, with convergence in $L^2(-1, 1)$.

- Consider the problem

$$\begin{cases} u'' - 2xu' + 2\lambda u = 0 & \text{in } (-\infty, +\infty) \\ e^{-x^2/2}u(x) \to 0 & \text{as } x \to \pm\infty. \end{cases}$$

---

[6] See also Problem 8.5.

The differential equation is known as *Hermite's* equation (see Problem 6.6) and may be written in the form (6.21):

$$\left(e^{-x^2} u'\right)' + 2\lambda e^{-x^2} u = 0$$

which shows the proper weight function $w(x) = e^{-x^2}$. The eigenvalues are $\lambda_n = n$, $n = 0, 1, 2, \ldots$. The corresponding eigenfunctions are the *Hermite polynomials* defined by *Rodrigues' formula*

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} \qquad (n \geq 0).$$

For instance

$$H_0(x) = 1, \quad H_1(x) = 2x, \quad H_2(x) = 4x^2 - 2, \quad H_3(x) = 8x^3 - 12x.$$

The normalized polynomials $\pi^{-1/4} (2^n n!)^{-1/2} H_n$ constitute an orthonormal basis in $L_w^2(\mathbb{R})$, with $w(x) = e^{-x^2}$. Every $f \in L_w^2(\mathbb{R})$ has an expansion

$$f(x) = \sum_{n=0}^{\infty} f_n H_n(x)$$

where $f_n = [\pi^{1/2} 2^n n!]^{-1} \int_{\mathbb{R}} f(x) H_n(x) e^{-x^2} dx$, with convergence in $L_w^2(\mathbb{R})$.

• After separating variables in the model for the vibration of a circular membrane the following *parametric Bessel equation of order p* arises (see Problem 6.8):

$$x^2 u'' + x u' + \left(\lambda x^2 - p^2\right) u = 0 \qquad x \in (0, a) \tag{6.22}$$

where $p \geq 0$, $\lambda \geq 0$, with the boundary conditions

$$u(0) \text{ finite}, \quad u(a) = 0. \tag{6.23}$$

Equation (6.22) may be written in Sturm-Liouville form as

$$(x u')' + \left(\lambda x - \frac{p^2}{x}\right) u = 0$$

which shows the proper weight function $w(x) = x$. The simple rescaling $z = \sqrt{\lambda} x$ reduces (6.22) to the *Bessel equation of order p*

$$z^2 \frac{d^2 u}{dz^2} + z \frac{du}{dz} + \left(z^2 - p^2\right) u = 0 \tag{6.24}$$

where the dependence on the parameter λ is removed. The only bounded solutions of (6.24) are the *Bessel functions of first kind and order p*, given by

$$J_p(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(k+1) \Gamma(k+p+1)} \left(\frac{z}{2}\right)^{p+2k}$$
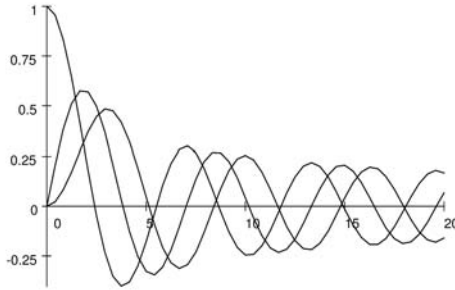
Fig. 6.2. Graphs of J_0, J_1 and J_2

where

$$\Gamma\left(s\right) = \int_0^\infty e^{-t} t^{s-1} dt \tag{6.25}$$

is the Euler $\Gamma-$ function. In particular, if $p = n \geq 0$, integer:

$$J_n\left(z\right) = \sum_{k=0}^\infty \frac{\left(-1\right)^k}{k!\left(k+n\right)!} \left(\frac{z}{2}\right)^{n+2k}.$$

For every p, there exists an infinite, increasing sequence $\{\alpha_{pj}\}_{j \geq 1}$ of positive zeroes of J_p:

$$J_p\left(\alpha_{pj}\right) = 0 \qquad \left(j = 1, 2, ...\right).$$

Then, the eigenvalues of problem (6.22), (6.23) are given by $\lambda_{pj} = \left(\frac{\alpha_{pj}}{a}\right)^2$, with corresponding eigenfunctions $u_{pj}\left(x\right) = J_p\left(\frac{\alpha_{pj}}{a}x\right)$. The normalized eigenfunctions

$$\frac{\sqrt{2}}{aJ_{p+1}\left(\alpha_{pj}\right)} J_p\left(\frac{\alpha_{pj}}{a}x\right)$$

constitute an orthonormal basis in $L_w^2\left(0, a\right)$, with $w\left(x\right) = x$. Every function $f \in L_w^2\left(0, a\right)$ has an expansion in *Fourier-Bessel series*

$$f\left(x\right) = \sum_{j=1}^\infty f_j J_p\left(\frac{\alpha_{pj}}{a}x\right),$$

where

$$f_j = \frac{2}{a^2 J_{p+1}^2\left(\alpha_{pj}\right)} \int_0^a x f\left(x\right) J_p\left(\frac{\alpha_{pj}}{a}x\right) dx,$$

convergent in $L_w^2\left(0, a\right)$.

6.5 Linear Operators and Duality

6.5.1 Linear operators

Let H_1 and H_2 be Hilbert spaces. A **linear operator from** H_1 into H_2 is a function

$$L : H_1 \to H_2$$

such that[7], $\forall \alpha, \beta \in \mathbb{R}$ and $\forall x, y \in H_1$

$$L(\alpha x + \beta y) = \alpha L x + \beta L y.$$

For every linear operator we define its *Kernel*, $\mathcal{N}(L)$ and *Range*, $\mathcal{R}(L)$, as follows:

Definition 6.3. *The **kernel** of L, is the pre-image of the null vector in H_2:*

$$\mathcal{N}(L) = \{x \in H_1 : Lx = 0\}.$$

*The **range** of L is the set of all outputs from points in H_1:*

$$\mathcal{R}(L) = \{y \in H_2 : \exists x \in H_1, Lx = y\}.$$

$\mathcal{N}(L)$ and $\mathcal{R}(L)$ are linear subspaces of H_1 and H_2, respectively.

Our main objects will be linear bounded operators.

Definition 6.4. *A linear operator $L : H_1 \to H_2$ is **bounded** if there exists a number C such that*

$$\|Lx\|_{H_2} \leq C \|x\|_{H_1}, \qquad \forall x \in H_1. \tag{6.26}$$

The number C controls the expansion rate operated by L on the elements of H_1. In particular, if $C < 1$, L contracts the sizes of the vectors in H_1.

If $x \neq 0$, using the linearity of L, we may write (6.26) in the form

$$\left\| L \left(\frac{x}{\|x\|_{H_1}} \right) \right\|_{H_2} \leq C$$

which is equivalent to

$$\sup_{\|x\|_{H_1}=1} \|Lx\|_{H_2} = K < \infty, \tag{6.27}$$

since $x / \|x\|_{H_1}$ is a unit vector in H_1. Clearly $K \leq C$.

Proposition 6.3. *A linear operator $L : H_1 \to H_2$ is bounded if and only if it is continuous.*

---

[7] Notation: if L is linear, when no confusion arises, we may write Lx instead of $L(x)$.

Proof. Let L be bounded. From (6.26) we have, $\forall x, x_0 \in H_1$,

$$\|L(x - x_0)\|_{H_2} \leq C \|x - x_0\|_{H_1}$$

so that, if $\|x - x_0\|_{H_1} \to 0$, also $\|Lx - Lx_0\|_{H_2} = \|L(x - x_0)\|_{H_2} \to 0$. This shows the continuity of L.

Let L be continuous. In particular, L is continuous at $x = 0$ so that there exists δ such that

$$\|Lx\|_{H_2} \leq 1 \quad \text{if } \|x\|_{H_1} \leq \delta.$$

Choose now $y \in H_1$ with $\|y\|_{H_1} = 1$ and let $z = \delta y$. We have $\|z\|_{H_1} = \delta$ which implies

$$\delta \|Ly\|_{H_2} = \|Lz\|_{H_2} \leq 1$$

or

$$\|Ly\|_{H_2} \leq \frac{1}{\delta}$$

and (6.27) holds with $K \leq C = \frac{1}{\delta}$. $\square$

Given two Hilbert spaces H_1 and H_2, we denote by

$$\mathcal{L}(H_1, H_2)$$

the *family of all linear bounded operators from H_1 into H_2*. If $H_1 = H_2$ we simply write $\mathcal{L}(H)$. $\mathcal{L}(H_1, H_2)$ becomes a linear space if we define, for $x \in H_1$ and $\lambda \in \mathbb{R}$,

$$(G + L)(x) = Gx + Lx$$
$$(\lambda L)x = \lambda Lx.$$

Also, we may use the number K in (6.27) as a norm in $\mathcal{L}(H_1, H_2)$:

$$\|L\|_{\mathcal{L}(H_1, H_2)} = \sup_{\|x\|_{H_1}=1} \|Lx\|_{H_2}. \tag{6.28}$$

When no confusion arises we will write simply $\|L\|$ instead of $\|L\|_{\mathcal{L}(H_1, H_2)}$. Thus, for every $L \in \mathcal{L}(H_1, H_2)$, we have

$$\|Lx\|_{H_2} \leq \|L\| \|x\|_{H_1}.$$

The resulting space is complete, so that:

Proposition 6.4. *Endowed with the norm (6.28), $\mathcal{L}(H_1, H_2)$ is a Banach space.*

Example 6.6. Let $\mathbf{A}$ be an $m \times n$ real matrix. The map

$$L : \mathbf{x} \longmapsto \mathbf{A}\mathbf{x}$$

is a linear operator from $\mathbb{R}^n$ into $\mathbb{R}^m$. To compute $\|L\|$, note that

$$\|\mathbf{A}\mathbf{x}\|^2 = \mathbf{A}\mathbf{x} \cdot \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{A}\mathbf{x} \cdot \mathbf{x}.$$

The matrix $\mathbf{A}^\top \mathbf{A}$ is symmetric and nonnegative and therefore, from Linear Algebra,

$$\sup_{\|\mathbf{x}\|=1} \mathbf{A}^\top \mathbf{A}\mathbf{x} \cdot \mathbf{x} = \Lambda_M$$

where Λ_M is the maximum eigenvalue of $\mathbf{A}^\top \mathbf{A}$. Thus, $\|L\| = \sqrt{\Lambda_M}$.

Example 6.7. Let V be a *closed subspace* of a Hilbert space H. The projections

$$x \longmapsto P_V x, \qquad x \longmapsto Q_V x,$$

defined in Theorem 6.2, are bounded linear operators from H into H. In fact, from $\|x\|^2 = \|P_V x\|^2 + \|Q_V x\|^2$, it follows immediately that

$$\|P_V x\| \leq \|x\|, \qquad \|Q_V x\| \leq \|x\|$$

so that (6.26) holds with $C = 1$. Since $P_V x = x$ when $x \in V$ and $Q_V x = x$ when $x \in V^\perp$, it follows that $\|P_V\| = \|Q_V\| = 1$. Finally, observe that

$$\mathcal{N}(P_V) = \mathcal{R}(Q_V) = V^\perp \quad \text{and} \quad \mathcal{N}(Q_V) = \mathcal{R}(P_V) = V.$$

Example 6.8. Let V and H be Hilbert spaces with[8] $V \subset H$. Considering an element in V as an element of H, we define the operator $I_{V \to H} : V \to H$,

$$I_{V \to H}(u) = u,$$

which is called *embedding of V into H*. $I_{V \to H}$ is clearly a linear operator and it is also bounded if there exists a constant C such that

$$\|u\|_H \leq C \|u\|_V, \qquad \text{for every } u \in V$$

In this case, we say that V *is continuously embedded in H* and we write

$$V \hookrightarrow H.$$

For instance, $H^1_{per}(0, 2\pi) \hookrightarrow L^2(0, 2\pi)$.

6.5.2 Functionals and dual space

When $H_2 = \mathbb{R}$ (or $\mathbb{C}$, for complex Hilbert spaces), a linear operator $L : H \to \mathbb{R}$ takes the name of **functional**.

Definition 6.5. *The collection of all bounded linear functionals on a Hilbert space H is called **dual space** of H and denoted by H^* (instead of $\mathcal{L}(H, \mathbb{R})$).*

---

[8] The inner products in V and H may be different.

Example 6.9. Let $H = L^2(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ and fix $g \in L^2(\Omega)$. The functional defined by

$$L_g : f \longmapsto \int_\Omega fg$$

is linear and bounded. In fact, Schwarz's inequality yields

$$|L_g f| = \left| \int_\Omega fg \right| \leq \left(\int_\Omega |f|^2 \right)^{1/2} \left(\int_\Omega |g|^2 \right)^{1/2} = \|g\|_0 \|f\|_0$$

so that $L_g \in L^2(\Omega)^*$ and $\|L_g\| \leq \|g\|_0$. Actually $\|L_g\| = \|g\|_0$ since, choosing $f = g$, we have

$$\|g\|_0^2 = L_g(g) \leq \|L_g\| \|g\|_0$$

whence also $\|L_g\| \geq \|g\|_0$.

Example 6.10. The functional in Example 6.13 is induced by the inner product with a fixed element in $L^2(\Omega)$. More generally, let H be a Hilbert space. For fixed $y \in H$, the functional

$$L_1 : x \longmapsto (x, y)$$

is continuous. In fact Schwarz's inequality yields $|(x,y)| \leq \|x\| \|y\|$, whence $L_1 \in H^*$ and $\|L_1\| \leq \|y\|$. Actually $\|L_1\| = \|y\|$ since, choosing $x = y$, we have

$$\|y\|^2 = |L_1 y| \leq \|L_1\| \|y\|,$$

or $\|L_1\| \geq \|y\|$. Observe that this argument provides the following alternative definition of the norm of an element $y \in H$:

$$\|y\| = \sup_{\|x\|=1} (x, y). \tag{6.29}$$

To identify the dual space of a Hilbert space H is crucial in many instances. Example 6.14 shows that the inner product with a fixed element y in H defines an element of H^*, whose norm is exactly $\|y\|$. From Linear Algebra it is well known that *all* linear functionals in a finite-dimensional space can be represented in that way. Precisely, if L is linear in $\mathbb{R}^n$, there exists a vector $\mathbf{a} \in \mathbb{R}^n$ such that, for every $\mathbf{h} \in \mathbb{R}^n$,

$$L\mathbf{h} = \mathbf{a} \cdot \mathbf{h}$$

and $\|L\| = |\mathbf{a}|$. The following theorem says that an analogous result holds in Hilbert spaces.

Theorem 6.3. (Riesz's Representation Theorem). *Let H be a Hilbert space. For every $L \in H^*$ there exists a unique $u_L \in H$ such that:*

1. $Lx = (u_L, x)$ *for every $x \in H$,*
2. $\|L\| = \|u_L\|$.

Proof. Let $\mathcal{N}$ be the kernel of L. If $\mathcal{N} = H$, then L is the *null operator* and $u_L = 0$. If $\mathcal{N} \subset H$, then $\mathcal{N}$ is a *closed* subspace of H. In fact, if $\{x_n\} \subset \mathcal{N}$ and $x_n \to x$, then $0 = Lx_n \to Lx$ so that $x \in \mathcal{N}$; thus $\mathcal{N}$ contains all its limit points and therefore is closed.

Then, by the Projection Theorem, there exists $z \in \mathcal{N}^\perp$, $z \neq 0$. Thus $Lz \neq 0$ and, given any $x \in H$, the element

$$w = x - \frac{Lx}{Lz} z$$

belongs to $\mathcal{N}$. In fact

$$Lw = L\left(x - \frac{Lx}{Lz} z\right) = Lx - \frac{Lx}{Lz} Lz = 0.$$

Since $z \in \mathcal{N}^\perp$, we have

$$0 = (z, w) = (z, x) - \frac{Lx}{Lz} \|z\|^2$$

which entails

$$Lx = \frac{L(z)}{\|z\|^2} (z, x).$$

Therefore if $u_L = L(z) \|z\|^{-2} z$, then $Lx = (u_L, x)$.

For the uniqueness, observe that, if $v \in H$ and

$$Lx = (v, x) \quad \text{for every } x \in H,$$

subtracting this equation from $Lx = (u_L, x)$, we infer

$$(u_L - v, x) = 0 \qquad \text{for every } x \in H$$

which forces $v = u_L$.

To show $\|L\| = \|u_L\|$, use Schwarz's inequality

$$|(u_L, x)| \leq \|x\| \|u_L\|$$

to get

$$\|L\| = \sup_{\|x\|=1} |Lx| = \sup_{\|x\|=1} |(u_L, x)| \leq \|u_L\|.$$

On the other hand,

$$\|u_L\|^2 = (u_L, u_L) = Lu_L \leq \|L\| \|u_L\|$$

whence

$$\|u_L\| \leq \|L\|.$$

Thus $\|L\| = \|u_L\|$. $\square$

The Riesz's map $R : H^* \to H$ given by

$$L \longmapsto u_L$$

is a *canonical isometry*, since it preserves the norm:

$$\|L\| = \|u_L\|.$$

We say that u_L is *the Riesz element associated with L*, with respect to the scalar product $(\cdot, \cdot)$. Moreover, H^* endowed with the inner product

$$(L_1, L_2)_{H^*} = (u_{L_1}, u_{L_2})$$

is clearly a Hilbert space. Thus, in the end, the Representation Theorem allows the **identification of a Hilbert space with its dual**.

Typically, $L^2(\Omega)$ or l_2 are identified with their duals.

Remark 6.5. **Warning**: there are situations in which the above canonical identification requires some care. A typical case we shall meet later occurs when dealing with a pair of Hilbert spaces V, H such that

$$V \hookrightarrow H \quad \text{and} \quad H^* \hookrightarrow V^*.$$

As we will see in subsection 6.8.1, in this conditions it is possible to identify H and H^* and write

$$V \hookrightarrow H \hookrightarrow V^*,$$

but at this point the identification of V with V^* is forbidden, since it would give rise to nonsense!

Remark 6.6. A few words about **notations.** The symbol $(\cdot, \cdot)$ or $(\cdot, \cdot)_H$ denotes the inner product in a Hilbert space H. Let now $L \in H^*$. For the *action* of the functional L on an element $x \in H$ we used the symbol Lx. Sometimes, when it is useful or necessary to emphasize the *duality (or pairing)* between H and H^*, we shall use the notation $\langle L, x \rangle_*$ or even $_{H^*}\langle L, x \rangle_H$.

6.5.3 The adjoint of a bounded operator

The concept of *adjoint operator* extends the notion of transpose of an $m \times n$ matrix $\mathbf{A}$ and plays a crucial role in determining compatibility conditions for the solvability of several problems. The transpose $\mathbf{A}^\top$ is characterized by the identity

$$(\mathbf{A}\mathbf{x}, \mathbf{y})_{\mathbb{R}^m} = (\mathbf{x}, \mathbf{A}^\top \mathbf{y})_{\mathbb{R}^n}, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \forall \mathbf{y} \in \mathbb{R}^m.$$

We extend precisely this relation to define the adjoint of a bounded linear operator. Let $L \in \mathcal{L}(H_1, H_2)$. If $y \in H_2$ is fixed, the real map

$$T_y : x \longmapsto (Lx, y)_{H_2}$$

defines an element of H_1^*. In fact

$$|T_y x| = \left|(Lx,y)_{H_2}\right| \leq \|Lx\|_{H_2} \|y\|_{H_2} \leq \|L\|_{\mathcal{L}(H_1,H_2)} \|y\|_{H_2} \|x\|_{H_1}$$

so that $\|T_y\| \leq \|L\|_{\mathcal{L}(H_1,H_2)} \|y\|_{H_2}$.

From Riesz's Theorem, there exists a unique $w \in H_1$ depending on y, which we denote by $w = L^* y$, such that

$$T_y x = (x, L^* y)_{H_1} \qquad \forall x \in H_1, \forall y \in H_2.$$

This defines L^* as an operator from H_2 into H_1, which is called the *adjoint of* L. Precisely:

Definition 6.6. *The operator* $L^* : H_2 \to H_1$ *defined by the identity*

$$(Lx,y)_{H_2} = (x, L^* y)_{H_1}, \qquad \forall x \in H_1, \forall y \in H_2 \tag{6.30}$$

is called the **adjoint of** L.

Example 6.11. Let $R : H^* \to H$ the Riesz operator. Then $R^* = R^{-1} : H \to H^*$. In fact, for every $F \in H^*$ and $v \in H$, we have:

$$(RF, v)_H = \langle F, v \rangle_* = \left(F, R^{-1}v\right)_{H^*}.$$

Example 6.12. Let $T : L^2(0,1) \to L^2(0,1)$ be the linear map

$$Tu(x) = \int_0^x u(t)\, dt.$$

Schwarz's inequality gives

$$\left|\int_0^x u\right|^2 \leq x \int_0^x u^2,$$

whence

$$\|Tu\|_0^2 = \int_0^1 |Tu|^2 = \int_0^1 \left|\int_0^x u\right|^2 dx \leq \int_0^1 (x \int_0^x u^2) dx \leq \frac{1}{2}\int_0^1 u^2 \leq \frac{1}{2}\|u\|_0^2$$

and therefore T is bounded. To compute T^*, observe that

$$(Tu,v)_0 = \int_0^1 [v(x)\int_0^x u(y)\, dy]\, dx = \text{exchanging the order of integration}$$

$$= \int_0^1 [u(y)\int_y^1 v(x)\, dx]\, dy = (u, T^* v)_0.$$

Thus,

$$T^* v(x) = \int_x^1 v(t)\, dt.$$

Symmetric matrices correspond to selfadjoint operators. We say that L is **self-adjoint** if $H_1 = H_2$ and $L^* = L$. Then, (6.30) reduces to

$$(Lx,y) = (x,Ly).$$

An example of a selfadjoint operator in a Hilbert space H is the projection P_V on a closed subspace of H; in fact, recalling the Projection Theorem:

$$(P_V x, y) = (P_V x, P_V y + Q_V y) = (P_V x, P_V y) = (P_V x + Q_V x, P_V y) = (x, P_V y).$$

Important self-adjoint operators are associated with *inverses of differential operators*, as we will see in Chapter 8.

The following properties are immediate consequences of the definition of adjoint (for the proof, see Problem 6.10).

Proposition 6.5. *Let $L, L_1 \in \mathcal{L}(H_1, H_2)$ and $L_2 \in \mathcal{L}(H_2, H_3)$. Then:*

(a) $L^* \in \mathcal{L}(H_2, H_1)$. *Moreover* $L^{**} = L$ *and*

$$\|L^*\|_{\mathcal{L}(H_2, H_1)} = \|L\|_{\mathcal{L}(H_1, H_2)}.$$

(b) $(L_2 L_1)^* = L_1^* L_2^*$. *In particular, if L is an isomorphism, then*

$$\left(L^{-1}\right)^* = (L^*)^{-1}.$$

The next theorem extends relations well known in the finite-dimensional case.

Theorem 6.4. *Let $L \in \mathcal{L}(H_1, H_2)$. Then*

a) $\overline{\mathcal{R}(L)} = \mathcal{N}(L^*)^\perp$

b) $\mathcal{N}(L) = \mathcal{R}(L^*)^\perp$.

Proof. a) Let $z \in \mathcal{R}(L)$. Then, there exists $x \in H_1$ such that $z = Lx$ and, if $y \in \mathcal{N}(L^*)$, we have

$$(z,y)_{H_2} = (Lx,y)_{H_2} = (x,L^*y)_{H_1} = 0.$$

Thus, $\mathcal{R}(L) \subseteq \mathcal{N}(L^*)^\perp$. Since $\mathcal{N}(L^*)^\perp$ is closed[9], it follows that

$$\overline{\mathcal{R}(L)} \subseteq \mathcal{N}(L^*)^\perp$$

as well. On the other hand, if $z \in \mathcal{R}(L)^\perp$, for every $x \in H_1$ we have

$$0 = (Lx,z)_{H_2} = (x,L^*z)_{H_1}$$

whence $L^*z = 0$. Therefore

$$\mathcal{R}(L)^\perp \subseteq \mathcal{N}(L^*),$$

equivalent to

$$\mathcal{N}(L^*)^\perp \subseteq \overline{\mathcal{R}(L)}.$$

b) letting $L = L^*$ in *a)* we deduce

$$\overline{\mathcal{R}(L^*)} = \mathcal{N}(L)^\perp,$$

equivalent to $\mathcal{R}(L^*)^\perp = \mathcal{N}(L)$. $\square$

---

[9] Remark 6.8.

6.6 Abstract Variational Problems

6.6.1 Bilinear forms and the Lax-Milgram Theorem

In the variational formulation of boundary value problems a key role is played by *bilinear forms*. Given two linear spaces V_1, V_2, a **bilinear form in** $V_1 \times V_2$ is a function

$$a : V_1 \times V_2 \to \mathbb{R}$$

satisfying the following properties:

i) For every $y \in V_2$, the function

$$x \longmapsto a(x, y)$$

is linear in V_1.

ii) For every $x \in V_1$, the function

$$y \longmapsto a(x, y)$$

is linear in V_2.

When $V_1 = V_2$, we simply say that a is a *bilinear form in* V.

Remark 6.7. In complex inner product spaces we define *sesquilinear forms*, instead of bilinear forms, replacing **ii)** by:

ii$_{bis}$) for every $x \in V_1$, the function

$$y \longmapsto a(x, y)$$

is *anti-linear*[10] in V_2.

Here are some examples.

- A typical example of bilinear form in a Hilbert space is its inner product.
- The formula

$$a\,(u, v) = \int_a^b \left(p(x)u'v' + q(x)u'v + r(x)uv \right)\ dx$$

where p, q, r are bounded functions, defines a bilinear form in $C^1\left([a, b]\right)$.
More generally, if Ω is a bounded domain in $\mathbb{R}^n$,

$$a(u,v) = \int_\Omega \left(\alpha\ \nabla u \cdot \nabla v + u\mathbf{b}\,(\mathbf{x}) \cdot \nabla v + a_0\,(\mathbf{x})\, uv \right)\ d\mathbf{x} \qquad (\alpha > 0)\,,$$

---

[10] That is

$$a\,(x, \alpha y + \beta z) = \overline{\alpha}a\,(x, y) + \overline{\beta}a\,(x, z)$$

or

$$a(u,v) = \int_\Omega \alpha \, \nabla u \cdot \nabla v \, d\mathbf{x} + \int_{\partial\Omega} huv \, d\sigma \qquad (\alpha > 0) \, ,$$

($\mathbf{b}$, a_0, h bounded) are bilinear forms in $C^1\left(\overline{\Omega}\right)$.

• A bilinear form in $C^2\left(\overline{\Omega}\right)$ involving higher order derivatives is

$$a\left(u, v\right) = \int_\Omega \Delta u \, \Delta v \, d\mathbf{x}.$$

Let V be a Hilbert space, a be a bilinear form in V and $F \in V^*$. Consider the following problem, called *abstract variational problem*:

$$\begin{cases} \quad \text{Find } u \in V \\ \qquad \text{such that} \\ a\left(u, v\right) = \langle F, v \rangle_* \quad \forall v \in V. \end{cases} \tag{6.31}$$

As we shall see, many boundary values problems can be recast in this form. The fundamental result is:

Theorem 6.5. (Lax $-$ Milgram). *Let V be a real Hilbert space endowed with inner product $(\cdot, \cdot)$ and norm $\|\cdot\|$. Let $a = a\left(u, v\right)$ be a bilinear form in V. If:*

*i) a is **continuous**, i.e. there exists a constant M such that*

$$|a(u, v)| \leq M \, \|u\| \, \|v\| \, , \qquad \forall u, v \in V;$$

*ii) a is $V-$**coercive**, i.e. there exists a constant $\alpha > 0$ such that*

$$a(v, v) \geq \alpha \, \|v\|^2 \, , \qquad \forall v \in V, \tag{6.32}$$

then there exists a unique solution $\overline{u} \in V$ of problem (6.31). Moreover, the following stability estimate holds:

$$\|\overline{u}\| \leq \frac{1}{\alpha} \, \|F\|_{V^*} \, . \tag{6.33}$$

Remark 6.8. The coercivity inequality (6.32) may be considered as an abstract version of the *energy* or *integral estimates* we met in the previous chapters. Usually, it is the key estimate to prove in order to apply Theorem 6.5. We shall come back to the general solvability of a variational problem in Section 6.8, when a is not $V-$coercive.

Remark 6.9. Inequality (6.61) is called *stability estimate* for the following reason. The functional F, element of V^*, encodes the "data" of the problem (6.31). Since for every F there is a unique solution $u(F)$, the map

$$F \longmapsto u(F)$$

is a well defined *function* from V^* onto V. Also, everything here has a linear nature, so that the solution map is linear as well. To check it, let λ, $\mu \in \mathbb{R}$, F_1, $F_2 \in V^*$ and u_1, u_2 the corresponding solutions. The bilinearity of a, gives

$$a(\lambda u_1 + \mu u_2, v) = \lambda a(u_1, v) + \mu a(u_2, v) =$$
$$= \lambda F_1 v + \mu F_2 v.$$

Therefore, the same linear combination of the solutions corresponds to a linear combination of the data; this expresses the *principle of superposition* for problem (6.31). Applying now (6.33) to $u_1 - u_2$, we obtain

$$\|u_1 - u_2\| \leq \frac{1}{\alpha} \|F_1 - F_2\|_{V^*}.$$

Thus, close data imply close solutions. The stability constant $1/\alpha$ plays an important role, since it controls the norm-variation of the solutions in terms of the variations on the data, measured by $\|F_1 - F_2\|_{V^*}$. This entails, in particular, that the more the coercivity constant α is large, the more "stable" is the solution.

Proof of theorem 6.5. We split it into several steps.

1. *Reformulation of problem* (6.31). For every fixed $u \in V$, by the continuity of a, the linear map

$$v \mapsto a(u, v)$$

is bounded in V and therefore it defines an element of V^*. From Riesz's Representation Theorem, there exists a unique $A[u] \in V$ such that

$$a(u, v) = (A[u], v), \qquad \forall v \in V. \tag{6.34}$$

Since $F \in V^*$ as well, there exists a unique $z_F \in V$ such that

$$Fv = (z_F, v) \qquad \forall v \in V$$

and moreover $\|F\|_{V^*} = \|z_F\|$. Then, problem (6.31) can be recast in the following way:

$$\begin{cases} \text{Find } u \in V \\ \quad \text{such that} \\ (A[u], v) = (z_F, v), \quad \forall v \in V \end{cases}$$

which, in turn, is equivalent to **finding** u **such that**

$$A[u] = z_F. \tag{6.35}$$

We want to show that (6.35) has exactly one solution. To do this we show that

$$A : V \to V$$

is a *linear, continuous, one-to-one, surjective* map.

2. *Linearity and continuity of* A. We repeatedly use the definition of A and the bilinearity of a. To show linearity, we write, for every $u_1, u_2, v \in V$ and $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$(A\left[\lambda_1 u_1 + \lambda_2 u_2\right], v) = a\left(\lambda_1 u_1 + \lambda_2 u_2, v\right) = \lambda_1 a\left(u_1, v\right) + \lambda_2 a\left(u_2, v\right)$$
$$= \lambda_1 \left(A\left[u_1\right], v\right) + \lambda_2 \left(A\left[u_2\right], v\right) = \left(\lambda_1 A\left[u_1\right] + \lambda_2 A\left[u_2\right], v\right)$$

whence

$$A\left[\lambda_1 u_1 + \lambda_2 u_2\right] = \lambda_1 A\left[u_1\right] + \lambda_2 A\left[u_2\right].$$

Thus A is linear and we may write Au instead of $A\left[u\right]$. For the continuity, observe that

$$\|Au\|^2 = (Au, Au) = a(u, Au)$$
$$\leq M\|u\|\|Au\|$$

whence

$$\|Au\| \leq M\|u\|.$$

3. *A is one-to-one and has closed range*, i.e.

$$\mathcal{N}(A) = \{0\} \quad \text{and} \quad \mathcal{R}(A) \text{ is a closed subspace of } V.$$

In fact, the coercivity of a yields

$$\alpha\|u\|^2 \leq a(u, u) = (Au, u) \leq \|Au\|\|u\|$$

whence

$$\|u\| \leq \frac{1}{\alpha}\|Au\|. \tag{6.36}$$

Thus, $Au = 0$ implies $u = 0$ and hence $\mathcal{N}(A) = \{0\}$.

To prove that $\mathcal{R}(A)$ is closed we have to consider a sequence $\{y_m\} \subset \mathcal{R}(A)$ such that

$$y_m \to y \in V$$

as $m \to \infty$, and show that $y \in \mathcal{R}(A)$. Since $y_m \in \mathcal{R}(A)$, there exists u_m such that $Au_m = y_m$. From (6.36) we infer

$$\|u_k - u_m\| \leq \frac{1}{\alpha}\|y_k - y_m\|$$

and therefore, since $\{y_m\}$ is convergent, $\{u_m\}$ is a Cauchy sequence. Since V is complete, there exists $u \in V$ such that

$$u_m \to u$$

and the continuity of A yields $y_m = Au_m \to Au$. Thus $Au = y$, so that $y \in \mathcal{R}(A)$ and $\mathcal{R}(A)$ is closed.

4. *A is surjective, that is* $\mathcal{R}(A) = V$. Suppose $\mathcal{R}(A) \subset V$. Since $\mathcal{R}(A)$ is a closed subspace, by the Projection Theorem there exists $z \neq 0$, $z \in \mathcal{R}(A)^{\perp}$. In particular, this implies

$$0 = (Az, z) = a(z, z) \geq \alpha \|z\|^2$$

whence $z = 0$. Contradiction. Therefore $\mathcal{R}(A) = V$.

5. *Solution of problem* (6.31). Since A is one-to-one and $\mathcal{R}(A) = V$, there exists exactly one solution $\overline{u} \in V$ of equation

$$Au = z_F.$$

From point **1**, $\overline{u}$ is the unique solution of problem (6.31) as well.

6. *Stability estimate.* From (6.36) with $u = \overline{u}$, we obtain

$$\|\overline{u}\| \leq \frac{1}{\alpha} \|A\overline{u}\| = \frac{1}{\alpha} \|z_F\| = \frac{1}{\alpha} \|F\|_{V^*}$$

and the proof is complete. $\square$

Remark 6.10. Some applications require the solution to be in some Hilbert space W, while asking the variational equation

$$a(u, v) = \langle F, v \rangle_*$$

to hold for every $v \in V$, with $V \neq W$. A variant of Theorem 6.5 deals with this asymmetric situation. Let $F \in V^*$ and $a = a(u, v)$ be a bilinear form in $W \times V$ satisfying the following three hypotheses:

i) *there exists M such that*

$$|a(u, v)| \leq M \|u\|_W \|v\|_V, \qquad \forall u \in W, \forall v \in V;$$

ii) *there exists $\alpha > 0$ such that*

$$\sup_{\|v\|_V = 1} a(u, v) \geq \alpha \|u\|_W, \qquad \forall u \in W;$$

iii)

$$\sup_{w \in W} a(w, v) > 0, \qquad \forall v \in V.$$

Condition *ii)* is an asymmetric coercivity, while *iii)* assures that, for every fixed $v \in V$, $a(v, \cdot)$ is positive at some point in W. We have (for the proof see Problem 6.11):

Theorem 6.6. (Neças). *If i), ii), iii) hold, there exists a unique $u \in W$ such that*

$$a(u, v) = \langle F, v \rangle_*, \qquad \forall v \in V.$$

Moreover

$$\|u\|_W \leq \frac{1}{\alpha} \|F\|_{V^*}. \tag{6.37}$$

6.6.2 Minimization of quadratic functionals

When a is *symmetric*, i.e. if

$$a\left(u,v\right)=a\left(v,u\right)\qquad\forall u,v\in V,$$

the abstract variational problem (6.31) is equivalent to a *minimization* problem. In fact, consider the quadratic functional

$$E\left(v\right)=\frac{1}{2}a\left(v,v\right)-\left\langle F,v\right\rangle_{*}.$$

We have:

Theorem 6.7. *Let a be symmetric. Then $\overline{u}$ is solution of problem (6.31) if and only if $\overline{u}$ is a minimizer of E, that is*

$$E\left(\overline{u}\right)=\min_{v\in V}E\left(v\right).$$

Proof. For every $\varepsilon\in\mathbb{R}$ and every "variation" $v\in V$ we have

$$E\left(\overline{u}+\varepsilon v\right)-E\left(\overline{u}\right)$$
$$=\left\{\frac{1}{2}a\left(\overline{u}+\varepsilon v,\overline{u}+\varepsilon v\right)-\left\langle F,\overline{u}+\varepsilon v\right\rangle_{*}\right\}-\left\{\frac{1}{2}a\left(\overline{u},\overline{u}\right)-\left\langle F,\overline{u}\right\rangle_{*}\right\}$$
$$=\varepsilon\left\{a\left(\overline{u},v\right)-\left\langle F,v\right\rangle_{*}\right\}+\frac{1}{2}\varepsilon^{2}a\left(v,v\right).$$

Now, if $\overline{u}$ is the solution of problem (6.31), then $a\left(\overline{u},v\right)-\left\langle F,v\right\rangle_{*}=0$. Therefore

$$E\left(\overline{u}+\varepsilon v\right)-E\left(\overline{u}\right)=\frac{1}{2}\varepsilon^{2}a\left(v,v\right)\geq0$$

so that $\overline{u}$ minimizes E. On the other hand, if $\overline{u}$ is a minimizer of E, then

$$E\left(\overline{u}+\varepsilon v\right)-E\left(\overline{u}\right)\geq0,$$

which entails

$$\varepsilon\left\{a\left(\overline{u},v\right)-\left\langle F,v\right\rangle_{*}\right\}+\frac{1}{2}\varepsilon^{2}a\left(v,v\right)\geq0.$$

This inequality forces (why?)

$$a\left(\overline{u},v\right)-\left\langle F,v\right\rangle_{*}=0\qquad\forall v\in V\tag{6.38}$$

and $\overline{u}$ is a solution of problem (6.31)). $\square$

Letting $\varphi\left(\varepsilon\right)=E\left(\overline{u}+\varepsilon v\right)$, from the above calculations we have

$$\varphi'(0)=a\left(\overline{u},v\right)-\left\langle F,v\right\rangle_{*}.$$

Thus, the linear functional

$$v \longmapsto a\left(\overline{u}, v\right) - \langle F, v\rangle_*$$

appears as the **derivative of** E at $\overline{u}$ along the direction v and we write

$$E'\left(\overline{u}\right)v = a\left(\overline{u}, v\right) - \langle F, v\rangle_* . \tag{6.39}$$

In Calculus of Variation E' is called **first variation** and denoted by δE.

If a is symmetric, the *variational equation*

$$E'\left(u\right)v = a\left(u, v\right) - \langle F, v\rangle_* = 0, \qquad \forall v \in V \tag{6.40}$$

is called **Euler equation** for the functional E.

Remark 6.11. A bilinear form a, symmetric and coercive, induces in V the inner product

$$\left(u, v\right)_a = a\left(u, v\right).$$

In this case, existence, uniqueness and stability for problem (6.31) follow directly from Riesz's Representation Theorem. In particular, *there exists a unique minimizer* $\overline{u}$ *of* E.

6.6.3 Approximation and Galerkin method

The solution u of the abstract variational problem (6.31), satisfies the equation

$$a\left(u, v\right) = \langle F, v\rangle_* \tag{6.41}$$

for *every* v in the Hilbert space V. In concrete applications, it is important to compute approximate solutions with a given degree of accuracy and the infinite dimension of V is the main obstacle. Often, however, V may be written as a *union of finite-dimensional subspaces*, so that, in principle, it could be reasonable to obtain approximate solutions by "projecting" equation (6.41) on those subspaces. This is the idea of **Galerkin's method**. In principle, the higher the dimension of the subspace the better should be the degree of approximation. More precisely, the idea is to construct a sequence $\{V_k\}$ of subspaces of V with the following properties:

a) Every V_k is *finite-dimensional*: $\dim V_k = k$,

b) $V_k \subset V_{k+1}$ (actually, not strictly necessary),

c) $\overline{\cup V_k} = V$.

To realize the projection, assume that the vectors $\psi_1, \psi_2, ..., \psi_k$ span V_k. Then, we look for an approximation of the solution u in the form

$$u_k = \sum_{j=1}^{k} c_j \psi_j, \tag{6.42}$$

by solving the *projected* problem

$$a\left(u_k, v\right) = \langle F, v \rangle_* \qquad \forall v \in V_k. \tag{6.43}$$

Since $\{\psi_1, \psi_2, ..., \psi_k\}$ constitutes a basis in V_k, (6.43) amounts to requiring

$$a\left(u_k, \psi_r\right) = \langle F, \psi_r \rangle_* \qquad r = 1, ..., k. \tag{6.44}$$

Substituting (6.42) into (6.44), we obtain the k linear algebraic equations

$$\sum_{j=1}^{k} c_j a\left(\psi_j, \psi_r\right) = \langle F, \psi_r \rangle_* \qquad r = 1, 2, ..., k \tag{6.45}$$

for the unknown coefficients $c_1, c_2, ..., c_k$. Introducing the vectors

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \langle F, \psi_1 \rangle_* \\ \langle F, \psi_2 \rangle_* \\ \vdots \\ \langle F, \psi_k \rangle_* \end{pmatrix}$$

and the matrix $\mathbf{A} = (a_{rj})$, with entries

$$a_{rj} = a\left(\psi_j, \psi_r\right), \qquad j, r = 1, ..., k,$$

we may write (6.45) in the compact form

$$\mathbf{A}\mathbf{c} = \mathbf{F}. \tag{6.46}$$

The matrix $\mathbf{A}$ is called *stiffness matrix* and clearly plays a key role in the numerical analysis of the problem.

If the bilinear form a is coercive, $\mathbf{A}$ is *strictly positive*. In fact, let $\boldsymbol{\xi} \in \mathbb{R}^k$. Then, by linearity and coercivity:

$$\mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} = \sum_{r,j=1}^{k} a_{rj} \xi_r \xi_j = \sum_{r,j=1}^{k} a\left(\psi_j, \psi_r\right) \xi_r \xi_j$$

$$= \sum_{r,j=1}^{k} a\left(\xi_j \psi_j, \xi_r \psi_r\right) = a\left(\sum_{i=1}^{k} \xi_j \psi_j, \sum_{j=1}^{k} \xi_r \psi_r\right)$$

$$\geq \alpha \left\| \mathbf{v} \right\|^2$$

where

$$\mathbf{v} = \sum_{j=1}^{k} \xi_j \psi_j \in V_k.$$

Since $\{\psi_1, \psi_2, ..., \psi_k\}$ is a basis in V_k, we have $\mathbf{v} = \mathbf{0}$ if and only if $\boldsymbol{\xi} = \mathbf{0}$. Therefore $\mathbf{A}$ is strictly positive and, in particular, non singular.

Thus, for each $k \geq 1$, there exists a unique solution $u_k \in V_k$ of (6.46). We want to show that $u_k \to u$, as $k \to \infty$, i.e. the *convergence of the method,* and give a control of the approximation error.

For this purpose, we prove the following lemma, which also shows the role of the continuity and the coercivity constants (M and α, respectively) of the bilinear form a.

Lemma 6.1. (Céa). *Assume that the hypotheses of the Lax-Milgram Theorem hold and let u be the solution of problem (6.31). If u_k is the solution of problem (6.44), then*

$$\|u - u_k\| \leq \frac{M}{\alpha} \inf_{v \in V_k} \|u - v\|. \tag{6.47}$$

Proof. We have

$$a(u_k, v) = \langle F, v \rangle_*, \qquad \forall v \in V_k$$

and

$$a(u, v) = \langle F, v \rangle_*, \qquad \forall v \in V_k.$$

Subtracting the two equations we obtain

$$a(u - u_k, v) = 0, \qquad \forall v \in V_k.$$

In particular, since $v - u_k \in V_k$, we have

$$a(u - u_k, v - u_k) = 0, \qquad \forall v \in V_k$$

which implies

$$a(u - u_k, u - u_k) = a(u - u_k, u - v) + a(u - u_k, v - u_k)$$
$$= a(u - u_k, u - v).$$

Then, by the coercivity of a,

$$\alpha \|u - u_k\|^2 \leq a(u - u_k, u - u_k) \leq M \|u - u_k\| \|u - v\|$$

whence,

$$\|u - u_k\| \leq \frac{M}{\alpha} \|u - v\|. \tag{6.48}$$

This inequality holds for every $v \in V_k$, with $\frac{M}{\alpha}$ independent of k. Therefore (6.48) still holds if we take in the right hand side the infimum over all $v \in V_k$. $\square$

Convergence of Galerkin's method. Since we have assumed that

$$\overline{\cup V_k} = V,$$

there exists a sequence $\{w_k\} \subset V_k$ such that $w_k \to u$ as $k \to \infty$. Céa's Lemma gives, for every k:

$$\|u - u_k\| \leq \frac{M}{\alpha} \inf_{v \in V_k} \|u - v\| \leq \frac{M}{\alpha} \|u - w_k\|$$

whence

$$\|u - u_k\| \to 0.$$

6.7 Compactness and Weak Convergence

6.7.1 Compactness

The solvability of boundary value problems and the analysis of numerical methods involve several questions of convergence. In typical situations one is able to construct a sequence of approximations and the main task is to prove that this sequence converges to a solution of the problem in a suitable sense. It is often the case that, through *energy type estimates*[11], one is able to show that these sequences of approximations are *bounded* in some Hilbert space. How can we use this information? Although we cannot expect these sequences to converge, we may reasonably look for *convergent subsequences*, which is already quite satisfactory. In technical words, we are asking to our sequences to have a *compactness property.* Let us spend a few words on this important topological concept [12]. Once more, the difference between finite and infinite dimension plays a big role.

Let X be a normed space. The general definition of compact set involves open coverings: an *open covering* of $E \subseteq X$ is a family of *open* sets whose union contains E.

Definition 6.7. (Compactness 1)*. We say that $E \subseteq X$ is **compact** if from every open covering of E it is possible to extract a finite subcovering of E.*

It is somewhat more convenient to work with **pre-compact** sets as well, i.e. sets whose **closure** is compact. In finite-dimensional spaces the characterization of pre-compact sets is well known: $E \subset \mathbb{R}^n$ is pre-compact if and only if E is bounded. What about infinitely many dimensions? Let us introduce a characterization of pre-compact sets in normed spaces in terms of convergent sequences, much more comfortable to use.

First, let us agree that a subset E of a normed space X is *sequentially pre-compact* (resp. *compact*), if for every sequence $\{x_k\} \subset E$ there exists a subsequence $\{x_{k_s}\}$, convergent in X (resp. in E).

We have:

Theorem 6.8. (Compactness 2)*. Let X be a normed space and $E \subset X$. Then E is **pre-compact (compact)** if and only if it is sequentially pre-compact (compact).*

While a compact set is always *closed and bounded* (see Problem 6.12), the following example exhibits a closed and bounded set which is *not* compact in l^2.

Consider the real Hilbert space

$$l^2 = \left\{ \mathbf{x} = \{x_k\}_{k=1}^\infty : \sum_{k=1}^\infty x_k^2 < \infty, x_k \in \mathbb{R} \right\}$$

---

[11] I.e. estimates for a function and its gradient in L^2.

[12] For the proofs see e.g. *Rudin*, 1964 or *Yhosida*, 1968.

endowed with

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\infty} x_k y_k \qquad \text{and} \qquad \|\mathbf{x}\|^2 = \sum_{k=1}^{\infty} x_k^2.$$

Let $E = \left\{\mathbf{e}^k\right\}_{k \geq 1}$, where $\mathbf{e}^1 = \{1, 0, 0, ...\}$, $\mathbf{e}^2 = \{0, 1, 0, ...\}$, etc.. Observe that E constitutes an orthonormal basis in l^2. Then, E is closed and bounded in l^2.

However, E *is not sequentially compact*. Indeed, $\left\|\mathbf{e}^j - \mathbf{e}^k\right\| = \sqrt{2}$, if $j \neq k$, and therefore no subsequence of $\left\{\mathbf{e}^k\right\}_{k \geq 1}$ can be convergent.

Thus, in infinite-dimensions, closed and bounded does not imply compact. Actually, this can only happen in finite-dimensional spaces. In fact:

Theorem 6.9. *Let B be a Banach space. B is finite-dimensional if and only if the unit ball $\{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ is compact.*

Criterion of compactness in L^2. To recognize that a subset of a Hilbert space is compact is usually a hard task. The following theorem gives a criterion for recognizing pre-compact sets $S \subset L^2(\Omega)$.

Theorem 6.10. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and $S \subset L^2(\Omega)$. If:*

i) S is bounded: i.e. there exists K such that $\|u\|_{L^2(\Omega)} \leq K, \forall u \in S$,

ii) there exist α and L, positive, such that, if u is extended by zero outside Ω,

$$\|u\left(\cdot + \mathbf{h}\right) - u\left(\cdot\right)\|_{L^2(\Omega)} \leq L\left|\mathbf{h}\right|^{\alpha}, \quad \text{for every } \mathbf{h} \in \mathbb{R}^n \text{ and } u \in S,$$

then S is pre-compact.

The second condition expresses an *equicontinuity in norm* of all the elements in S. We shall meet this condition in subsection 7.10.1.

6.7.2 Weak convergence and compactness

We have seen that the compactness in a normed space is equivalent to sequential compactness. In the applications, this translates into a very strong requirement for approximating sequences.

Fortunately, in normed spaces, and in particular in Hilbert spaces, there is another notion of convergence, much more flexible, which turns out to be perfectly adapted to the variational formulation of boundary value problems.

Let H be a Hilbert space with inner product $(\cdot, \cdot)$ and norm $\|\cdot\|$. If $F \in H^*$, we know that $\langle F, x_k \rangle_* \to \langle F, x \rangle_*$ when $\|x_k - x\| \to 0$. However, it could be that

$$\langle F, x_k \rangle_* \to \langle F, x \rangle_*$$

for every $F \in H^*$, even if $\|x_k - x\| \not\to 0$. Then, we say that x_k converges *weakly* to x. Precisely:

Definition 6.8. *A sequence* $\{x_k\} \subset H$ ***converges weakly to*** $x \in H$, *and we write*

$$x_k \rightharpoonup x$$

(with an "half arrow"), if

$$\langle F, x_k \rangle_* \to \langle F, x \rangle_* \,, \qquad \forall F \in H^*.$$

The convergence in norm is then called *strong convergence*. From Riesz's Representation Theorem, it follows that $\{x_k\} \subset H$ *converges weakly to* $x \in H$ if and only if

$$(x_k, y) \to (x, y) \,, \qquad \forall y \in H.$$

The weak limit is unique, since $x_k \rightharpoonup x$ and $x_k \rightharpoonup z$ implies

$$(x - z, y) = 0 \qquad \forall y \in H,$$

whence $x = z$. Moreover, Schwarz's inequality gives

$$|(x_k - x, y)| \le \|x_k - x\| \, \|y\|$$

so that *strong convergence* implies *weak convergence*, which should not be surprising.

The two notions of convergence are equivalent in finite-dimensional spaces. It is not so in infinite dimensions, as the following example shows.

Example 6.13. Let $H = L^2(0, 2\pi)$. The sequence $v_k(x) = \cos kx$, $k \ge 1$, is *weakly convergent to zero*. In fact, for every $f \in L^2(0, 2\pi)$, the Riemann-Lebesgue Theorem on the Fourier coefficients of f implies that

$$(f, v_k)_0 = \int_0^{2\pi} f(x) \cos kx \; dx \to 0$$

as $k \to \infty$. However

$$\|v_k\|_0 = \sqrt{\pi}$$

and therefore $\{v_k\}_{k \ge 1}$ does not converge strongly.

Remark 6.12. If $L \in \mathcal{L}(H_1, H_2)$ and $x_k \rightharpoonup x$ in H_1 we cannot say that $Lx_k \to Lx$ in H_2. However, by definition of weak convergence, $Lx_k \rightharpoonup Lx$ is true. Thus, if L is (strongly) continuous then it is *weakly continuous* as well.

Remark 6.13. **Warning**: Not always *strong implies weak*! Take a strongly closed set $E \subset H$. Can we deduce that E is *weakly closed* as well? The answer is *no*. Indeed, "strongly closed" means that E contains all the limits of strongly convergent sequences $\{x_k\} \subset E$. But suppose that $x_k \rightharpoonup x$ (only weakly); since the convergence is not strong, we can not affirm that $x \in E$. Thus, E is *not weakly closed*, in general[13].

For instance, let $E = \{v_k\}$ where $v_k(x) = \cos kx$, as in Example 6.23. Then, E is a strongly closed subset of $L^2(0, 2\pi)$ and contained in the set $\{\|v\|_0 = \sqrt{\pi}\}$. However $v_k \rightharpoonup 0 \notin E$, so that E is *not* weakly closed.

---

[13] See Problem 6.14.

We have observed that the norm in a Hilbert space is strongly continuous. With respect to weak convergence, the norm is only *lower semicontinuous,* as property 2 in the following theorem shows.

Theorem 6.11. *Let $\{x_k\} \subset H$ such that $x_k \rightharpoonup x$. Then*

   1) $\{x_k\}$ is bounded,

   2) $\|x\| \leq \liminf_{k\to\infty} \|x_k\|$.

We omit the proof of 1. For the second point, it is enough to observe that

$$\|x\|^2 = \lim_{k\to\infty} (x_k, x) \leq \|x\| \liminf_{k\to\infty} \|x_k\|$$

and simplify by $\|x\|$.

The usefulness of weak convergence is revealed by the following compactness result. Basically, it says that if we substitute *strong* with *weak* convergence, any bounded sequence in a Hilbert space is weakly pre-compact. Precisely:

Theorem 6.12. *Every bounded sequence in a Hilbert space H contains a subsequence which is weakly convergent to an element $x \in H$.*

Proof. We give it under the additional hypothesis that H is separable. Thus, there exists a sequence $\{z_k\}$ dense in H. Let now $\{x_j\} \subset H$ be a bounded sequence: $\|x_j\| \leq M, \forall j \geq 1$. We split the proof into three steps.

1. Using a "diagonal" process, we construct a subsequence $\{x_s^{(s)}\}$ such that the real sequence $(x_s^{(s)}, z_k)$ is convergent for every fixed z_k. To do this, observe that the sequence

$$(x_j, z_1)$$

is bounded in $\mathbb{R}$ and therefore there exists $\{x_j^{(1)}\} \subset \{x_j\}$ such that

$$(x_j^{(1)}, z_1)$$

is convergent. For the same reason, from $\{x_j^{(1)}\}$ we may extract a subsequence $\{x_j^{(2)}\}$ such that

$$(x_j^{(2)}, z_2)$$

is convergent. By induction, we construct $\{x_j^{(k)}\}$ such that

$$(x_j^{(k)}, z_k)$$

converges. Consider the diagonal sequence $\{x_s^{(s)}\}$, obtained by selecting $x_1^{(1)}$ from $\{x_j^{(1)}\}$, $x_2^{(2)}$ from $\{x_j^{(2)}\}$ and so on. Then,

$$(x_s^{(s)}, z_k)$$

is convergent for every fixed $k \geq 1$.

2. We use the density of $\{z_k\}$ to show that $(x_s^{(s)}, z_k)$ converges for every $z \in H$. In fact, for fixed $\varepsilon > 0$ and $z \in H$, we may find z_k such that $\|z - z_k\| < \varepsilon$. Write

$$(x_s^{(s)} - x_m^{(m)}, z) = (x_s^{(s)} - x_m^{(m)}, z - z_k) + (x_s^{(s)} - x_m^{(m)}, z_k).$$

If j and m are large enough

$$\left| (x_s^{(s)} - x_m^{(m)}, z_k) \right| < \varepsilon$$

since $(x_s^{(s)}, z_k)$ is convergent. Moreover, from Schwarz's inequality,

$$\left| (x_s^{(s)} - x_m^{(m)}, z - z_k) \right| \leq \left\| x_s^{(s)} - x_m^{(m)} \right\| \|z - z_k\| \leq 2M\varepsilon.$$

Thus, if j and m are large enough, we have

$$\left| (x_s^{(s)} - x_m^{(m)}, z) \right| \leq (2M + 1)\varepsilon,$$

hence the sequence $(x_s^{(s)} - x_m^{(m)}, z)$ is a Cauchy sequence in $\mathbb{R}$ and therefore convergent.

3. From **2**, we may define a linear functional T in H by setting

$$Tz = \lim_{s \to \infty} (x_s^{(s)}, z).$$

Since $\left\| x_s^{(s)} \right\| \leq M$, we have

$$|Tz| \leq M \|z\|$$

whence $T \in H^*$. From the Riesz Representation theorem, there exists a unique $x_\infty \in H$ such that

$$Tz = (x_\infty, z), \qquad \forall z \in H.$$

Thus

$$(x_s^{(s)}, z) \to (x_\infty, z), \qquad \forall z \in H$$

or

$$x_s^{(s)} \rightharpoonup x_\infty.$$

$\square$

Example 6.14. Let $H = L^2(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ and consider a sequence $\{u_k\}_{k \geq 1} \subset L^2(\Omega)$. To say that $\{u_k\}$ is bounded means that

$$\|u_k\|_0 \leq M, \qquad \text{for every } k \geq 1.$$

Theorem 6.11 implies the existence of a subsequence $\{u_{k_m}\}_{m \geq 1}$ and of $u \in L^2(\Omega)$ such that, as $m \to +\infty$,

$$\int_\Omega u_{k_m} v \to \int_\Omega uv, \qquad \text{for every } v \in L^2(\Omega).$$

6.7.3 Compact operators

By definition, every operator in $\mathcal{L}(H_1, H_2)$ transforms bounded sets in H_1 into bounded sets in H_2. The subclass of operators that transform *bounded sets* into *pre-compact* sets is particularly important.

Definition 6.9. *Let H_1 and H_2 Hilbert spaces and $L \in \mathcal{L}(H_1, H_2)$. We say that L is **compact** if, for every bounded $E \subset H_1$, the image $L(E)$ is **pre-compact** in H_2.*

An equivalent characterization of compact operators may be given in terms of weak convergence. Indeed, an operator is compact if and only if "converts weak convergence into strong convergence". Precisely:

Proposition 6.6. *Let $L \in \mathcal{L}(H_1, H_2)$. L is compact if and only if, for every sequence $\{x_k\} \subset H_1$,*

$$x_k \rightharpoonup 0 \quad in \ H_1 \qquad implies \qquad Lx_k \to 0 \quad in \ H_2. \tag{6.49}$$

Proof. Assume that (6.49) holds. Let $E \subset H_1$, bounded, and $\{z_k\} \subset L(E)$. Then $z_k = Lx_k$ with $x_k \in E$.

From theorem 6.11, there exists a subsequence $\{x_{k_s}\}$ weakly convergent to $x \in H_1$. Then $y_s = x_{k_s} - x \rightharpoonup 0$ in H_1 and, from (6.49), $Ly_{k_s} \to 0$ in H_2, that is $z_{k_s} = Lx_{k_s} \to Lx \equiv z$ in H_2.

Thus, $L(E)$ is sequentially pre-compact, and therefore pre-compact in H_2.

Viceversa, let L be compact and $x_k \rightharpoonup 0$ in H_1. Suppose $Lx_k \nrightarrow 0$. Then, for some $\bar{\varepsilon} > 0$ and infinitely many indexes k_j, we have $\|Lx_{k_j}\| > \bar{\varepsilon}$. Since

$$x_{k_j} \rightharpoonup 0,$$

by Theorem 6.10 $\{x_{k_j}\}$ is bounded in H_1, so that $\{Lx_{k_j}\}$ contains a subsequence (that we still call) $\{Lx_{k_j}\}$ strongly (and therefore weakly) convergent to some $y \in H_2$. On the other hand, we have $Lx_{k_j} \rightharpoonup 0$ as well, which entails $y = 0$. Thus $\|Lx_{k_j}\| \to 0$. Contradiction. $\square$

Example 6.15. Let $H^1_{per}(0, 2\pi)$ be the Hilbert space introduced in Example 6.4. The embedding

$$I_{H^1_{per} \to L^2} \colon H^1_{per}(0, 2\pi) \to L^2(0, 2\pi)$$

is compact (see Problem 6.15).

Example 6.16. From Theorem 6.8, the identity operator $I : H \to H$ is compact if and only if $\dim H < \infty$. Also, any bounded operator with finite dimensional range is compact.

Example 6.17. Let $Q = (0, 1) \times (0, 1)$ and $g \in C(\overline{Q})$. Consider the integral operator

$$Tv(x) = \int_0^1 g(x, y) v(y) \, dy. \tag{6.50}$$

We want to show that T is compact from $L^2(0,1)$ into $L^2(0,1)$. In fact, for every $x \in (0,1)$, Schwarz's inequality gives

$$|Tv(x)| \leq \int_0^1 |g(x,y)v(y)|\, dy \leq \|g(x,\cdot)\|_{L^2(0,1)} \|v\|_{L^2(0,1)}, \qquad (6.51)$$

whence

$$\int_0^1 |Tv(x)|^2\, dx \leq \|g\|_{L^2(Q)}^2 \|v\|_{L^2(0,1)}^2$$

which implies that $Tv \in L^2(0,1)$ and that T is bounded.

To check compactness, we use Proposition 6.7. Let $\{v_k\} \subset L^2(0,1)$ such that $v_k \rightharpoonup 0$, that is

$$\int_0^1 v_k w \to 0, \qquad \text{for every } w \in L^2(0,1). \qquad (6.52)$$

We have to show that $Tv_k \to 0$ in $L^2(0,1)$. Being weakly convergent, $\{v_k\}$ is bounded so that

$$\|v_k\|_{L^2(0,1)} \leq M, \qquad (6.53)$$

for some M and every k. From (6.51) we have

$$|Tv_k(x)| \leq M \|g(x,\cdot)\|_{L^2(0,1)}.$$

Moreover, inserting $w(\cdot) = g(x,\cdot)$ into (6.52), we infer that

$$Tv_k(x) = \int_0^1 g(x,y)v_k(y)\, dy \to 0 \qquad \text{for every } x \in (0,1).$$

From the Dominated Convergence Theorem[14] we infer that $Tv_k \to 0$ in $L^2(0,1)$. Therefore T is compact.

The following proposition is useful.

Proposition 6.7. *Let $L : H_1 \to H_2$ be compact. Then:*

a) $L^* : H_2 \to H_1$ is compact;

b) if $G \in \mathcal{L}(H_2, H_3)$ or $G \in \mathcal{L}(H_0, H_1)$, the operator $G \circ L$ or $L \circ G$ is compact.

Proof. a). We use Proposition 6.7. Let $\{x_k\} \subset H_2$ and $x_k \rightharpoonup 0$. Let us show that $\|L^* x_k\|_{H_1} \to 0$. We have:

$$\|L^* x_k\|_{H_1}^2 = (L^* x_k, L^* x_k)_{H_1} = (x_k, LL^* x_k)_{H_2}.$$

Since $L^* \in \mathcal{L}(H_2, H_1)$, we have

$$L^* x_k \rightharpoonup 0$$

----

[14] Appendix B.

in H_1 and the compactness of L entails $LL^*x_n \to 0$ in H_2. Since $\|x_k\| \le M$, we finally have

$$\|L^*x_k\|^2_{H_1} = (x_k, LL^*x_k)_{H_2} \le M \|LL^*x_k\|^2_{H_2} \to 0.$$

b). We leave it as an exercise. □

6.8 The Fredholm Alternative

6.8.1 Solvability for abstract variational problems

Let us go back to the variational problem

$$a\,(u, v) = \langle F, v \rangle_* \qquad \forall v \in V, \tag{6.54}$$

and suppose that Lax-Milgram Theorem cannot be applied, since, for instance, a is not $V-$coercive. In this situation it may happen that the problem does not have a solution, unless certain compatibility conditions on F are satisfied. A typical example is given by the Neumann problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u = g \text{ on } \partial\Omega. \end{cases}$$

A necessary and sufficient solvability condition is given by.

$$\int_{\Omega} f + \int_{\partial\Omega} g = 0. \tag{6.55}$$

Moreover, if (6.55) holds, there are infinitely many solutions, differing among each other by an additive constant. Condition (6.55) has both a precise physical interpretation in terms of a resultant of forces at equilibrium and a deep mathematical meaning, with roots in Linear Algebra!

Indeed, the results we are going to present are extensions of well known facts concerning the solvability of linear algebraic systems of the form

$$\mathbf{Ax} = \mathbf{b} \tag{6.56}$$

where $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$. The following dichotomy holds: *either* (6.56) *has a unique solution for every* $\mathbf{b}$ *or the homogeneous equation* $\mathbf{Ax} = \mathbf{0}$ *has non trivial solutions.*

More precisely, system (6.56) is solvable if and only if $\mathbf{b}$ belongs to the *column space of* $\mathbf{A}$, which is the orthogonal complement of $\ker(\mathbf{A}^\top)$. If $\mathbf{w}_1, ..., \mathbf{w}_s$ span $\ker(\mathbf{A}^\top)$, this amounts to asking the s compatibility conditions, $0 \le s \le n$,

$$\mathbf{b} \cdot \mathbf{w}_j = \mathbf{0} \qquad j = 1, ..., s.$$

Finally, $\ker(\mathbf{A})$ and $\ker(\mathbf{A}^\top)$ have the same dimension and if $\mathbf{v}_1, ..., \mathbf{v}_s$ span $\ker(\mathbf{A})$, the general solution of (6.56) is given by

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{j=1}^{s} c_j \mathbf{v}_j$$

where $\overline{\mathbf{x}}$ is a particular solution of (6.56) and $c, ..., c_s$ are arbitrary constants.

The extension to infinite-dimensional spaces requires some care. In particular, in order to state an analogous dichotomy theorem for the variational problem (6.54), we need to clarify the general setting, to avoid confusion.

The problem involves two Hilbert spaces: V, the space where we seek the solution, and V^*, which the data F belongs to. Let us introduce a third space H, intermediate between V and V^*. In boundary value problems, usually $H = L^2(\Omega)$, with Ω bounded domain in $\mathbb{R}^n$, while V is a Sobolev space. In practice, we often meet a pair of Hilbert spaces V, H with the following properties:

1. $V \hookrightarrow H$, i.e. V is *continuously embedded in H*. Recall that this simply means that the identity operator $I_{V \to H}$, from V into H, is continuous or, equivalently that there exists C such that

$$\|u\|_H \leq C \|u\|_V \qquad \forall u \in V. \tag{6.57}$$

2. V is *dense in H*.

Using Riesz's Theorem, we may identify H with H^*. Also, we may *continuously embed H into V^**, so that any element in H can be thought as an element of V^*. It is enough to observe that, for any fixed $u \in H$, the functional T_u defined by

$$\langle T_u, v \rangle_* = (u, v)_H \qquad v \in V, \tag{6.58}$$

is continuous in V. In fact, Schwarz's inequality and (6.57) give

$$|(u, v)_H| \leq \|u\|_H \|v\|_H \leq C \|u\|_H \|v\|_V. \tag{6.59}$$

Then, we have a continuous map $u \to T_u$, from H into V^*, with $\|T_u\|_{V^*} \leq C \|u\|_H$. If $T_u = 0$ then

$$(u, v)_H = 0 \quad \forall v \in V$$

which forces $u = 0$, by the density of V in H.

Thus, the map $u \longmapsto T_u$ is one to one and defines a continuous embedding $I_{H \to V^*}$. This allows the *identification* of u with an element of V^*, which means that, instead of (6.58), we can write

$$\langle u, v \rangle_* = (u, v)_H \qquad \forall v \in V,$$

regarding u on the left as an element of V^* and on the right as an element of H.

Finally, it can be shown that V and H *are dense in V^**. Thus, we have

$$V \hookrightarrow H \hookrightarrow V^*$$

with *dense embeddings*. We call (V, H, V^*) a **Hilbert triplet**.

This is the right setting. We use the symbols

$$(\cdot, \cdot) = (\cdot, \cdot)_H, \qquad \|\cdot\| = \|\cdot\|_H$$

to denote inner product and norm in H, respectively, while $\langle \cdot, \cdot \rangle_* = {}_{V^*}\langle \cdot, \cdot \rangle_V$ is reserved for the duality between V^* and V.

To state the main result we need to introduce weakly coercive forms and their adjoints.

Definition 6.10. *We say that the bilinear form $a(u, v)$ is weakly coercive with respect to the pair (V, H) if there exist $\lambda_0 \in \mathbb{R}$ and $\alpha > 0$ such that*

$$a(v, v) + \lambda_0 \|v\|^2 \geq \alpha \|v\|_V^2 \qquad \forall v \in V.$$

The *adjoint* form a^* of a is given by

$$a^*(u, v) = a(v, u),$$

obtained by interchanging the arguments in the analytical expression of a. In the applications to boundary value problems, a^* is associated with the so called *formal adjoint* of a differential operator (see subsection 8.5.1).

We shall denote by $\mathcal{N}(a)$ and $\mathcal{N}(a^*)$, the set of solutions u and w, respectively, of the variational problems

$$a(u, v) = 0, \quad \forall v \in V \qquad \text{and} \qquad a^*(w, v) = 0, \quad \forall v \in V.$$

Observe that $\mathcal{N}(a)$ and $\mathcal{N}(a^*)$ are both subspaces of V, playing the role of *kernels* for a and a^*

Theorem 6.13. *Let (V, H, V^*) be a Hilbert triplet, with V compactly embedded in H. Let $F \in V^*$ and a be a bilinear form in V, continuous and weakly coercive with respect to (V, H). Then:*

a) Either equation

$$a(u, v) = \langle F, v \rangle_* \qquad \forall v \in V \tag{6.60}$$

has a unique solution $\overline{u}$ and

$$\|\overline{u}\| \leq C \|F\|_{V^*} \tag{6.61}$$

b) or

$$\dim \mathcal{N}(a) = \dim \mathcal{N}(a_*) = d < \infty.$$

and (6.60) is solvable if and only if $\langle F, w \rangle_* = 0$ for every $w \in \mathcal{N}(a^*)$.

The proof of Theorem 6.12 relies on a more general result, known as Fredholm's Alternative, presented in the next section.

Some comments are in order. The following dichotomy holds: either (6.60) has a unique solution for every $F \in V^*$ or the homogeneous equation $a(u, v) = 0$ has non trivial solutions. The same conclusions hold for the adjoint equation

$$a^*(u, v) = \langle F, v \rangle_*, \qquad \forall v \in V.$$

If $w_1, w_2, ..., w_d$ span $\mathcal{N}(a^*)$, (6.60) is solvable if and only if the d *compatibility conditions*

$$\langle F, w_j \rangle_* = 0, \qquad j = 1, ..., d$$

hold. In this case, equation (6.60) has infinitely many solutions given by

$$u = \overline{u} + \sum_{j=1}^{d} c_j z_j$$

where $\overline{u}$ is a particular solution of (6.60), $z_1, ..., z_d$ span $\mathcal{N}(a)$ and $c_1, ..., c_d$ are arbitrary constants.

We shall apply Theorem 6.12 to boundary value problems in Chapter 8. Here is however a preliminary example.

Example 6.18. Let $V = H^1_{per}(0, 2\pi)$, $H = L^2(0, 2\pi)$ and assume that $w = w(t)$ is a *positive,* continuous function in $[0, 2\pi]$. We know (Example 6.27) that V is compactly embedded in H. Moreover, it can be shown that V is dense in H.

Thus (V, H, V^*) is a Hilbert triplet.

Given $f \in H$, consider the variational problem

$$\int_0^{2\pi} u' v' w \, dt = \int_0^{2\pi} f v \, dt, \qquad \forall v \in V. \tag{6.62}$$

The bilinear form $a(u, v) = \int_0^{2\pi} u' v' w \, dt$ is continuous in V but it is not $V-$coercive. In fact

$$|a(u, v)| \leq w_{\max} \|u'\|_0 \|v'\|_0 \leq w_{\max} \|u\|_1 \|v\|_1,$$

but $a(u, u) = 0$ if u is constant. However it is weakly coercive with respect to (V, H), since

$$a(u, u) + \|u\|_0^2 = \int_0^{2\pi} (u')^2 w \, dt + \int_0^{2\pi} u^2 \, dt \geq \min\{w_{\min}, 1\} \|u\|_{1,2}^2.$$

Moreover,

$$\left| \int_0^{2\pi} f v \, dt \right| \leq \|f\|_0 \|v\|_0 \leq \|f\|_0 \|v\|_{1,2}$$

hence the functional $F : v \mapsto \int_0^{2\pi} f v \, dt$ defines an element of V^*.

We are under the hypotheses of Theorem 6.12. The bilinear form is symmetric, so that $\mathcal{N}(a) = \mathcal{N}(a_*)$. The solutions of the homogeneous equation

$$a(u, v) = \int_0^{2\pi} u' v' w \, dt = 0, \qquad \forall v \in V \tag{6.63}$$

are the constant functions. In fact, letting $v = u$ in (6.63) we obtain

$$\int_0^{2\pi} (u')^2 w \, dt = 0$$

which forces $u(t) \equiv c$, constant, since $w > 0$. Then, $\dim \mathcal{N}(a) = 1$. Thus, from Theorem 6.11 we can draw the following conclusions: equation (6.62) is solvable if and only if

$$\langle F, 1 \rangle_* = \int_0^{2\pi} f \, dt = 0.$$

Moreover, in this case, (6.62) has infinitely many solutions of the form $u = \overline{u} + c$.

The variational problem has a simple interpretation as a boundary value problem. By an integration by parts, recalling that $v(0) = v(2\pi)$, we may rewrite (6.62) as

$$\int_0^{2\pi} [(-wu')' - f]v \, dt + v(0)[w(2\pi)u'(2\pi) - w(0)u'(0)] = 0, \qquad \forall v \in V.$$

Choosing v vanishing at 0 we are left with

$$\int_0^{2\pi} [-(wu')' - f]v \, dt = 0, \qquad \forall v \in V, \, v(0) = v(2\pi) = 0.$$

which forces

$$(u'w)' = -f.$$

Then

$$v(0)[w(2\pi)u'(2\pi) - w(0)u'(0)] = 0, \qquad \forall v \in V$$

which, in turn, forces

$$w(2\pi)u'(2\pi) = w(0)u'(0).$$

Thus, problem (6.62) constitutes the variational formulation of the following boundary value problem:

$$\begin{cases} (wu')' = -f & \text{in } (0, 2\pi) \\ u(0) = u(2\pi) \\ w(2\pi)u'(2\pi) = w(0)u'(0). \end{cases}$$

It is important to point out that the periodicity condition $u(0) = u(2\pi)$ is forced by the choice of the space V while the Neuman type periodicity condition is encoded in the variational equation (6.62).

6.8.2 Fredholm's Alternative

We introduce some terminology. Let V_1, V_2 Hilbert spaces and $\Phi : V_1 \to V_2$. We say that Φ is a *Fredholm operator* if $\mathcal{N}(\Phi)$ and $\mathcal{R}(\Phi)^\perp$ *have finite dimension*. The *index of* Φ is the integer

$$\text{ind}(\Phi) = \dim \mathcal{N}(\Phi) - \dim \mathcal{R}(\Phi)^\perp = \dim \mathcal{N}(\Phi) - \dim \mathcal{N}(\Phi^*).$$

We have[15]:

---

[15] For the proof, see e.g. *Brezis, 1983*.

Theorem 6.14. (Fredholm's Alternative). *Let V be a Hilbert space and $K \in \mathcal{L}(V)$ be a compact operator. Then*

$$\Phi = I - K$$

is a Fredholm operator with zero index. Moreover $\Phi^* = I - K^*$,

$$\mathcal{R}(\Phi) = \mathcal{N}(\Phi^*)^\perp \tag{6.64}$$

and

$$\mathcal{N}(\Phi) = \{0\} \iff \mathcal{R}(\Phi) = V. \tag{6.65}$$

The last formula shows that Φ is one-to-one if and only if it is onto. In other words, uniqueness for the equation

$$x - Kx = f \tag{6.66}$$

is equivalent to existence for every $f \in V$ and viceversa. The same thing holds for the adjoint $\Phi^* = I - K^*$ and the associated equation

$$y - K^*y = g.$$

Let $d = \dim \mathcal{R}(\Phi)^\perp = \dim \mathcal{N}(\Phi^*) > 0$. Then, (6.64) says that equation (6.66) is solvable if and only if $f \perp \mathcal{N}(\Phi^*)$, that is, if and only if $(f, y) = 0$ for every solution y of

$$y - K^*y = 0. \tag{6.67}$$

If $y_1, y_2, ..., y_d$ span $\mathcal{N}(\Phi^*)$, this amounts to asking the d *compatibility relations*

$$(f, y_j) = 0, \qquad j = 1, ..., d$$

as necessary and sufficient conditions for the solvability of (6.66).

Remark 6.14. Clearly, Theorem 6.13 holds for operators $K - \lambda I$ with $\lambda \neq 0$. The case $\lambda = 0$ cannot be included. Trivially, for the operator $K = 0$ (which is compact), we have $\mathcal{N}(K) = V$, hence, if $\dim V = \infty$, Theorem 6.13 does not hold. A more significant example is the one-dimensional range operator

$$Kx = L(x)x_0$$

where $L \in \mathcal{L}(V)$ and x_0 is fixed in V. Assume $\dim V = \infty$. From Riesz's Theorem, there exists $z \in V$ such that $Lx = (z, x)$ for every $x \in V$. Thus, $\mathcal{N}(K)$ is given by the subspace of the elements in V orthogonal to z, which has infinitely many dimensions.

- *Proof of Theorem* 6.12 (sketch). The strategy is to write equation

$$a(u, v) = \langle F, v \rangle_* \tag{6.68}$$

in the form

$$(I_V - K)u = g.$$

where I_V is the identity operator in V and $K : V \to V$ is compact.

Let $J : V \to V^*$ the embedding of V into V^*. Recall that J is the composition of the embeddings $I_{V \to H}$ and $I_{H \to V^*}$. Since $I_{V \to H}$ is compact and $I_{H \to V^*}$ is continuous, we infer from Proposition 6.8 that J **is compact**. We write (6.68) in the form

$$a_{\lambda_0}(u, v) \equiv a(u, v) + \lambda_0 (u, v)_H = \langle \lambda_0 Ju + F, v \rangle_*$$

where $\lambda_0 > 0$ is such that $a_{\lambda_0}(u, v)$ is coercive. Since, for fixed $u \in V$, the linear map

$$v \mapsto a_{\lambda_0}(u, v)$$

is continuous in V, there exists $L \in \mathcal{L}(V, V^*)$ such that

$$\langle Lu, v \rangle_* = a_{\lambda_0}(u, v) \qquad \forall u, v \in V.$$

Thus, equation $a(u, v) = \langle F, v \rangle_*$ is equivalent to

$$\langle Lu, v \rangle_* = \langle \lambda_0 Ju + F, v \rangle_* \qquad \forall v \in V$$

and therefore to

$$Lu = \lambda_0 Ju + F. \tag{6.69}$$

Since a_{λ_0} is $V-$coercive, from the Lax-Milgram Theorem, the operator L is an isomorphism between V and V^* and (6.69) can be written in the form

$$u - \lambda_0 L^{-1} Ju = L^{-1} F.$$

Letting $g = L^{-1}F \in V$ and $K = \lambda_0 L^{-1}J$, (6.69) becomes

$$(I_V - K)u = g.$$

where $K : V \to V$.

Since J is compact and L^{-1} is continuous, K is compact (Proposition 6.8). Applying the Fredholm Alternative Theorem and rephrasing the conclusions in terms of bilinear forms we conclude the proof[16]. $\square$

6.9 Spectral Theory for Symmetric Bilinear Forms

6.9.1 Spectrum of a matrix

Let $\mathbf{A}$ be an $n \times n$ matrix and $\lambda \in \mathbb{C}$. Then, either the equation

$$\mathbf{Ax} - \lambda \mathbf{x} = \mathbf{b}$$

---

[16] We omit the rather long and technical details.

has a unique solution for every $\mathbf{b}$ or there exists $\mathbf{u} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{u}.$$

In the last case we say that $\lambda, \mathbf{u}$ constitutes an *eigenvalue-eigenvector pair*. The set of eigenvalues of $\mathbf{A}$ is called *spectrum of* $\mathbf{A}$, denoted by $\sigma_P(\mathbf{A})$. If $\lambda \notin \sigma_P(\mathbf{A})$ the *resolvent matrix* $(\mathbf{A} - \lambda \mathbf{I})^{-1}$ is well defined. The set

$$\rho(\mathbf{A}) = \mathbb{C} \backslash \sigma_P(\mathbf{A})$$

is called the *resolvent of* $\mathbf{A}$. If $\lambda \in \sigma_P(\mathbf{A})$, the kernel $\mathcal{N}(\mathbf{A} - \lambda \mathbf{I})$ is the subspace spanned by the eigenvectors corresponding to λ and it is called the *eigenspace* of λ. Note that $\sigma_P(\mathbf{A}) = \sigma_P(\mathbf{A}^\top)$.

The *symmetric* matrices are particularly important: all the eigenvalues $\lambda_1, ... \lambda_n$ are real (possibly of multiplicity greater than 1) and there exists in $\mathbb{R}^n$ an orthonormal basis of eigenvectors $\mathbf{u}_1, ..., \mathbf{u}_n$.

We are going to extend these concepts in the Hilbert space setting. A motivation is .... the method of separation of variables.

6.9.2 Separation of variables revisited

Using the method of separation of variables, in the first chapters we have constructed solutions of boundary value problems by superposition of special solutions. However, explicit computations can be performed only when the geometry of the relevant domain is quite particular. What may we say in general? Let us consider an example from diffusion.

Suppose we have to solve the problem

$$\begin{cases} u_t = \Delta u & (x, y) \in \Omega,\, t > 0 \\ u(x, y, 0) = g(x, y) & (x, y) \in \Omega \\ u(x, y, t) = 0 & (x, y) \in \partial\Omega,\, t > 0 \end{cases}$$

where Ω is a bounded bi-dimensional domain. Let us look for solutions of the form

$$u(x, y, t) = v(x, y)\, w(t).$$

Substituting into the differential equation, with some elementary manipulations, we obtain

$$\frac{w'(t)}{w(t)} = \frac{\Delta v(x, y)}{v(x, y)} = -\lambda,$$

where λ is a constant, which leads to the two problems

$$w' + \lambda w = 0 \qquad t > 0 \tag{6.70}$$

and

$$\begin{cases} -\Delta v = \lambda v & \text{in } \Omega \\ v = 0 & \text{on } \partial\Omega. \end{cases} \tag{6.71}$$

A number λ such that there exists a non trivial solution v of (6.71) is called a *Dirichlet eigenvalue of the operator* $-\Delta$ *in* Ω and v is a corresponding *eigenfunction*. Now, the original problem can be solved if the following two properties hold:

a) There exists a sequence of (real) eigenvalues λ_k with corresponding eigenvectors u_k. Solving (6.70) for $\lambda = \lambda_k$ yields

$$w_k(t) = ce^{-\lambda_k t} \qquad c \in \mathbb{R}.$$

b) The initial data g can be expanded is series of eigenfunctions:

$$u(x, y) = \sum g_k u_k(x, y).$$

Then, the solution is given by

$$u(x, y, t) = \sum g_k e^{-\lambda_k t} u_k(x, y)$$

where the series converges in some suitable sense.

Condition **b**) requires that the set of Dirichlet eigenfunctions of $-\Delta$ constitutes a basis in the space of initial data. This leads to the problem of determining the *spectrum* of a linear operator in a Hilbert space and, in particular, of self-adjoint compact operators. Indeed, it turns out that the solution map of a symmetric variational boundary value problem is often a self-adjoint compact operator. We will go back to the above problem in subsection 8.4.3.

6.9.3 Spectrum of a compact self-adjoint operator

We define *resolvent and spectrum* for a bounded linear operator. Although the natural setting is the complex field $\mathbb{C}$, we limit ourselves to $\mathbb{R}$, mainly for simplicity but also because this is the interesting case for us.

Definition 6.11. *Let H be a Hilbert space, $L \in \mathcal{L}(H)$, and I the identity in H.*

a) The resolvent set $\rho(L)$ of L is the set of real numbers λ such that $L - \lambda I$ is one-to-one and onto:

$$\rho(L) = \{\lambda \in \mathbb{R} : L - \lambda I \text{ is one-to-one and onto}\}.$$

b) The (real) spectrum $\sigma(L)$ of L is

$$\sigma(L) = \mathbb{R} \backslash \rho(L).$$

Remark 6.15. If $\lambda \in \rho(L)$, the *resolvent* $(L - \lambda I)^{-1}$ is bounded[17].

---

[17] It is a consequence of the *Closed Graph Theorem: If the graph of a linear operator $A : H_1 \to H_2$ is closed in $H_1 \times H_2$ then A is bounded.*

If H has finite dimension, any linear operator is represented by a matrix, so that its spectrum is given by the set of its eigenvalues. In infinitely many dimensions the spectrum may be divided in three subsets. In fact, if $\lambda \in \sigma(L)$, different things can go wrong with $(L - \lambda I)^{-1}$.

First of all, it may happen that $L - \lambda I$ is not one-to-one so that $(L - \lambda I)^{-1}$ does not even exist. This means that $\mathcal{N}(L - \lambda I) \neq \emptyset$, i.e. that the equation

$$Lx = \lambda x \tag{6.72}$$

has non trivial solutions. Then, we say that λ is an *eigenvalue of* L and that the non zero solutions of (6.72) are the *eigenvectors* corresponding to λ. The linear space spanned by these eigenvectors is called the *eigenspace of* λ and denoted by $\mathcal{N}(L - \lambda I)$.

Definition 6.12. *The set* $\sigma_P(L)$ *of the eigenvalues of L is called the point spectrum of L.*

Other things can occur. $L - \lambda I$ is one-to-one, $\mathcal{R}(L - \lambda I)$ is dense in H, but $(L - \lambda I)^{-1}$ is unbounded. Then, we say that λ belongs to the *continuous spectrum of* L, denoted by $\sigma_C(L)$.

Finally, $L - \lambda I$ is one-to-one but $\mathcal{R}(L - \lambda I)$ is not dense in H. This defines the *residual spectrum of* L.

Example 6.19. Let $H = l^2$ and $L : l^2 \to l^2$ be the *shift* operator which maps $\mathbf{x} = \{x_1, x_2, ...\} \in l^2$ into $\mathbf{y} = \{0, x_1, x_2, ...\}$. We have

$$(L - \lambda I) x = \{-\lambda x_1, x_1 - \lambda x_2, x_2 - \lambda x_3, ...\}.$$

If $\lambda \neq 0$, then $\lambda \in \rho(L)$. In fact for every $\mathbf{z} = \{z_1, z_2, ...\} \in l^2$,

$$(L - \lambda I)^{-1} \mathbf{z} = \left\{ -\frac{z_1}{\lambda}, -\frac{z_2}{\lambda} + \frac{z_1}{\lambda^2}, ... \right\}.$$

Since $\mathcal{R}(L)$ contains only sequences whose first element is zero, $\mathcal{R}(L)$ is *not dense* in l^2, therefore $0 \in \sigma_R(L) = \sigma(L)$.

We are mainly interested in the spectrum of a *compact self-adjoint* operator. The following theorem is fundamental[18].

Theorem 6.15. *Let K be a compact, self-adjoint operator on a separable Hilbert space H. Then:*

a) $0 \in \sigma(K)$ *and* $\sigma(K) \setminus \{0\} = \sigma_P(K) \setminus \{0\}$.

b) H has an orthonormal basis $\{u_m\}$ *consisting of eigenvectors for K.*

c) If dim $H = \infty$, *the corresponding eigenvalues different from zero* $\{\lambda_m\}$ *can be arranged in a decreasing sequence* $|\lambda_1| \geq |\lambda_2| \geq \cdots$, *with* $\lambda_m \to 0$, *as* $m \to \infty$.

---

[18] For the proof, see *Brezis, 1983.*

Thus, the spectrum of a compact self-adjoint operator contains always $\lambda = 0$, which is not necessarily an eigenvalue. The other elements in $\sigma(L)$ are eigenvalues, arranged in a sequence converging to zero if H is infinite dimensional.

If $\lambda \neq 0$ is an eigenvalue, Fredholm's Alternative applies to $L - \lambda I$, so that, in particular, the eigenspace $\mathcal{N}(L - \lambda I)$ has finite dimension.

A consequence of Theorem 6.15 is the *spectral decomposition formula* for K. If $x \in H$ and $\{u_m\}_{m \geq 1}$ is an orthonormal set of eigenvectors corresponding to all non-zero eigenvalues $\{\lambda_m\}_{m \geq 1}$, we can describe the action of K as follows:

$$Kx = \sum_{m \geq 1}(Kx, u_m)u_m = \sum_{m \geq 1}\lambda_m(x, u_m)u_m, \quad \forall x \in H. \tag{6.73}$$

6.9.4 Application to abstract variational problems

We now apply theorems 6.13 and 6.14 to our abstract variational problems. The setting is the same of theorem 6.12, given by a Hilbert triplet (V, H, V^*), with compact embedding of V into H. We assume that H is also separable. Let a be a bilinear form in V, continuous and weakly coercive; in particular:

$$a_{\lambda_0}(u, v) \equiv a(v, v) + \lambda_0 \|v\|^2 \geq \alpha \|v\|_V^2 \qquad \forall v \in V$$

The notion of *resolvent and spectrum* can be easily defined. Consider the problem

$$a(u, v) = \lambda(u, v) + \langle F, v \rangle_* \qquad \forall v \in V. \tag{6.74}$$

The *resolvent* $\rho(a)$ is the set of real numbers λ such that (6.74) has a unique solution $u(F) \in V$ for every $F \in V^*$ and the solution map

$$S_\lambda : F \longmapsto u(F)$$

is an isomorphism between V^* and V.

The (real) *spectrum* is $\sigma(a) = \mathbb{R} \setminus \rho(a)$, while the *point spectrum* $\sigma_P(a)$ is the subset of the spectrum given by the *eigenvalues,* i.e. the numbers λ such that the homogeneous problem

$$a(u, v) = \lambda(u, v) \qquad \forall v \in V \tag{6.75}$$

has non-trivial solutions (*eigenfunctions*). We call *eigenspace of* λ the space spanned by the corresponding eigenfunctions and we denote it by $\mathcal{N}(a, \lambda)$.

The following theorem is a consequence of the Fredholm Alternative and Theorem 7.4. and it is based on the following relation between $\sigma_P(S_{\lambda_0})$ and $\sigma_P(a_{\lambda_0})$. Note that $0 \notin \sigma(S_{\lambda_0})$ and that $\sigma(a_{\lambda_0}) \subset (0, +\infty)$.

Let $\mu \in \sigma_P(S_{\lambda_0})$ and f be a corresponding eigenvector, that is,

$$S_{\lambda_0} f = \mu f.$$

Thus, necessarily $f \in V$ and

$$a_{\lambda_0}\left(S_{\lambda_0}f, v\right) = \mu a_{\lambda_0}\left(f, v\right) = (f, v)$$

or

$$a_{\lambda_0}\left(f, v\right) = \frac{1}{\mu}\left(f, v\right)$$

for all $v \in V$. Therefore $\lambda = 1/\mu$ is an eigenvalue of a_{λ_0}, with the same eigenspace. As a consequence

$$\mathcal{N}\left(a_{\lambda_0}\right) = \mathcal{N}\left(S_{\lambda_0}\right) \subset V.$$

Moreover, since the eigenvalues of a_{λ_0} are all positive, if follows that $\mu > 0$ as well.

Theorem 6.16. *Let (V, H, V^*) be a Hilbert triplet with H separable and V compactly embedded in H. Let $F \in V^*$ and a be a symmetric bilinear form in V, continuous and weakly coercive. We have:*

(a) $\sigma\left(a\right) = \sigma_P\left(a\right) \subset (-\lambda_0, +\infty)$. Moreover, if the sequence of eigenvalues $\{\lambda_m\}$ is infinite, then $\lambda_m \to +\infty$.

(b) If u, v are eigenfunctions corresponding to different eigenvalues, then $a\left(u, v\right) = (u, v) = 0$. Moreover, H has an orthonormal basis of eigenvectors u_m.

(c) $\{u_m/\sqrt{\lambda_m + \lambda_0}\}$ constitutes an orthonormal basis in V, with respect to the scalar product

$$\left(\left(u, v\right)\right) = a\left(u, v\right) + \lambda_0\left(u, v\right). \tag{6.76}$$

Proof. By hypothesis, S_{λ_0} is an isomorphism between V^* and V. In particular, it is well defined as a map from H into $V \subset H$. Since the embedding of V in H is compact, then S_{λ_0} is compact as an operator from H into H. Also, by the symmetry of a, S_{λ_0} is selfadjoint, that is

$$\left(S_{\lambda_0}f, g\right) = \left(f, S_{\lambda_0}g\right) \qquad \text{for all } f, g \in H.$$

In fact, let $u = S_{\lambda_0}f$ and $w = S_{\lambda_0}g$. Then, for every $v \in V$,

$$a_{\lambda_0}\left(u, v\right) = (f, v) \quad \text{and} \quad a_{\lambda_0}\left(w, v\right) = (g, v).$$

In particular,

$$a_{\lambda_0}\left(u, w\right) = (f, w) \quad \text{and} \quad a_{\lambda_0}\left(w, u\right) = (g, u)$$

so that, since $a_{\lambda_0}\left(u, w\right) = a_{\lambda_0}\left(w, u\right)$ and $(g, u) = (u, g)$, we can write

$$\left(S_{\lambda_0}f, g\right) = (u, g) = (f, w) = \left(f, S_{\lambda_0}g\right).$$

Since $0 \notin \sigma\left(S_{\lambda_0}\right)$, from Theorem 6.15 it follows that $\sigma\left(S_{\lambda_0}\right) = \sigma_P\left(S_{\lambda_0}\right)$ and the eigenvalues form a sequence $\{\mu_m\}$ with $\mu_m \downarrow 0$. Using Theorem 6.15 and the relation between $\sigma_P\left(S_{\lambda_0}\right)$ and $\sigma_P\left(a_{\lambda_0}\right)$, (a) and (b) follow easily.

Finally if $\{u_m\}$ is an orthonormal basis of eigenvectors for a in H, then[19]

$$a\left(u_m, u_k\right) = \lambda_m\left(u_m, u_k\right) = \lambda_m\delta_{mk}$$

so that

$$((u_m, u_k)) \equiv a_{\lambda_0}\left(u_m, u_k\right) = \left(\lambda_m + \lambda_0\right)\delta_{mk}$$

which easily gives (c). $\square$

Problems

6.1. *Heisenberg Uncertainty Principle.* Let $\psi \in C^1\left(\mathbb{R}\right)$ such that $x\left[\psi\left(x\right)\right]^2 \to 0$ as $|x| \to \infty$ and $\int_{\mathbb{R}}\left[\psi\left(x\right)\right]^2 dx = 1$. Show that

$$1 \le 2 \int_{\mathbb{R}} x^2\left|\psi\left(x\right)\right|^2 dx \int_{\mathbb{R}} \left|\psi'\left(x\right)\right|^2 dx.$$

(If ψ is a Schrödinger wave function, the first factor in the right hand side measures the spread of the density of a particle, while the second one measures the spread of its momentum).

6.2. Let H be a Hilbert space and $a\left(u, v\right)$ be a symmetric and non negative bilinear form in H:

$$a\left(u, v\right) = a\left(v, u\right) \quad \text{and} \quad a\left(u, v\right) \ge 0 \quad \forall u, v \in H.$$

Show that

$$\left|a\left(u, v\right)\right| \le \sqrt{a\left(u, u\right)}\sqrt{a\left(v, v\right)}.$$

[*Hint.* Mimic the proof of Schwarz's inequality].

6.3. Show the completeness of l^2.

[*Hint.* Take a Cauchy sequence $\{\mathbf{x}^k\}$ where $\mathbf{x}^k = \{x_m^k\}$. In particular, $\left|x_m^k - x_m^h\right| \to 0$ as $h, k \to \infty$ and therefore $x_m^h \to x_m$ for every m. Define $\mathbf{x} = \{x_m\}$ and show that $\mathbf{x}^k \to \mathbf{x}$ in l^2].

6.4. Let H be a Hilbert space and V a closed subspace of H. Show that $u = P_V x$ if and only if

$$\begin{cases} \textbf{1.} \ u \in V \\ \textbf{2.} \ (x - u, v) = 0, \forall v \in V. \end{cases}$$

6.5. Let $f \in L^2\left(-1, 1\right)$. Find the polynomial of degree $\le n$ that gives the best approximation of f in the least squares sense, that is, the polynomial p that minimizes

$$\int_{-1}^{1}(f - q)^2$$

among all polynomials q with degree $\le n$.

---

[19] δ_{mk} is Kronecker symbol.

[*Answer:* $p(x) = a_0 L_0(x) + a_1 L_1(x) + ... + a_n L_n(x)$, where L_n is the $n - th$ Legendre polynomials and $a_j = (n + 1/2)(f, L_n)_{L^2(-1,1)}$].

6.6. *Hermite's equation and the quantum mechanics harmonic oscillator.* Consider the equation

$$w'' + (2\lambda + 1 - x^2) w = 0 \qquad x \in \mathbb{R} \qquad (6.77)$$

with $w(x) \to 0$ as $x \to \pm\infty$.

a) Show that the change of variables $z = we^{x^2/2}$ transforms (6.77) into Hermite's equation for z :

$$z'' - 2xz' + 2\lambda z = 0$$

with $e^{-x^2/2} z(x) \to 0$ as $x \to \pm\infty$.

b) Consider the Schrödinger wave equation for the harmonic oscillator

$$\psi'' + \frac{8\pi^2 m}{h^2} \left(E - 2\pi^2 m\nu^2 x^2 \right) \psi = 0 \qquad x \in \mathbb{R}$$

where m is the mass of the particle, E is the total energy, h is the Plank constant and ν is the vibrational frequency. The physically admissible solutions are those satisfying the following conditions:

$$\psi \to 0 \ \text{ as } x \to \pm\infty \qquad \text{and} \qquad \|\psi\|_{L^2(\mathbb{R})} = 1.$$

Show that there is a solution if and only if

$$E = h\nu \left(n + \frac{1}{2} \right) \qquad n = 0, 1, 2....$$

and, for each n, the corresponding solution is given by

$$\psi_n(x) = k_n H_n \left(2\pi \sqrt{\nu m/h} x \right) \exp \left(-\frac{2\pi^2 \nu m}{h} x^2 \right)$$

where $k_n = \left(\dfrac{4\pi\nu m}{2^{2n} (n!)^2 h} \right)^{1/2}$ and H_n is the $n - th$ Hermite polynomial.

6.7. Using separation of variables, solve the following steady state diffusion problem in three dimensions (r, θ, φ spherical coordinates, $0 \le \theta \le 2\pi, 0 \le \varphi \le \pi$):

$$\begin{cases} \Delta u = 0 & r < 1, 0 < \varphi < \pi \\ u(1, \varphi) = g(\varphi) & 0 \le \varphi \le \pi. \end{cases}$$

[*Answer:*

$$u(r, \varphi) = \sum_{n=0}^{\infty} a_n r^n L_n(\cos \varphi),$$

where L_n is the $n - th$ Legendre polynomial and

$$a_n = \frac{2n+1}{2} \int_{-1}^{1} g\left(\cos^{-1} x\right) L_n\left(x\right) dx.$$

At a certain point, the change of variable $x = \cos\varphi$ is required].

6.8. The vertical displacement u of a circular membrane of radius a satisfies the bidimensional wave equation $u_{tt} = \Delta u$, with boundary condition $u\left(a, \theta, t\right) = 0$. Supposing the membrane initially at rest, write a formal solution of the problem. [*Hint:*

$$u\left(r, \theta, t\right) = \sum_{p,j=0}^{\infty} J_p\left(\alpha_{pj} r\right) \left\{A_{pj} \cos p\theta + B_{pj} \sin p\theta\right\} \cos(\sqrt{\alpha_{pj}} t)$$

where the coefficients A_{pj} and B_{pj} are determined by the expansion of the initial condition $u\left(r, \theta, 0\right) = g\left(r, \theta\right)$].

6.9. In calculus, we say that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}_0$ if there exists a linear mapping $L : \mathbb{R}^n \to \mathbb{R}$ such that

$$f\left(\mathbf{x}_0 + \mathbf{h}\right) - f\left(\mathbf{x}_0\right) = L\mathbf{h} + o\left(\|\mathbf{h}\|\right) \qquad \text{as } \mathbf{h} \to \mathbf{0}.$$

Determine the Riesz elements associated with L, with respect to the inner products:

a) $\left(\mathbf{x}, \mathbf{y}\right) = \mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{n} x_j y_j,$ b) $\left(\mathbf{x}, \mathbf{y}\right)_{\mathbf{A}} = \mathbf{A}\mathbf{x} \cdot \mathbf{y} = \sum_{i,j=1}^{n} a_{ij} x_i y_j,$

where $\mathbf{A} = \left(a_{ij}\right)$ is a *positive and symmetric matrix* (see Example 6.3).

6.10. Prove Proposition 6.5.
[*Hint.* First show that
$$\|L^*\|_{\mathcal{L}(H_2, H_1)} \leq \|L\|_{\mathcal{L}(H_1, H_2)}$$
and then that $L^{**} = L$. Reverse the role of L and L^* to show that $\|L^*\|_{\mathcal{L}(H_2, H_1)} \geq \|L\|_{\mathcal{L}(H_1, H_2)}$].

6.11. Prove Neças Theorem 6.6.
[*Hint.* Try to follow the same steps in the proof of the Lax-Milgram Theorem].

6.12. Let $E \subset X$, X Banach space. Prove the following facts:

a) If E is compact, then it is closed and bounded.

b) Let $E \subset F$ and F be compact; if E closed then E is compact.

6.13. *Projection on a closed convex set.* Let H be a Hilbert space and $E \subset H$, closed and convex.

a) Show that, for every $x \in H$, there is a unique element $P_E x \in E$ (the *projection of x on E*) such that

$$\|P_E x - x\| = \inf_{v \in E} \|v - x\|. \tag{6.78}$$

b) Show that $x^* = P_E x$ if and only if

$$(x^* - x, v - x^*) \geq 0 \qquad \text{for every } v \in E. \tag{6.79}$$

c) Give a geometrical interpretation of (6.79).

[*Hint. a*) Follow the proof of the Projection Theorem 6.2. *b*) Let $0 \leq t \leq 1$ and define

$$\varphi(t) = \|x^* + t(v - x^*) - x\|^2 \qquad v \in E.$$

Show that $x^* = P_E x$ if and only if $\varphi'(0) \geq 0$. Check that $\varphi'(0) \geq 0$ is equivalent to (6.79)].

6.14. Let H be a Hilbert space and $E \subset H$ be closed and convex. Show that E is *weakly closed.*

[*Hint.* Let $\{x_k\} \subset E$ such that $x_k \rightharpoonup x$. Use (6.79) to show that $P_E x = x$, so that $x \in E$].

6.15. Show that the embedding of $H^1_{per}(0, 2\pi)$ into $L^2(0, 2\pi)$ is compact.

[*Hint.* Let $\{u_k\} \subset H^1_{per}(0, 2\pi)$ with

$$\|u_k\|^2 = \sum_{m \in \mathbb{Z}} (1 + m^2) |\widehat{u_{km}}|^2 < M.$$

Show that, by a diagonal process, it is possible to select indexes k_j such that, for each m, $\widehat{u_{k_j m}}$ converges to some number U_m. Let

$$u(x) = \sum_{m \in \mathbb{Z}} U_m e^{imx}$$

and show that $u_{k_j} \to u$ in $L^2(0, 2\pi)$].

9.16. Let $L : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be defined by $Lv(x) = v(-x)$. Show that $\sigma(L) = \sigma_P(L) = \{1\}$.

6.17. Let V and W be two closed subspaces of a Hilbert space H, with inner product $(\cdot, \cdot)$. Let $x_0 \in H$ and define the following sequence of projections (see Fig. 6.3):

$$x_{2n+1} = P_W(x_{2n}), \quad x_{2n+2} = P_V(x_{2n+1}), \qquad n \geq 0.$$

Prove that:

(a) If $V \cap W = \{0\}$ then $x_n \to 0$.

(b) If $V \cap W \neq \{0\}$, then $x_n \to P_{V \cap W}(x_0)$

by filling in the details in the following steps.

1. Observe that

$$\|x_{n+1}\|^2 = (x_{n+1}, x_n).$$

Computing $\|x_{n+1} - x_n\|^2$, show that $\|x_n\|$ is decreasing (hence $\|x_n\| \downarrow l \geq 0$) and $\|x_{n+1} - x_n\| \to 0$.

2. If $V \cap W = \{0\}$, show that if a subsequence $x_{2n_k} \rightharpoonup x$, then $x_{2n_k+1} \rightharpoonup x$ as well. Deduce that $x = 0$ (so that the entire sequence converges weakly to 0).

3. Show that

$$\|x_n\|^2 = (x_{n+1}, x_{n-2}) = (x_{n+2}, x_{n-3}) = \cdots = (x_{2n-1}, x_0)$$

and deduce that $x_n \to 0$.

4. If $V \cap W \neq \{0\}$, let

$$z_n = x_n - P_{V \cap W}(x_0)$$

and reduce to the case (a).
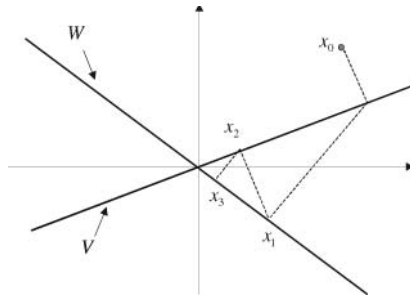
Fig. 6.3. The sequence of projections in problem 6.17 (a)

7

Distributions and Sobolev Spaces

Distributions. Preliminary Ideas – Test Functions and Mollifiers – Distributions – Calculus – Multiplication, Composition, Division, Convolution – Fourier Transform – Sobolev Spaces – Approximations by Smooth Functions and Extensions – Traces – Compactness and Embeddings – Spaces Involving Time

7.1 Distributions. Preliminary Ideas

We have seen the concept of *Dirac measure* arising in connection with the fundamental solutions of the diffusion and the wave equations. Another interesting situation is the following, where the Dirac measure models a mechanical impulse.
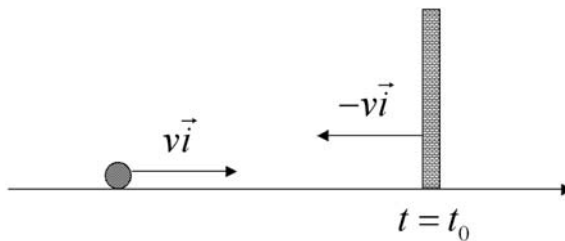
Fig. 7.1. Elastic collision at time $t = t_0$

Consider a mass m moving along the $x-$axis with constant speed $v\mathbf{i}$ (see Fig. 7.1). At time $t = t_0$ an *elastic* collision with a vertical wall occurs. After the collision, the mass moves with opposite speed $-v\mathbf{i}$. If v_2, v_1 denote the scalar speeds at times t_1, t_2, $t_1 < t_2$, by the laws of mechanics we should have

$$m(v_2 - v_1) = \int_{t_1}^{t_2} F(t)\, dt,$$

where F denotes the intensity of the force acting on m. When $t_1 < t_2 < t_0$ or $t_0 < t_1 < t_2$, then $v_2 = v_1 = v$ or $v_2 = v_1 = -v$ and therefore $F = 0$: no force is acting on m before and after the collision. However, if $t_1 < t_0 < t_2$, the left hand side is equal to $2mv \neq 0$. If we insist to model the intensity of the force by a function F, the integral in the right hand side is zero and we obtain a contradiction.

Indeed, in this case, F is a force concentrated at time t_0, of intensity $2mv$, that is

$$F(t) = 2mv \; \delta(t - t_0).$$

In this chapter we see how the Dirac delta is perfectly included in the theory of *distributions or Schwartz generalized functions*. We already mentioned in subsection 2.3.3 that the key idea in this theory is to describe a mathematical object through its action on smooth test functions φ, with compact support. In the case of the Dirac δ, such action is expressed by the formula (see Definition 2.2)

$$\int \delta(x)\,\varphi(x)\,dx = \varphi(0)$$

where, we recall, the integral symbol is purely formal. As we shall shortly see, the appropriate notation is $\langle \delta, \varphi \rangle = \varphi(0)$.

Of course, by a principle of coherence, among the *generalized functions* we should be able to recover the *usual* functions of Analysis. This fact implies that the choice of the test functions cannot be arbitrary. In fact, let $\Omega \subseteq \mathbb{R}^n$ be a domain and take for instance a function $u \in L^2(\Omega)$. A natural way to define the *action* of u on a test φ is

$$\langle u, \varphi \rangle = (u, \varphi)_0 = \int_\Omega u\varphi \; d\mathbf{x}.$$

If we let φ be varying over all $L^2(\Omega)$, we know from the last chapter, that $\langle u, \varphi \rangle$ identifies uniquely u. Indeed, if $v \in L^2(\Omega)$ is such that $\langle u, \varphi \rangle = \langle v, \varphi \rangle$ for every $\varphi \in L^2(\Omega)$, we have

$$0 = \langle u - v, \varphi \rangle = \int_\Omega (u - v)\varphi \; d\mathbf{x} \qquad \forall \varphi \in L^2(\Omega) \qquad (7.1)$$

which forces (why?) $u = v$ a.e. in Ω.

On the other hand, we cannot use L^2-functions as test functions since, for instance, $\langle \delta, \varphi \rangle = \varphi(0)$ does not have any meaning.

We ask: is it possible to reconstruct u from the knowledge of $(u, \varphi)_0$, when φ varies on a set of *nice* functions?

Certainly this is impossible if we use only a restricted set of *test* functions. However, it is possible to recover u from the value of $(u, \varphi)_0$, when φ varies in a **dense** set in $L^2(\Omega)$. In fact, let $(u, \varphi)_0 = (v, \varphi)_0$ for every test function. Given $\psi \in L^2(\Omega)$, there exists a sequence of test functions $\{\varphi_k\}$ such that $\|\varphi_k - \psi\|_0 \to$

0. Then[1],

$$0 = \int_\Omega (u - v)\varphi_k \, d\mathbf{x} \to \int_\Omega (u - v)\psi \, d\mathbf{x}$$

so that (7.1) still holds *for every* $\psi \in L^2(\Omega)$ and $(u, \varphi)_0$ identifies a unique element in $L^2(\Omega)$.

Thus, the set of test functions must be *dense in* $L^2(\Omega)$ if we want L^2-functions to be seen as *distributions*. In the next section we construct an appropriate set of test functions.

However, the main purpose of introducing the Schwartz distributions is not restricted to a mere extension of the notion of function but it relies on the possibility of broadening the domain of *calculus* in a significant way, opening the door to an enormous amount of new applications. Here the key idea is to use integration by parts to carry the derivatives onto the test functions. Actually, this is not a new procedure. For instance, we have used it in subsection 2.3.3, when have interpreted the Dirac delta at $x = 0$ as the derivative of the Heaviside function $\mathcal{H}$, (see formula (2.63) and footnote 24).

Also, the weakening of the notion of solution of conservation laws (subsection 4.4.2) or of the wave equation (subsection 5.4.2) follows more or less the same pattern.

In the first part of this chapter we give the basic concepts of the theory of Schwartz distributions, mainly finalized to the introduction of Sobolev spaces. The basic reference is the book of *L. Schwartz*, 1966, to which we refer for the proofs we do not present here.

7.2 Test Functions and Mollifiers

Recall that, given a continuous function v, defined in a domain $\Omega \subseteq \mathbb{R}^n$, the *support of v* is given by *the closure of the set of points where v is different from zero*:

$$\text{supp}(v) = \Omega \; \cap \; \text{closure of } \{\mathbf{x} \in \Omega : v(\mathbf{x}) \neq 0\}.$$

Actually, the support or, better, the *essential support*, is defined also for measurable functions, not necessarily continuous in Ω. Namely, let Z be the union of the open sets on which $v = 0$ a.e. Then, $\Omega \backslash Z$ is called *the essential support of v* and we use the same symbol $\text{supp}(v)$ to denote it.

We say that v is *compactly supported* in Ω, if $\text{supp}(v)$ is a *compact* subset of Ω.

Definition 7.1. *Denote by* $C_0^\infty(\Omega)$ *the set of functions belonging to* $C^\infty(\Omega)$, *compactly supported in* Ω. *We call* **test functions** *the elements of* $C_0^\infty(\Omega)$.

---

[1] From

$$\left| \int_\Omega (u - v)(\varphi_k - \psi) \, d\mathbf{x} \right| \leq \|u - v\|_0 \|\varphi_k - \psi\|_0.$$

Example 7.1. The reader can easily check that the function given by

$$
\eta\left(\mathbf{x}\right) = \begin{cases} c\exp\left(\frac{1}{|\mathbf{x}|^2-1}\right) & 0 \le |\mathbf{x}| < 1 \\ 0 & |\mathbf{x}| \ge 1 \end{cases} \qquad (c \in \mathbb{R}). \tag{7.2}
$$

belongs to $C_0^\infty\left(\Omega\right)$.

The function (7.2) is a typical and important example of test function. Indeed, we will see below that many other test functions can be generated by convolution with (7.2).

Let us briefly recall the definition and the main properties of the convolution of two functions. Given two functions u and v defined in $\mathbb{R}^n$, the *convolution* $u*v$ of u and v is given by the formula:

$$
(u*v)\left(\mathbf{x}\right) = \int_{\mathbb{R}^n} u\left(\mathbf{x}-\mathbf{y}\right)v\left(\mathbf{y}\right)d\mathbf{y} = \int_{\mathbb{R}^n} u\left(\mathbf{y}\right)v\left(\mathbf{x}-\mathbf{y}\right)d\mathbf{y}.
$$

It can be proved that (*Young's Theorem*): if $u \in L^p(\mathbb{R}^n)$ and $v \in L^q(\mathbb{R}^n)$, $p,q \in [1,\infty]$, then $u*v \in L^r(\mathbb{R}^n)$ where $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$ and

$$
\|u*v\|_{L^r(\mathbb{R}^n)} \le \|u\|_{L^p(\mathbb{R}^n)}\|u\|_{L^q(\mathbb{R}^n)}.
$$

The convolution is a very useful device to *regularize* "wild functions". Indeed, consider the function η defined in (7.2). We have:

$$
\eta \ge 0 \quad \text{and} \quad \operatorname{supp}\left(\eta\right) = \overline{B}_1\left(\mathbf{0}\right)
$$

where, we recall, $B_R\left(\mathbf{0}\right) = \{\mathbf{x} \in \mathbb{R}^n \colon |\mathbf{x}| < R\}$. Choose

$$
c = \left(\int_{B_1(\mathbf{0})} \exp\left(\frac{1}{|\mathbf{x}|^2-1}\right) d\mathbf{x}\right)^{-1}
$$

so that $\int_{\mathbb{R}^n} \eta = 1$. Set, for $\varepsilon > 0$,

$$
\eta_\varepsilon\left(\mathbf{x}\right) = \frac{1}{\varepsilon^n}\eta\left(\frac{|\mathbf{x}|}{\varepsilon}\right). \tag{7.3}
$$

This function belongs to $C_0^\infty\left(\mathbb{R}^n\right)$ (and therefore to all $L^p\left(\mathbb{R}^n\right)$), with support equal to $\overline{B}_\varepsilon\left(\mathbf{0}\right)$, and still $\int_{\mathbb{R}^n} \eta_\varepsilon = 1$.

Let now $f \in L^p(\Omega)$. If we set $f \equiv 0$ outside Ω, we obtain a function in $L^p(\mathbb{R}^n)$, still denoted by f, for which the convolution $f*\eta_\varepsilon$ is well defined in all $\mathbb{R}^n$:

$$
f_\varepsilon\left(\mathbf{x}\right) = (f*\eta_\varepsilon)\left(\mathbf{x}\right) = \int_\Omega \eta_\varepsilon\left(\mathbf{x}-\mathbf{y}\right)f\left(\mathbf{y}\right)d\mathbf{y}
$$

$$
= \int_{B_\varepsilon(\mathbf{0})} \eta\left(\mathbf{z}\right)f\left(\mathbf{x}-\mathbf{z}\right)d\mathbf{z}.
$$

Observe that, since $\int_{\mathbb{R}^n} \eta_\varepsilon = 1$, $f * \eta_\varepsilon$ may be considered as a *convex weighted average* of f and, as such, we expect a smoothing effect on f. Indeed, even if f is very irregular, f_ε is a $C^\infty-$function. For this reason η_ε is called a *mollifier*. Moreover, as $\varepsilon \to 0$, f_ε is an approximation of f in the sense of the following important lemma.

Lemma 7.1. *Let $f \in L^p(\Omega)$; then f_ε has the following properties:*

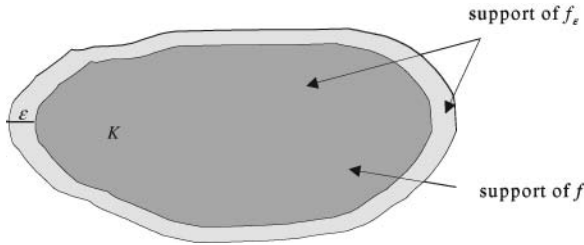**a.** *The support of f_ε is a $\varepsilon-$neighborhood of the support of f:*

$$\operatorname{supp}(f_\epsilon) \subseteq \{\mathbf{x} \in \mathbb{R}^n : \operatorname{dist}(\mathbf{x}, \operatorname{supp}(f)) \le \varepsilon\}.$$

b. *$f_\varepsilon \in C^\infty(\mathbb{R}^n)$ and if the support of f is a compact $K \subset \Omega$, then $f_\varepsilon \in C_0^\infty(\Omega)$, for $\varepsilon \ll 1$.*

c. *If $f \in C(\Omega)$, $f_\varepsilon \to f$ uniformly in every compact $K \subset \Omega$ as $\varepsilon \to 0$.*

d. *If $1 \le p < \infty$, then*

$$\|f_\varepsilon\|_{L^p(\Omega)} \le \|f\|_{L^p(\Omega)} \quad and \quad \|f_\varepsilon - f\|_{L^p(\Omega)} \to 0 \quad as \ \varepsilon \to 0.$$

Fig. 7.2. Support of the convolution with a ε-mollifier

Proof. **a.** Let $K = \operatorname{supp}(f)$. If $|\mathbf{z}| \le \varepsilon$ and $\operatorname{dist}(\mathbf{x}, K) > \varepsilon$, then $f(\mathbf{x} - \mathbf{z}) = 0$ so that $f_\varepsilon(\mathbf{x}) = 0$.

b. Since $\eta_\varepsilon(\mathbf{x} - \mathbf{y}) \in C_0^\infty(\mathbb{R}^n)$, f_ε is continuous and there is no problem in differentiating under the integral sign, obtaining all the time a continuous function. Thus $f_\varepsilon \in C^\infty(\mathbb{R}^n)$. From **a.**, if K is compact, the support of f_ε is compact as well and contained in Ω if $\varepsilon \ll 1$. Therefore $f_\varepsilon \in C_0^\infty(\Omega)$.

c. Since $\int_{\mathbb{R}^n} \eta_\varepsilon = 1$, We can write

$$f_\varepsilon(\mathbf{x}) - f(\mathbf{x}) = \int_{\{|\mathbf{z}| \le \varepsilon\}} \eta(\mathbf{z})\left[f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})\right] d\mathbf{z}.$$

Then, if $\mathbf{x} \in K \subset \Omega$, compact,

$$|f_\varepsilon(\mathbf{x}) - f(\mathbf{x})| \le \sup_{|\mathbf{z}| \le \varepsilon} |f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})|.$$

Since f is uniformly continuous in K we have that $\sup_{|\mathbf{z}| \leq \varepsilon} |f(\mathbf{x} - \mathbf{z}) - f(\mathbf{x})| \to 0$, uniformly in $\mathbf{x}$, as $\varepsilon \to 0$. Thus $f_\varepsilon \to f$ uniformly in K.

d. From Hölder's inequality, we have, for $q = p/(p-1)$,

$$f_\varepsilon(\mathbf{x}) = \int_\Omega \eta_\varepsilon(\mathbf{x} - \mathbf{y}) f(\mathbf{y}) \, d\mathbf{y} = \int_\Omega \eta_\varepsilon(\mathbf{x} - \mathbf{y})^{1/q} \eta_\varepsilon(\mathbf{x} - \mathbf{y})^{1/p} f(\mathbf{y}) \, d\mathbf{y}$$

$$\leq \left(\int_\Omega \eta_\varepsilon(\mathbf{x} - \mathbf{y}) |f(\mathbf{y})|^p \, d\mathbf{y} \right)^{1/p}.$$

This inequality and Fubini's Theorem[2] yield

$$\|f_\varepsilon\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega)}. \tag{7.4}$$

In fact:

$$\|f_\varepsilon\|_{L^p(\Omega)}^p = \int_\Omega |f_\varepsilon(\mathbf{x})|^p \, d\mathbf{x} \leq \int_\Omega \left(\int_\Omega \eta_\varepsilon(\mathbf{x} - \mathbf{y}) |f(\mathbf{y})|^p \, d\mathbf{y} \right) d\mathbf{x}$$

$$= \int_\Omega |f(\mathbf{y})|^p \left(\int_\Omega \eta_\varepsilon(\mathbf{x} - \mathbf{y}) \, d\mathbf{x} \right) d\mathbf{y} \leq \int_\Omega |f(\mathbf{y})|^p \, d\mathbf{y} = \|f\|_{L^p(\Omega)}^p.$$

From Theorem B.6, given any $\delta > 0$, there exists $g \in C_0(\Omega)$ such that $\|g - f\|_{L^p(\Omega)} < \delta$. Then, (7.4) implies

$$\|g_\varepsilon - f_\varepsilon\|_{L^p(\Omega)} \leq \|g - f\|_{L^p(\Omega)} < \delta.$$

Moreover, since the support of g is compact, $g_\varepsilon \to g$ uniformly in Ω, by **c**, so that, we have, in particular, $\|g_\varepsilon - g\|_{L^p(\Omega)} < \delta$, for ε small. Thus,

$$\|f - f_\varepsilon\|_{L^p(\Omega)} \leq \|f - g\|_{L^p(\Omega)} + \|g - g_\varepsilon\|_{L^p(\Omega)} + \|g_\varepsilon - f_\varepsilon\|_{L^p(\Omega)} \leq 3\delta.$$

This shows that $\|f - f_\varepsilon\|_{L^p(\Omega)} \to 0$ as $\varepsilon \to 0$. $\square$

Remark 7.1. Let $f \in L^1_{loc}(\Omega)$, i.e. $f \in L^1(\Omega')$ for every[3] $\Omega' \subset\subset \Omega$. The convolution $f_\varepsilon(\mathbf{x})$ is well defined if $\mathbf{x}$ stays $\varepsilon-$away from $\partial\Omega$, that is if $\mathbf{x}$ belongs to the set

$$\Omega_\varepsilon = \{\mathbf{x} \in \Omega : \text{dist}(\mathbf{x}, \partial\Omega) > \varepsilon\}.$$

Moreover, $f_\varepsilon \in C^\infty(\Omega_\varepsilon)$.

Remark 7.2. In general $\|f - f_\varepsilon\|_{L^\infty(\Omega)} \nrightarrow 0$ as $\varepsilon \to 0$. However $\|f_\varepsilon\|_{L^\infty(\Omega)} \leq \|f\|_{L^\infty(\Omega)}$ is clearly true.

---

[2] Appendix B.
[3] $\Omega' \subset\subset \Omega$ means that the closure of Ω' is a compact subset of Ω.

Example 7.2. Let $\Omega' \subset\subset \Omega$ and $f = \chi_{\Omega'}$ be the characteristic function of Ω'. Then, $f_\varepsilon = \chi_{\Omega'} * \eta_\varepsilon \in C_0^\infty (\Omega)$ as long as $\varepsilon < \text{dist}(\Omega', \partial\Omega)$. Note that $0 \le f_\varepsilon \le 1$. In fact

$$f_\varepsilon (\mathbf{x}) = \int_\Omega \eta_\varepsilon (\mathbf{x} - \mathbf{y}) \chi_{\Omega'} (\mathbf{y}) \, d\mathbf{y} = \int_{\Omega' \cap B_\varepsilon (\mathbf{x})} \eta_\varepsilon (\mathbf{x} - \mathbf{y}) \, d\mathbf{y} = \int_{\Omega' \cap B_\varepsilon (\mathbf{0})} \eta_\varepsilon (\mathbf{y}) \, d\mathbf{y} \le 1.$$

Moreover, $f \equiv 1$ in Ω'_ε . In fact, if $\mathbf{x} \in \Omega'_\varepsilon$, the ball $B_\varepsilon (\mathbf{x})$ is contained in Ω' and therefore

$$\int_{\Omega' \cap B_\varepsilon (\mathbf{x})} \eta_\varepsilon (\mathbf{x} - \mathbf{y}) \, d\mathbf{y} = \int_{B_\varepsilon (\mathbf{0})} \eta_\varepsilon (\mathbf{y}) \, d\mathbf{y} = 1.$$

A consequence of Lemma 7.1 is the following approximation theorem.

Theorem 7.1. $C_0^\infty (\Omega)$ *is dense in* $L^p (\Omega)$ *for every* $1 \le p < \infty$.

Proof. Denote by $L_c^p (\Omega)$ the space of functions in $L^p (\Omega)$, with (essential) support compactly contained in Ω. Let $f \in L_c^p (\Omega)$ and $K = \text{supp}(f)$. From Lemma 7.1.**a**, we know that $\text{supp}(f_\varepsilon)$ is a $\varepsilon-neighborhood$ of K, which is still a compact subset of Ω, for ε small.

Since by Lemma 7.1.**d**, $f_\varepsilon \to f$ in $L^p (\Omega)$, we deduce that $C_0^\infty (\Omega)$ is dense in $L_c^p (\Omega)$, if $1 \le p < \infty$. On the other hand, $L_c^p (\Omega)$ is dense in $L^p (\Omega)$; in fact, let $\{K_m\}$ be a sequence of compact subsets of Ω such that

$$K_m \subset K_{m+1} \qquad \text{and} \qquad \cup K_m = \Omega.$$

Denote by χ_{K_m} the characteristic function of K_m. Then, we have

$$\left\{ \chi_{K_m} f \right\} \subset L_c^p (\Omega) \quad \text{and} \quad \left\| \chi_{K_m} f - f \right\|_{L^p} \to 0 \quad \text{as } m \to +\infty$$

by the Dominated Convergence Theorem[4], since $\left| \chi_{K_m} f \right| \le |f|$. $\square$

7.3 Distributions

We now endow $C_0^\infty (\Omega)$ with a suitable notion of convergence. Recall that the symbol

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} ... \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}, \qquad \alpha = (\alpha_1, ..., \alpha_n),$$

denotes a derivative of order $|\alpha| = \alpha_1 + ... + \alpha_n$.

Definition 7.2. *Let* $\{\varphi_k\} \subset C_0^\infty (\Omega)$ *and* $\varphi \in C_0^\infty (\Omega)$. *We say that*

$$\varphi_k \to \varphi \qquad in \ C_0^\infty (\Omega) \quad as \ k \to +\infty$$

if:
1. *$D^\alpha \varphi_k \to D^\alpha \varphi$ uniformly in Ω, $\forall \alpha = (\alpha_1, ..., \alpha_n)$;*
2. *there exists a compact set $K \subset \Omega$ containing the support of every φ_k.*

---

[4] Appendix B.

It is possible to show that the limit so defined is *unique*. The space $C_0^\infty(\Omega)$ is denoted by $\mathcal{D}(\Omega)$, when endowed with the above notion of convergence.

Following the discussion in the first section, we focus on the linear functionals in $\mathcal{D}(\Omega)$. If L is one of those, we shall use the *bracket* (or *pairing*) $\langle L, \varphi \rangle$ to denote the action of L on a test function φ.

We say that linear functional

$$L : \mathcal{D}(\Omega) \to \mathbb{R}$$

is *continuous* in $\mathcal{D}(\Omega)$ if

$$\langle L, \varphi_k \rangle \to \langle L, \varphi \rangle, \qquad \text{whenever } \varphi_k \to \varphi \text{ in } \mathcal{D}(\Omega). \tag{7.5}$$

Note that, given the linearity of L, it would be enough to check (7.5) in the case $\varphi = 0$.

Definition 7.3. *A **distribution** in Ω is a linear continuous functional in $\mathcal{D}(\Omega)$. The set of distributions is denoted by $\mathcal{D}'(\Omega)$.*

Two distributions F and G coincide when their action on every test function is the same, i.e. if

$$\langle F, \varphi \rangle = \langle G, \varphi \rangle, \qquad \forall \varphi \in \mathcal{D}(\Omega).$$

To every $u \in L^2(\Omega)$ corresponds the functional I_u whose action on φ is

$$\langle I_u, \varphi \rangle = \int_\Omega u\varphi \, dx,$$

which is certainly continuous in $\mathcal{D}(\Omega)$. Therefore I_u is a distribution in $\mathcal{D}'(\Omega)$ and we have seen at the end of Section 7.1 that I_u may be identified with u.

Thus, the notion of distribution generalizes the notion of function (in $L^2(\Omega)$) and the pairing $\langle \cdot, \cdot \rangle$ between $\mathcal{D}(\Omega)$ and $\mathcal{D}'(\Omega)$ generalizes the inner product in $L^2(\Omega)$.

The same arguments show that every function $u \in L_{loc}^1(\Omega)$ belongs to $\mathcal{D}'(\Omega)$ and

$$\langle u, \varphi \rangle = \int_\Omega u\varphi \, dx.$$

On the other hand, if $u \notin L_{loc}^1$, u **cannot** represent a distribution. A typical example is $u(x) = 1/x$ which does not belongs to $L_{loc}^1(\mathbb{R})$. However, there is a distribution closely related to $1/x$ as we show in Example 7.6.

Example 7.3. (*Dirac delta*). The *Dirac delta* at the point $\mathbf{y}$, i.e. $\delta_{\mathbf{y}} : \mathcal{D}(\mathbb{R}^n) \to \mathbb{R}$, whose action is

$$\langle \delta_{\mathbf{y}}, \varphi \rangle = \varphi(\mathbf{y}),$$

is a distribution $\mathcal{D}'(\mathbb{R}^n)$, as it is easy to check.

$\mathcal{D}'(\Omega)$ is a linear space. Indeed if α, β are real (or complex) scalars, $\varphi \in \mathcal{D}(\Omega)$ and $L_1, L_2 \in \mathcal{D}'(\Omega)$, we define $\alpha L_1 + \beta L_2 \in \mathcal{D}'(\Omega)$ by means of the formula

$$\langle \alpha L_1 + \beta L_2, \varphi \rangle = \alpha \langle L_1, \varphi \rangle + \beta \langle L_2, \varphi \rangle.$$

In $\mathcal{D}'(\Omega)$ we may introduce a notion of (weak) convergence: $\{L_k\}$ converges to L in $\mathcal{D}'(\Omega)$ if

$$\langle L_k, \varphi \rangle \to \langle L, \varphi \rangle, \qquad \forall \varphi \in \mathcal{D}(\Omega).$$

If $1 \le p \le \infty$, we have the **continuous embeddings**:

$$L^p(\Omega) \hookrightarrow L^1_{loc}(\Omega) \hookrightarrow \mathcal{D}'(\Omega).$$

This means that, if $u_k \to u$ in $L^p(\Omega)$ or in $L^1_{loc}(\Omega)$, then[5] $u_k \to u$ in $\mathcal{D}'(\Omega)$ as well.

With respect to this convergence, $\mathcal{D}'(\Omega)$ possesses a *completeness* property that may be used to construct a distribution or to recognize that some linear functional in $\mathcal{D}(\Omega)$ is a distribution. Precisely, one can prove the following result.

Proposition 7.1. *Let* $\{F_k\} \subset \mathcal{D}'(\Omega)$ *such that*

$$\lim_{k \to \infty} \langle F_k, \varphi \rangle$$

exists and is finite for all $\varphi \in \mathcal{D}(\Omega)$. *Call* $F(\varphi)$ *this limit. Then,* $F \in \mathcal{D}'(\Omega)$ *and* $F_k \to F$ *in* $\mathcal{D}'(\Omega)$.

In particular, if the numerical series

$$\sum_{k=1}^{\infty} \langle F_k, \varphi \rangle$$

converges for all $\varphi \in \mathcal{D}(\Omega)$, then $\sum_{k=1}^{\infty} F_k = F \in \mathcal{D}'(\Omega)$.

Example 7.4. (*Dirac comb*). For every $\varphi \in \mathcal{D}(\mathbb{R})$, the numerical series

$$\sum_{k=-\infty}^{\infty} \langle \delta(x-k), \varphi \rangle = \sum_{k=-\infty}^{\infty} \varphi(k)$$

is convergent, since only a finite number of terms is different from zero[6]. From Proposition 7.1, we deduce that the series

$$comb(x) = \sum_{k=-\infty}^{\infty} \delta(x-k). \tag{7.6}$$

---

[5] For instance, let $\varphi \in \mathcal{D}(\Omega)$. We have, by Hőlder's inequality:

$$\left| \int_{\Omega} (u_k - u)\varphi d\mathbf{x} \right| \le \|u_k - u\|_{L^p(\Omega)} \|\varphi\|_{L^q(\Omega)}$$

where $q = p/(p-1)$. Then, if $\|u_k - u\|_{L^p(\Omega)} \to 0$, also $\int_{\Omega}(u_k - u)\varphi d\mathbf{x} \to 0$, showing the convergence of $\{u_k\}$ in $\mathcal{D}'(\Omega)$.

[6] Only a finite number of integers k belongs to the support of φ.

is convergent in $\mathcal{D}'(\mathbb{R})$ and its sum is a distribution called **Dirac comb**. This name is due to the fact it models a train of impulses concentrated at the integers (see Fig. 7.3, using some ...fantasy).
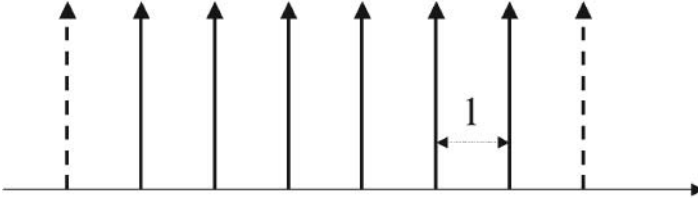
Fig. 7.3. A train of impluses

Example 7.5. Let $h_r(x) = 1 - \chi_{[-r,r]}(x)$ be the characteristic function of the set $\mathbb{R}\setminus[-r,r]$. Define

$$p.v.\frac{1}{x} = \lim_{r \to 0} \frac{1}{x} h_r(x).$$

We want to show that $p.v.\frac{1}{x}$ defines a distribution in $\mathcal{D}'(\mathbb{R})$, called *principal value of* $\frac{1}{x}$. By Proposition 7.1 it is enough to check that, for all $\varphi \in \mathcal{D}(\mathbb{R})$, the limit

$$\lim_{r \to 0} \int_{\mathbb{R}} \frac{1}{x} h_r(x) \varphi(x)\, dx$$

is finite. Indeed, assume that $\text{supp}(\varphi) \subset [-a, a]$. Then,

$$\int_{\mathbb{R}} \frac{1}{x} h_r(x) \varphi(x)\, dx = \int_{\{r < |x| < a\}} \frac{\varphi(x)}{x}\, dx = \int_{\{r < |x| < a\}} \frac{\varphi(x) - \varphi(0)}{x}\, dx$$

since

$$\int_{\{r < |x| < a\}} \frac{\varphi(0)}{x}\, dx = 0,$$

due to the odd symmetry of $1/x$. Now, we have

$$\varphi(x) - \varphi(0) = \varphi'(0) x + o(x), \qquad \text{as } x \to 0,$$

so that

$$\frac{\varphi(x) - \varphi(0)}{x} = \varphi'(0) + o(1) \qquad \text{as } x \to 0.$$

This implies that $[\varphi(x) - \varphi(0)]/x$ is summable in $[-a, a]$ and therefore

$$\lim_{r \to 0} \int_{\{r < |x| < a\}} \frac{\varphi(x) - \varphi(0)}{x}\, dx = \int_{\{|x| < a\}} \frac{\varphi(x) - \varphi(0)}{x}\, dx$$

is a finite number. Thus, $p.v.\frac{1}{x} \in \mathcal{D}'(\mathbb{R})$ and the above computations yield

$$\langle p.v.\frac{1}{x}, \varphi \rangle = \lim_{r \to 0} \int_{\{r < |x|\}} \frac{\varphi(x)}{x} dx \equiv p.v. \int_{\mathbb{R}} \frac{\varphi(x)}{x} dx$$

where $p.v.$ stays for *principal value*[7].

• *Support of a distribution.* The Dirac δ is *concentrated at a point.* More precisely, we say that its **support** coincides with a point. The support of a general distribution F may be defined in the following way. We want to characterize the smallest closed set outside of which F vanishes. However, we cannot proceed as in the case of a function, since a distribution is defined on the elements of $\mathcal{D}(\Omega)$, not on subsets of $\mathbb{R}^n$.

Thus, let us start saying that $F \in \mathcal{D}'(\Omega)$ *vanishes in an open set* $A \subset \Omega$ if

$$\langle F, \varphi \rangle = 0$$

for every $\varphi \in \mathcal{D}(\Omega)$ whose support is contained in A. Let $\mathcal{A}$ be the *union of all open sets where F vanishes.* $\mathcal{A}$ is open. Then, we define:

$$\text{supp}(F) = \Omega \backslash \mathcal{A}.$$

For example, $\text{supp}(comb) = \mathbb{Z}$.

Remark 7.3. Let $F \in \mathcal{D}'(\Omega)$ with compact support K. Then the bracket $\langle F, v \rangle$ is well defined for all $v \in C^\infty(\Omega)$, **not necessarily with compact support**. In fact, let $\varphi \in \mathcal{D}(\Omega)$, $0 \leq \varphi \leq 1$, such that $\varphi \equiv 1$ in an open neighborhood of K (see Remark 7.4). Then $v\varphi \in \mathcal{D}(\Omega)$ and we can define

$$\langle F, v \rangle = \langle F, v\varphi \rangle.$$

Note that $\langle F, v\varphi \rangle$ is independent of the choice of φ. Indeed if ψ has the same property of φ, then

$$\langle F, v\varphi \rangle - \langle F, v\psi \rangle = \langle F, v(\varphi - \psi) \rangle = 0$$

since $\varphi - \psi = 0$ in an open neighborhood of K.

7.4 Calculus

7.4.1 The derivative in the sense of distributions

A central concept in the theory of the Schwartz distributions is the notion of *weak* or *distributional derivative.* Clearly we have to abandon the classical definition,

---

[7] Whence the symbol $p.v.\frac{1}{x}$.

since, for instance, we are going to define the derivative for a function $u \in L^1_{loc}$, which may be quite irregular.

The idea is to carry the derivative onto the test functions, as if we were using the integration by parts formula.

Let us start from a function $u \in C^1(\Omega)$. If $\varphi \in \mathcal{D}(\Omega)$, denoting by $\boldsymbol{\nu} = (\nu_1, ..., \nu_n)$ the outward normal unit vector to $\partial\Omega$, we have

$$\int_\Omega \varphi \partial_{x_i} u \, d\mathbf{x} = \int_{\partial\Omega} \varphi u \, \nu_i \, d\mathbf{x} - \int_\Omega u \partial_{x_i} \varphi \, d\mathbf{x}$$

$$= -\int_\Omega u \partial_{x_i} \varphi \, d\mathbf{x}$$

since $\varphi = 0$ on $\partial\Omega$. The equation

$$\int_\Omega \varphi \, \partial_{x_i} u \, d\mathbf{x} = -\int_\Omega u \, \partial_{x_i} \varphi \, d\mathbf{x},$$

interpreted in $\mathcal{D}'(\Omega)$, becomes

$$\langle \partial_{x_i} u, \varphi \rangle = -\langle u, \partial_{x_i} \varphi \rangle. \tag{7.7}$$

Formula (7.7) shows that the action of $\partial_{x_i} u$ on the test function φ equals the action of u on the test function $-\partial_{x_i}\varphi$. On the other hand, formula (7.7) makes perfect sense if we replace u by any $F \in \mathcal{D}'(\Omega)$ and it is not difficult to check that it defines a continuous linear functional in $\mathcal{D}(\Omega)$. This leads to the following fundamental notion:

Definition 7.4. *Let $F \in \mathcal{D}'(\Omega)$. The derivative $\partial_{x_i} F$ is the distribution defined by the formula*

$$\langle \partial_{x_i} F, \varphi \rangle = -\langle F, \partial_{x_i} \varphi \rangle, \qquad \forall \varphi \in \mathcal{D}(\Omega) \, .$$

From (7.7), if $u \in C^1(\Omega)$ its derivatives in the sense of distributions coincide with the classical ones. This is the reason we keep the same notations in the two cases.

Note that the derivative of a distribution is **always defined**! Moreover, since any derivative of a distribution is a distribution, we deduce the convenient fact that **every distribution possesses derivatives of any order** (in $\mathcal{D}'(\Omega)$):

$$\langle D^\alpha F_k, \varphi \rangle = (-1)^{|\alpha|} \langle F_k, D^\alpha \varphi \rangle.$$

For example, the second order derivative

$$\partial_{x_i x_k} F = \partial_{x_i} (\partial_{x_k} F)$$

is defined by

$$\langle \partial_{x_i x_k} F, \varphi \rangle = \langle F, \partial_{x_i x_k} \varphi \rangle. \tag{7.8}$$

Not only. Since φ is smooth, then $\partial_{x_i x_k} \varphi = \partial_{x_k x_i} \varphi$ so that (7.8) yields

$$\partial_{x_i x_k} F = \partial_{x_k x_i} F.$$

Thus, for **all** $F \in \mathcal{D}'(\Omega)$ we may always reverse the order of differentiation *without any restriction*.

Example 7.6. Let $u(x) = \mathcal{H}(x)$, the Heaviside function. In $\mathcal{D}'(\mathbb{R})$ we have $\mathcal{H}' = \delta$. In fact, let $\varphi \in \mathcal{D}(\mathbb{R})$. By definition,

$$\langle \mathcal{H}', \varphi \rangle = -\langle \mathcal{H}, \varphi' \rangle.$$

On the other hand, $\mathcal{H} \in L^1_{loc}(\mathbb{R})$, hence

$$\langle \mathcal{H}, \varphi' \rangle = \int_{\mathbb{R}} \mathcal{H}(x)\, \varphi'(x)\, dx = \int_0^\infty \varphi'(x)\, dx = -\varphi(0)$$

whence

$$\langle \mathcal{H}', \varphi \rangle = \varphi(0) = \langle \delta, \varphi \rangle$$

or $\mathcal{H}' = \delta$.

Another aspect of the idyllic relationship between calculus and distributions is given by the following theorem, which expresses the continuity in $\mathcal{D}'(\Omega)$ of every derivative D^α.

Proposition 7.2. *If* $F_k \to F$ *in* $\mathcal{D}'(\Omega)$ *then* $D^\alpha F_k \to D^\alpha F$ *in* $\mathcal{D}'(\Omega)$ *for any multi-index* α.

Proof. $F_k \to F$ in $\mathcal{D}'(\Omega)$ means that $\langle F_k, \varphi \rangle \to \langle F, \varphi \rangle$, $\forall \varphi \in \mathcal{D}(\Omega)$.
In particular, since $D^\alpha \varphi \in \mathcal{D}(\Omega)$,

$$\langle D^\alpha F_k, \varphi \rangle = (-1)^{|\alpha|} \langle F_k, D^\alpha \varphi \rangle \to (-1)^{|\alpha|} \langle F, D^\alpha \varphi \rangle = \langle D^\alpha F, \varphi \rangle.$$

$\square$

As a consequence, if $\sum_{k=1}^\infty F_k = F$ in $\mathcal{D}'(\Omega)$, then

$$\sum_{k=1}^\infty D^\alpha F_k = D^\alpha F \qquad \text{in } \mathcal{D}'(\Omega).$$

Thus, term by term differentiation is **always** permitted in $\mathcal{D}'(\Omega)$.

More difficult is the proof of the following theorem, which expresses a well known fact for functions.

Proposition 7.3. *Let* Ω *be a domain in* $\mathbb{R}^n$. *If* $F \in \mathcal{D}'(\Omega)$ *and* $\partial_{x_j} F = 0$ *for every* $j = 1, ..., n$, *then* F *is a constant function.*

7.4.2 Gradient, divergence, laplacian

There is no problem to define *vector valued distributions*. The space of test functions is $\mathcal{D}(\Omega; \mathbb{R}^n)$, i.e. the set of vectors $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_n)$ whose components belong to $\mathcal{D}(\Omega)$.

A distribution $\mathbf{F} \in \mathcal{D}'(\Omega; \mathbb{R}^n)$ is given by $\mathbf{F} = (F_1, ..., F_n)$ with $F_j \in \mathcal{D}'(\Omega)$, $j = 1, ..., n$. The pairing between $\mathcal{D}(\Omega; \mathbb{R}^n)$ and $\mathcal{D}'(\Omega; \mathbb{R}^n)$ is defined by

$$\langle \mathbf{F}, \boldsymbol{\varphi} \rangle = \sum_{i=1}^n \langle F_i, \varphi_i \rangle. \tag{7.9}$$

• The gradient of $F \in \mathcal{D}'(\Omega)$, $\Omega \subset \mathbb{R}^n$, is simply

$$\nabla F = (\partial_{x_1} F, \partial_{x_2} F, ..., \partial_{x_n} F).$$

Clearly $\nabla F \in \mathcal{D}'(\Omega; \mathbb{R}^n)$. If $\boldsymbol{\varphi} \in \mathcal{D}(\Omega; \mathbb{R}^n)$, we have

$$\langle \nabla F, \boldsymbol{\varphi} \rangle = \sum_{i=1}^n \langle \partial_{x_i} F, \varphi_i \rangle = -\sum_{i=1}^n \langle F, \partial_{x_i} \varphi_i \rangle = -\langle F, \operatorname{div} \boldsymbol{\varphi} \rangle$$

whence

$$\langle \nabla F, \boldsymbol{\varphi} \rangle = -\langle F, \operatorname{div} \boldsymbol{\varphi} \rangle \tag{7.10}$$

which shows the action of ∇F on $\boldsymbol{\varphi}$.

• For $\mathbf{F} \in \mathcal{D}'(\Omega; \mathbb{R}^n)$, we set

$$\operatorname{div} \mathbf{F} = \sum_{i=1}^n \partial_{x_i} F_i.$$

Clearly $\operatorname{div} \mathbf{F} \in \mathcal{D}'(\Omega)$. If $\varphi \in \mathcal{D}(\Omega)$, then

$$\langle \operatorname{div} \mathbf{F}, \varphi \rangle = \langle \sum_{1=1}^n \partial_{x_i} F_i, \varphi \rangle = -\sum_{1=1}^n \langle F_i, \partial_{x_i} \varphi \rangle = -\langle \mathbf{F}, \nabla \varphi \rangle$$

whence

$$\langle \operatorname{div} \mathbf{F}, \varphi \rangle = -\langle \mathbf{F}, \nabla \varphi \rangle. \tag{7.11}$$

• The Laplace operator is defined in $\mathcal{D}'(\Omega)$ by

$$\Delta F = \sum_{i=1}^n \partial_{x_i x_i} F.$$

If $\varphi \in \mathcal{D}(\Omega)$, then

$$\langle \Delta F, \varphi \rangle = \langle F, \Delta \varphi \rangle.$$

Using (7.10), (7.11) we get

$$\langle \Delta F, \varphi \rangle = \langle F, \operatorname{div} \nabla \varphi \rangle = -\langle \nabla F, \nabla \varphi \rangle = \langle \operatorname{div} \nabla F, \varphi \rangle$$

whence $\Delta = \operatorname{div} \nabla$ also in $\mathcal{D}'(\Omega)$.

Example 7.7. Consider the **fundamental solution** for the Laplace operator in $\mathbb{R}^3$

$$u(\mathbf{x}) = \frac{1}{4\pi} \frac{1}{|\mathbf{x}|}.$$

Observe that $u \in L^1_{loc}(\mathbb{R}^3)$ so that $u \in \mathcal{D}'(\mathbb{R}^3)$. We want to show that, in $\mathcal{D}'(\mathbb{R}^3)$,

$$-\Delta u = \delta. \tag{7.12}$$

First of all, if $\Omega \subset \mathbb{R}^3$ and $\mathbf{0} \notin \Omega$, we know that u is *harmonic in* Ω, that is

$$\Delta u = 0 \qquad \text{in } \Omega$$

in the classical sense and therefore also in $\mathcal{D}'\left(\mathbb{R}^3\right)$. Thus, let $\varphi \in \mathcal{D}\left(\mathbb{R}^3\right)$ with $\mathbf{0} \in \text{supp}(\varphi)$. We have, since $u \in L^1_{loc}\left(\mathbb{R}^3\right)$:

$$\langle \Delta u, \varphi \rangle = \langle u, \Delta \varphi \rangle = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{1}{|\mathbf{x}|} \Delta \varphi(\mathbf{x}) \, d\mathbf{x}. \tag{7.13}$$

We would like to carry the laplacian onto $1/|\mathbf{x}|$. However, this cannot be done directly, since the integrand is not continuous at $\mathbf{0}$. Therefore we exclude a small sphere $B_r = B_r(\mathbf{0})$ from our integration region and write

$$\int_{\mathbb{R}^3} \frac{1}{|\mathbf{x}|} \Delta \varphi(\mathbf{x}) \, d\mathbf{x} = \lim_{r \to 0} \int_{B_R \setminus B_r} \frac{1}{|\mathbf{x}|} \Delta \varphi(\mathbf{x}) \, d\mathbf{x} \tag{7.14}$$

where $B_R = B_R(\mathbf{0})$ is a sphere containing the support of φ. An integration by parts in the spherical shell $C_{R,r} = B_R \setminus B_r$ yields[8]

$$\int_{C_{R,r}} \frac{1}{|\mathbf{x}|} \Delta \varphi(\mathbf{x}) \, d\mathbf{x} = \int_{\partial B_r} \frac{1}{r} \partial_{\boldsymbol{\nu}} \varphi(\mathbf{x}) \, d\sigma - \int_{C_{R,r}} \nabla \left(\frac{1}{|\mathbf{x}|}\right) \cdot \nabla \varphi(\mathbf{x}) \, d\mathbf{x}$$

where $\boldsymbol{\nu} = -\frac{\mathbf{x}}{|\mathbf{x}|}$ is the *outward* normal unit vector on ∂B_r. Integrating once more by parts the last integral, we obtain:

$$\int_{C_{R,r}} \nabla \left(\frac{1}{|\mathbf{x}|}\right) \cdot \nabla \varphi(\mathbf{x}) \, d\mathbf{x} = \int_{\partial B_r} \partial_{\boldsymbol{\nu}} \left(\frac{1}{|\mathbf{x}|}\right) \varphi(\mathbf{x}) \, d\sigma - \int_{C_{R,r}} \Delta \left(\frac{1}{|\mathbf{x}|}\right) \varphi(\mathbf{x}) \, d\mathbf{x}$$

$$= \int_{\partial B_r} \partial_{\boldsymbol{\nu}} \left(\frac{1}{|\mathbf{x}|}\right) \varphi(\mathbf{x}) \, d\sigma,$$

since $\Delta \left(\frac{1}{|\mathbf{x}|}\right) = 0$ inside $C_{R,r}$. From the above computations we infer

$$\int_{C_{R,r}} \frac{1}{|\mathbf{x}|} \Delta \varphi(\mathbf{x}) \, d\mathbf{x} = \int_{\partial B_r} \frac{1}{r} \partial_{\boldsymbol{\nu}} \varphi(\mathbf{x}) \, d\sigma - \int_{\partial B_r} \partial_{\boldsymbol{\nu}} \left(\frac{1}{|\mathbf{x}|}\right) \varphi(\mathbf{x}) \, d\sigma. \tag{7.15}$$

We have:

$$\frac{1}{r} \left| \int_{\partial B_r} \partial_{\boldsymbol{\nu}} \varphi(\mathbf{x}) \, d\sigma \right| \leq \frac{1}{r} \int_{\partial B_r} |\partial_{\boldsymbol{\nu}} \varphi(\mathbf{x})| \, d\sigma \leq 4\pi r \max_{\mathbb{R}^3} |\nabla \varphi|$$

and therefore

$$\lim_{r \to 0} \int_{\partial B_r} \frac{1}{r} \partial_{\boldsymbol{\nu}} \varphi(\mathbf{x}) \, d\sigma = 0.$$

---

[8] Recall that $\varphi = 0$ and $\nabla \varphi = \mathbf{0}$ on ∂B_R.

Moreover, since

$$\partial_{\boldsymbol{\nu}}\left(\frac{1}{|\mathbf{x}|}\right) = \nabla\left(\frac{1}{|\mathbf{x}|}\right) \cdot \left(-\frac{\mathbf{x}}{|\mathbf{x}|}\right) = \left(-\frac{\mathbf{x}}{|\mathbf{x}|^3}\right) \cdot \left(-\frac{\mathbf{x}}{|\mathbf{x}|}\right) = \frac{1}{|\mathbf{x}|^2},$$

we may write

$$\int_{\partial B_r} \partial_\nu\left(\frac{1}{|\mathbf{x}|}\right)\varphi(\mathbf{x})\,d\sigma = 4\pi\frac{1}{4\pi r^2}\int_{\partial B_r}\varphi(\mathbf{x})\,d\sigma \to 4\pi\varphi(\mathbf{0}).$$

Thus, from (7.15) we get

$$\lim_{r\to 0}\int_{B_R\backslash B_r}\frac{1}{|\mathbf{x}|}\Delta\varphi(\mathbf{x})\,d\mathbf{x} = -4\pi\varphi(\mathbf{0})$$

and finally (7.13) yields

$$\langle\Delta u,\varphi\rangle = -\varphi(\mathbf{0}) = -\langle\delta,\varphi\rangle$$

whence $-\Delta u = \delta$.

7.5 Multiplication, Composition, Division, Convolution

7.5.1 Multiplication. Leibniz rule

Let us analyze the multiplication between two distributions. Does it make any sense to define, for instance, the product $\delta \cdot \delta = \delta^2$ as a distribution in $\mathcal{D}'(\mathbb{R})$?

Things are not so smooth. An idea for defining δ^2 may be the following: take a sequence $\{u_k\}$ of functions in $L^1_{loc}(\mathbb{R})$ such that $u_k \to \delta$ in $\mathcal{D}'(\mathbb{R})$, compute u_k^2 and set

$$\delta^2 = \lim_{k\to\infty} u_k^2 \qquad \text{in } \mathcal{D}'(\mathbb{R}).$$

Since we may approximate δ in $\mathcal{D}'(\mathbb{R})$ in many ways (see Problem 7.1), it is necessary that the definition *does not depend* on the approximating sequence. In other words, to compute δ^2 we must be free to choose any approximating sequence. However, this is illusory. Indeed choose

$$u_k = k\chi_{[0,1/k]}.$$

We have $u_k \to \delta$ in $\mathcal{D}'(\mathbb{R})$ but, if $\varphi \in \mathcal{D}(\mathbb{R})$, by the Mean Value Theorem we have

$$\int_{\mathbb{R}} u_k^2\varphi = k^2\int_0^{1/k}\varphi = k\varphi(x_k)$$

for some $x_k \in [0, 1/k]$. Now, if $\varphi(0) > 0$, say, we deduce that

$$\int_{\mathbb{R}} u_k^2\varphi \to +\infty, \qquad k \to +\infty$$

so that $\left\{u_k^2\right\}$ *does not converge in* $\mathcal{D}'\left(\mathbb{R}\right)$.

The method does not work and it seems that there is no other reasonable way to define δ^2. Thus, we simply give up defining δ^2 as a distribution or, in general, the product of a pair of distributions. However, if $F \in \mathcal{D}'\left(\Omega\right)$ and $u \in C^\infty\left(\Omega\right)$, we may define the product uF by the formula

$$\langle uF, \varphi \rangle = \langle F, u\varphi \rangle, \qquad \forall \varphi \in \mathcal{D}\left(\Omega\right).$$

First of all, this makes sense since $u\varphi \in \mathcal{D}\left(\Omega\right)$. Also, if $\varphi_k \to \varphi$ in $\mathcal{D}\left(\Omega\right)$, then $u\varphi_k \to u\varphi$ in $\mathcal{D}\left(\Omega\right)$ and

$$\langle uF, \varphi_k \rangle = \langle F, u\varphi_k \rangle \to \langle F, u\varphi \rangle = \langle uF, \varphi \rangle.$$

so that uF is a well defined element of $\mathcal{D}'\left(\Omega\right)$.

Example 7.8. Let $u \in C^\infty\left(\mathbb{R}\right)$. We have

$$u\delta = u\left(0\right)\delta.$$

Indeed, if $\varphi \in \mathcal{D}\left(\mathbb{R}\right)$,

$$\langle u\delta, \varphi \rangle = \langle \delta, u\varphi \rangle = u\left(0\right)\varphi\left(0\right) = \langle u\left(0\right)\delta, \varphi \rangle.$$

Note that the product $u\delta$ makes sense even if u is only continuous. In particular

$$x\delta = 0.$$

The *Leibniz* rule holds: let $F \in \mathcal{D}'\left(\Omega\right)$ and $u \in C^\infty\left(\Omega\right)$; then

$$\partial_{x_i}\left(uF\right) = u\,\partial_{x_i}F + \partial_{x_i}u\;F\;. \tag{7.16}$$

In fact, let $\varphi \in \mathcal{D}\left(\Omega\right)$; we have:

$$\langle \partial_{x_i}\left(uF\right), \varphi \rangle = -\langle uF, \partial_{x_i}\varphi \rangle = -\langle F, u\partial_{x_i}\varphi \rangle$$

while

$$\langle u\,\partial_{x_i}F + \partial_{x_i}u\;F, \varphi \rangle = \langle \partial_{x_i}F, u\varphi \rangle + \langle F, \varphi\partial_{x_i}u \rangle$$
$$= -\langle F, \partial_{x_i}\left(u\varphi\right) \rangle + \langle F, \varphi\partial_{x_i}u \rangle = \langle F, u\partial_{x_i}\varphi \rangle$$

and (7.16) follows.

Example 7.9. From $x\delta = 0$ and Leibniz formula we obtain

$$\delta + x\delta' = 0.$$

More generally,

$$x^m\delta^{(k)} = 0 \quad \text{in } \mathcal{D}'\left(\mathbb{R}\right), \quad \text{if } 0 \leq k < m.$$

7.5.2 Composition

Composition in $\mathcal{D}'(\mathbb{R})$ requires caution as well. For instance, if $F = \delta$ and $u(x) = x^3$, is there a natural way to define $F \circ u$ as a distribution in $\mathcal{D}'(\mathbb{R})$?

As above, consider the sequence $u_k = k\chi_{[0,1/k]}$ and compute $w_k = u_k \circ u$. If $\varphi \in \mathcal{D}'(\mathbb{R})$, we have

$$\int_{\mathbb{R}} w_k \varphi = k \int_{\mathbb{R}} \chi_{[0,1/k]}(x^3)\varphi(x)\,dx = k \int_0^{k^{-1/3}} \varphi(x)\,dx = k^{2/3}\varphi(x_k)$$

for some $x_k \in [0, 1/k]$. Then, if $\varphi(0) > 0$, $\int_{\mathbb{R}} w_k \varphi \to +\infty$ and $F \circ u$ does not make any sense. Thus, it seems hard to define the composition between two general distributions. To see what can be done, let us start analyzing the case of two functions.

Let $\psi : \Omega' \to \Omega$ be **one to one, with ψ and ψ^{-1} of class C^∞**. If $F : \Omega \to \mathbb{R}$ is a C^1–function we may consider the composition

$$w = F \circ \psi.$$

For $\varphi \in \mathcal{D}(\Omega)$, we have, using the change of variables $\mathbf{y} = \psi(\mathbf{x})$:

$$\int_{\Omega'} w(\mathbf{x})\,\varphi(\mathbf{x})\,d\mathbf{x} = \int_{\Omega'} F(\psi(\mathbf{x}))\,\varphi(\mathbf{x})\,d\mathbf{x} = \int_{\Omega} F(\mathbf{y}))\,\varphi(\psi^{-1}(\mathbf{y}))\,\big|\det J_{\psi^{-1}}(\mathbf{y})\big|\,d\mathbf{y}$$

which becomes, in terms of distributions,

$$\langle F \circ \psi, \varphi \rangle = \langle F, \varphi \circ \psi^{-1} \cdot \big|\det J_{\psi^{-1}}\big| \rangle. \tag{7.17}$$

This formula makes sense also if $F \in \mathcal{D}'(\Omega)$ and leads to the following

Definition 7.5. *If $F \in D'(\Omega)$ and $\psi : \Omega' \to \Omega$ is one to one, with ψ and ψ^{-1} of class C^∞, then formula (7.17) defines the composition $F \circ \psi$ as an element of $\mathcal{D}'(\Omega)$.*

Abuses of notation are quite common, like $F(\psi(\mathbf{x}))$ to denote $F \circ \psi$. For instance, we have repeatedly used the (comfortable and incorrect) notation $\delta(\mathbf{x} - \mathbf{x}_0)$ instead of the (uncomfortable and correct....) notation $\delta \circ \psi$, with $\psi(\mathbf{x}) = \mathbf{x} - \mathbf{x}_0$.

Example 7.10. In $\mathcal{D}'(\mathbb{R}^n)$, we have

$$\delta(a\mathbf{x}) = \frac{1}{|a|^n}\delta(\mathbf{x}).$$

Using formula (7.17) we may extend to distributions some properties, typical of functions. We list some of them.

We say that $F \in \mathcal{D}'(\mathbb{R}^n)$ is:

- *radial*, if

$$F(\mathbf{Ax}) = F(\mathbf{x}), \qquad \text{for every orthogonal matrix } \mathbf{A};$$

- *homogeneous of degree* λ, if

$$F(t\mathbf{x}) = t^\lambda F(\mathbf{x}), \qquad \forall t > 0;$$

- *even, odd* if, respectively,

$$F(-\mathbf{x}) = F(\mathbf{x}), \ F(-\mathbf{x}) = -F(\mathbf{x});$$

- *periodic with period* $\mathbf{P}$, if

$$F(\mathbf{x} + \mathbf{P}) = F(\mathbf{x}).$$

Example 7.11. **a.** $\delta \in \mathcal{D}'(\mathbb{R}^n)$ is radial, even and homogeneous of degree $\lambda = -n$.

b. $v.p.\frac{1}{x} \in \mathcal{D}'(\mathbb{R})$ is odd and homogeneous of degree $\lambda = -1$.

c. In $\mathcal{D}'(\mathbb{R})$, *comb* is periodic with period 1.

7.5.3 Division

The division in $\mathcal{D}'(\Omega)$ is rather delicate, even restricting to $F \in \mathcal{D}'(\Omega)$ and $u \in C^\infty(\Omega)$. To divide F by u means to find $G \in \mathcal{D}'(\Omega)$ such that $uG = F$. If u *never vanishes* there is no problem, since in this case $1/u \in C^\infty(\Omega)$ and the answer is simply

$$G = \frac{1}{u}F.$$

If u vanishes somewhere, things get complicated. We only consider a particular case in one dimension.

Let $I \subseteq \mathbb{R}$ be an **open interval** and $u \in C^\infty(I)$. If u vanishes at z, we say that z is a *zero of order* $m(z)$ if the derivatives of u up to order $m(z) - 1$, included, vanish at z, while the derivative of order $m(z)$ does not vanish at z.

For instance, $z = 0$ is a zero of order 3 for $u(x) = \sin x - x$.

One can prove the following theorem.

Proposition 7.4. *Assume that u vanishes at isolated points $z_1, z_2, \ldots$ with order $m(z_1), m(z_2), \ldots$. Then, the equation*

$$uG = 0$$

has infinitely many solutions in $\mathcal{D}'(I)$, given by the following formula:

$$G = \sum_j \sum_{k=0}^{m(z_j)-1} c_{jk} \delta^{(k)}(x - z_j) \tag{7.18}$$

where c_{jk} are arbitrary constants and $\delta^{(k)}$ is the derivative of δ of order k.

Example 7.12. The solutions in $\mathcal{D}'(\mathbb{R})$ of the equation

$$xG = 0.$$

are the distributions of the form $G = c\delta$, with $c \in \mathbb{R}$. To solve the equation

$$xG = 1 \tag{7.19}$$

we add to the solutions of the homogeneous equation $xG = 0$ a particular solution of (7.19). It turns out that one of these is

$$G_1 = v.p.\frac{1}{x}.$$

In fact, if $\varphi \in \mathcal{D}(\mathbb{R})$, from Example 7.6 we get

$$\langle x \cdot (v.p.\frac{1}{x}), \varphi \rangle = \langle v.p.\frac{1}{x}, x\varphi \rangle =$$

$$= v.p. \int_{\mathbb{R}} \frac{x\varphi(x)}{x} dx = \int_{\mathbb{R}} \varphi(x) \, dx = \langle 1, \varphi \rangle$$

whence

$$x \cdot (v.p.\frac{1}{x}) = 1.$$

Therefore, the general solution of (7.19) is

$$G = v.p.\frac{1}{x} + c\delta, \qquad c \in \mathbb{R}.$$

7.5.4 Convolution

The convolution of two distributions may be defined with some restrictions as well. Let us see why. If $u, w \in L^1(\mathbb{R}^n)$ and $\varphi \in \mathcal{D}(\mathbb{R}^n)$ we may write:

$$\langle u * w, \varphi \rangle = \langle \int_{\mathbb{R}^n} u(\mathbf{x} - \mathbf{y}) w(\mathbf{y}) \, d\mathbf{y}, \varphi \rangle =$$

$$= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^n} u(\mathbf{x} - \mathbf{y}) w(\mathbf{y}) \, d\mathbf{y} \right] \varphi(\mathbf{x}) \, d\mathbf{x} =$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} u(\mathbf{x}) w(\mathbf{y}) \varphi(\mathbf{x} + \mathbf{y}) \, d\mathbf{y} d\mathbf{x}.$$

Now, the question is: may we give any meaning to this formula if u and v are generic distributions? The answer is negative, mainly because the function

$$\phi(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x} + \mathbf{y})$$

does **not** have compact support[9] in $\mathbb{R}^n \times \mathbb{R}^n$ (unless $\varphi \equiv 0$).

---

[9] For instance: if $\varphi \in \mathcal{D}'(\mathbb{R})$ and $\mathrm{supp}(\varphi) = [a, b]$, then the support of $\varphi(x + y)$ in $\mathbb{R}^2$ is the unbounded strip $a \leq x + y \leq b$.

However, a small modification of the above formula would give the possibility to define the convolution between two distributions, if *at least one of them has compact support.* Here we limit ourselves to define the convolution between a distribution T and a $C^\infty-$function u. For $\mathbf{x}$ fixed, let $\psi^{\mathbf{x}}(\mathbf{y}) = \mathbf{x} - \mathbf{y}$ so that $u(\mathbf{x} - \mathbf{y}) = u \circ \psi^{\mathbf{x}}$.

If $T \in L^1(\mathbb{R}^n)$, with compact support, then the usual definition of convolution is

$$(T * u)(\mathbf{x}) = \int_{\mathbb{R}^n} T(\mathbf{y}) u(\mathbf{x} - \mathbf{y}) d\mathbf{y} = \langle T, u \circ \psi^{\mathbf{x}} \rangle. \qquad (7.20)$$

Since $u \circ \psi^{\mathbf{x}}$ is a $C^\infty-$function, recalling Remark 7.4, the last bracket makes sense if T is a distribution with compact support as well. Precisely, we have:

Proposition 7.5. *Let* $T \in \mathcal{D}'(\mathbb{R}^n)$, *with compact support, and* $u \in C^\infty(\mathbb{R}^n)$. *Then, the following formula*

$$(T * u)(\mathbf{x}) = \langle T, u \circ \psi^{\mathbf{x}} \rangle \qquad (7.21)$$

defines a $C^\infty-$*function called* **convolution** *of* T *and* u.

Example 7.13. Let $u \in C_0^\infty(\mathbb{R}^n)$. Then

$$(\delta * u)(\mathbf{x}) = \langle \delta, u(\mathbf{x} - \cdot) \rangle = u(\mathbf{x})$$

i.e

$$\delta * u = u. \qquad (7.22)$$

Thus, the Dirac distribution at zero, acts as the **identity** with respect to the convolution. Formula (7.22) actually holds for all $u \in \mathcal{D}'(\mathbb{R}^n)$. In particular:

$$\delta * \delta = \delta.$$

Proposition 7.6. *The convolution commutes with derivatives. Actually, we have:*

$$\partial_{x_j}(T * u) = \partial_{x_j} T * u = T * \partial_{x_j} u.$$

The last equality is easy to prove, under the hypotheses of Proposition 7.5:

$$(\partial_{x_j} T * u)(\mathbf{z}) = \langle \partial_{x_j} T, u \circ \psi^{\mathbf{z}} \rangle = -\langle T, \partial_{x_j}(u \circ \psi^{\mathbf{z}}) \rangle$$
$$= \langle T, \partial_{x_j} u \circ \psi^{\mathbf{z}} \rangle = (T * \partial_{x_j} u)(\mathbf{z}).$$

In particular, if $T = \mathcal{H}$ and $u \in \mathcal{D}(\mathbb{R})$,

$$(\mathcal{H} * u)' = (\mathcal{H}' * u) = \delta * u = u.$$

Warning: The convolution of functions is associative. For distributions, the convolution is, in general, not *associative*. In fact, consider the three distributions $1, \delta', \mathcal{H}$; we have (formally):

$$\delta' * 1 = (\delta * 1)' = 1' = 0$$

whence

$$\mathcal{H} * \left(\delta' * 1 \right) = \mathcal{H} * 0 = 0.$$

However,

$$\left(\mathcal{H} * \delta' \right) * 1 = \left(\mathcal{H}' * \delta \right) * 1 = \left(\delta * \delta \right) * 1 = 1.$$

The problem is that two out of three factors (1 and $\mathcal{H}$) have non compact support. If at least *two factors have compact support* one can show that the convolution is associative.

7.6 Fourier Transform

7.6.1 Tempered distributions

We introduce now the Fourier transform $\hat{F}$ of a distribution. As usual, the idea is to define the action of $\hat{F}$ by carrying the transform onto the test functions. However a problem immediately arises: if $\varphi \in \mathcal{D}(\mathbb{R}^n)$ is not identically zero, then

$$\widehat{\varphi}\left(\boldsymbol{\xi} \right) = \int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\boldsymbol{\xi}} \varphi\left(\mathbf{x} \right) d\mathbf{x}$$

cannot belong[10] to $\mathcal{D}(\mathbb{R}^n)$. Thus, it is necessary to choose a larger space of test functions. It turns out that the correct one consists in the set of functions *rapidly vanishing at* ∞, which obviously contains $\mathcal{D}(\mathbb{R}^n)$. It is convenient to consider *functions and distributions with complex values*.

Definition 7.6. *Denote by $\mathcal{S}\left(\mathbb{R}^n\right)$ the space of functions $v \in C^{\infty}(\mathbb{R}^n)$* **rapidly vanishing at infinity***, i.e. such that*

$$D^{\alpha} v\left(\mathbf{x} \right) = o\left(|\mathbf{x}|^{-m} \right), \qquad |\mathbf{x}| \to \infty,$$

for all $m \in \mathbb{N}$ and every multi-index α.

---

[10] Let $n = 1$ and $\varphi \in \mathcal{D}\left(\mathbb{R}\right)$. Assume that

$$\text{supp}\left(\varphi \right) \subset \left(-a, a \right).$$

We may write

$$\widehat{\varphi}\left(\xi \right) = \int_{-a}^{a} e^{-ix\xi} \varphi\left(x \right) dx = \int_{-a}^{a} \sum_{n=0}^{\infty} \frac{(-ix\xi)^n}{n!} \varphi\left(x \right) dx = \sum_{n=0}^{\infty} \frac{(-i\xi)^n}{n!} \int_{-a}^{a} x^n \varphi\left(x \right) dx.$$

Since

$$\left| \int_{-a}^{a} x^n \varphi\left(x \right) dx \right| \leq \max |\varphi| \, a^n,$$

it follows that $\widehat{\varphi}$ is an *analytic* function in all $\mathbb{C}$. Therefore $\widehat{\varphi}$ cannot vanish outside a compact interval, unless $\widehat{\varphi} \equiv 0$. But then $\varphi \equiv 0$ as well.

Example 7.14. The function $v(\mathbf{x}) = e^{-|\mathbf{x}|^2}$ belongs to $\mathcal{S}(\mathbb{R}^n)$ while $v(\mathbf{x}) = e^{-|x|^2} \sin(e^{|\mathbf{x}|^2})$ does not (why?).

We endow $\mathcal{S}(\mathbb{R}^n)$ with an "ad hoc" notion of convergence. If $\beta = (\beta_1, ..., \beta_n)$ is a multi-index, we set

$$\mathbf{x}^\beta = x_1^{\beta_1} \cdots x_1^{n_n}.$$

Definition 7.7. *Let $\{v_k\} \subset \mathcal{S}(\mathbb{R}^n)$ and $v \in \mathcal{S}(\mathbb{R}^n)$. We say that*

$$v_k \to v \qquad \text{in } \mathcal{S}(\mathbb{R}^n)$$

if for every pair of multi-indexes α, β,

$$\mathbf{x}^\beta D^\alpha v_k \to \mathbf{x}^\beta D^\alpha v, \quad \text{uniformly in } \mathbb{R}^n.$$

Remark 7.4. If $\{v_k\} \subset \mathcal{D}(\mathbb{R}^n)$ and $v_k \to v$ in $\mathcal{D}(\mathbb{R}^n)$, then

$$v_k \to v \quad \text{in } \mathcal{S}(\mathbb{R}^n)$$

as well, since each v_k vanishes outside a common compact set so that the multiplication by $\mathbf{x}^\beta$ does not have any influence.

The Fourier transform will be defined for distributions in $\mathcal{D}'(\mathbb{R}^n)$, continuous with respect to the convergence in Definition 7.7. These are the so called *tempered distributions.* Precisely:

Definition 7.8. *We say that $T \in \mathcal{D}'(\mathbb{R}^n)$ is a **tempered** distribution if*

$$\langle T, v_k \rangle \to 0$$

for all sequences $\{v_k\} \subset \mathcal{D}(\mathbb{R}^n)$ such that $v_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$. The set of tempered distributions is denoted by $\mathcal{S}'(\mathbb{R}^n)$.

So far, a tempered distribution T is only defined on $\mathcal{D}(\mathbb{R}^n)$. To define T on $\mathcal{S}(\mathbb{R}^n)$, first we observe that $\mathcal{D}(\mathbb{R}^n)$ *is dense in* $\mathcal{S}(\mathbb{R}^n)$.

In fact, given $v \in \mathcal{S}(\mathbb{R}^n)$, let

$$v_k(\mathbf{x}) = v(\mathbf{x}) \rho(|\mathbf{x}|/k)$$

where $\rho = \rho(s)$, $s \geq 0$, is a non-negative $C^\infty-$function, equal to 1 in $[0, 1]$ and zero for $s \geq 2$ (see Fig. 7.4). We have $\{v_k\} \subset \mathcal{D}(\mathbb{R}^n)$ and $v_k \to v$ in $\mathcal{S}(\mathbb{R}^n)$, since $\rho(|\mathbf{x}|/k)$ is equal to 1 for $\{|\mathbf{x}| < k\}$ and zero for $\{|\mathbf{x}| > 2k\}$.

Then, we set

$$\langle T, v \rangle = \lim_{k \to \infty} \langle T, v_k \rangle. \tag{7.23}$$

It can be shown that this limit exists and is finite, and it is independent of the approximating sequence $\{v_k\}$. Thus, **a tempered distribution is a continuous functional on** $\mathcal{S}(\mathbb{R}^n)$.
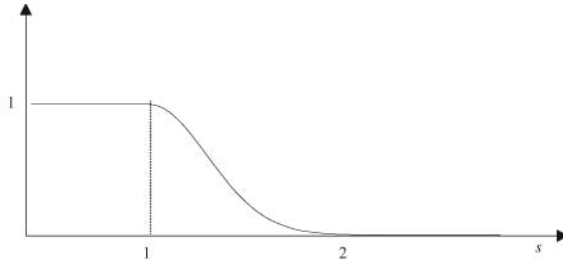
Fig. 7.4. A smooth decreasing and nonnegative function, equal to 1 in $[0, 1]$ and vanishing for $s \geq 2$

Example 7.15. We leave it as an exercise to show that the following distributions are tempered.

    a. Any polynomial.
    b. Any compactly supported distribution.
    c. Any periodic distribution (e.g. the Dirac comb).
    d. Any function $u \in L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$. Thus, we have

$$\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n).$$

On the contrary:

    e. $e^x \notin \mathcal{S}'(\mathbb{R})$ (why?).

Like $\mathcal{D}'(\Omega)$, $\mathcal{S}'(\mathbb{R}^n)$ possesses a *completeness* property that may be used to construct a tempered distribution or to recognize that some linear functional in $\mathcal{D}(\mathbb{R}^n)$ is a tempered distribution. First, we say that *a sequence* $\{T_k\} \subset \mathcal{S}'(\mathbb{R}^n)$ *converges to T in $\mathcal{S}'(\mathbb{R}^n)$ if*

$$\langle T_k, v \rangle \to \langle T, v \rangle, \quad \forall v \in \mathcal{S}(\mathbb{R}^n).$$

We have:

Proposition 7.7. *Let $\{T_k\} \subset \mathcal{S}'(\mathbb{R}^n)$ such that*

$$\lim_{k \to \infty} \langle T_k, v \rangle \text{ exists and is finite}, \quad \forall v \in \mathcal{S}(\mathbb{R}^n).$$

Then, this limit defines $T \in \mathcal{S}'(\mathbb{R}^n)$ and T_k converges to T in $\mathcal{S}'(\mathbb{R}^n)$.

Example 7.16. The Dirac *comb* is a tempered distribution. In fact, if $v \in \mathcal{S}(\mathbb{R})$, we have

$$\langle comb(x), v \rangle = \sum_{k=-\infty}^{\infty} v(k)$$

and the series is convergent since $v(k) \to 0$ more rapidly than $|k|^{-m}$ for every $m > 0$. From Proposition 7.7, $comb(x) \in \mathcal{S}'(\mathbb{R})$.

Remark 7.5. Convolution. If $T \in \mathcal{S}'(\mathbb{R}^n)$ and $v \in \mathcal{S}(\mathbb{R}^n)$, the convolution is well defined by formula (7.21). Then, $T * v \in \mathcal{S}'(\mathbb{R}^n)$ and coincides with a function in $C^\infty(\mathbb{R}^n)$.

7.6.2 Fourier transform in $\mathcal{S}'$

If $u \in L^1(\mathbb{R}^n)$, its Fourier transform is given by

$$\widehat{u}(\boldsymbol{\xi}) = \mathcal{F}[u](\boldsymbol{\xi}) = \int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\boldsymbol{\xi}} u(\mathbf{x}) \, d\mathbf{x}.$$

It could be that, even if u is compactly supported, $\widehat{u} \notin L^1(\mathbb{R}^n)$. For instance, if $p_a(x) = \chi_{[-a,a]}(x)$ then

$$\widehat{p}(\xi) = 2\frac{\sin(a\xi)}{\xi}$$

which is not[11] in $L^1(\mathbb{R})$. When also $\widehat{u} \in L^1(\mathbb{R}^n)$, u can be reconstructed from $\widehat{u}$ through the following *inversion* formula:

Theorem 7.2. *Let $u \in L^1(\mathbb{R}^n)$, $\widehat{u} \in L^1(\mathbb{R}^n)$. Then*

$$u(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\mathbf{x}\cdot\boldsymbol{\xi}}\widehat{u}(\boldsymbol{\xi})d\boldsymbol{\xi} \equiv \mathcal{F}^{-1}[\widehat{u}](\mathbf{x}). \qquad (7.24)$$

In particular, the inversion formula (7.24) holds for $u \in \mathcal{S}(\mathbb{R}^n)$, since (exercise) $\widehat{u} \in \mathcal{S}(\mathbb{R}^n)$ as well. Moreover, it may be proved that

$$u_k \to u \qquad \text{in } \mathcal{S}(\mathbb{R}^n)$$

if and only if

$$\widehat{u}_k \to \widehat{u} \qquad \text{in } \mathcal{S}(\mathbb{R}^n),$$

which means that

$$\mathcal{F}, \, \mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$$

are *continuous operators*

Now observe that, if $u, v \in \mathcal{S}(\mathbb{R}^n)$,

$$\langle \widehat{u}, v \rangle = \langle \int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\boldsymbol{\xi}} u(\mathbf{x})d\mathbf{x}, v \rangle = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\boldsymbol{\xi}} u(\mathbf{x})d\mathbf{x} \right) v(\boldsymbol{\xi}) \, d\boldsymbol{\xi}$$

$$= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\boldsymbol{\xi}} v(\boldsymbol{\xi}) \, d\boldsymbol{\xi} \right) u(\mathbf{x})d\mathbf{x} = \langle u, \widehat{v} \rangle$$

so that (*weak Parseval identity*):

$$\langle \widehat{u}, v \rangle = \langle u, \widehat{v} \rangle. \qquad (7.25)$$

The key point is that the last bracket makes sense for $u = T \in \mathcal{S}'(\mathbb{R}^n)$ as well, and defines a tempered distribution. In fact:

---

[11] Appendix B.

Lemma 7.2. *Let $T \in \mathcal{S}'(\mathbb{R}^n)$. The linear functional*

$$v \mapsto \langle T, \widehat{v} \rangle, \qquad \forall v \in \mathcal{S}(\mathbb{R}^n)$$

is a tempered distribution.

Proof. Let $v_k \to v$ in $\mathcal{D}(\mathbb{R}^n)$. Then $v_k \to v$ and $\widehat{v}_k \to \widehat{v}$ in $\mathcal{S}(\mathbb{R}^n)$ as well. Since $T \in \mathcal{S}'(\mathbb{R}^n)$, we have

$$\lim_{k \to \infty} \langle T, \widehat{v}_k \rangle = \langle T, \widehat{v} \rangle$$

so that $v \mapsto \langle T, \widehat{v} \rangle$ defines a distribution. If $v_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$, then $\widehat{v}_k \to 0$ in $\mathcal{S}(\mathbb{R}^n)$ and $\langle T, \widehat{v}_k \rangle \to 0$. Thus, $v \mapsto \langle T, \widehat{v} \rangle$ is a tempered distribution. $\square$

We are now in position to define the Fourier transform of $T \in \mathcal{S}'(\mathbb{R}^n)$.

Definition 7.9. *Let $T \in S'(\mathbb{R}^n)$. The Fourier transform $\widehat{T} = \mathcal{F}[T]$ is the tempered distribution defined by*

$$\langle \widehat{T}, v \rangle = \langle T, \widehat{v} \rangle, \qquad \forall v \in \mathcal{S}(\mathbb{R}^n).$$

We see that the transform has been carried onto the test function $v \in (\mathbb{R}^n)$. As a consequence, all the properties valid for functions, continue to hold for tempered distributions. We list some of them. Let $T \in \mathcal{S}'(\mathbb{R}^n)$ and $v \in \mathcal{S}(\mathbb{R}^n)$.

1. *Translation*. If $\mathbf{a} \in \mathbb{R}^n$,

$$\mathcal{F}[T(\mathbf{x} - \mathbf{a})] = e^{-i\mathbf{a} \cdot \boldsymbol{\xi}} \widehat{T} \quad \text{and} \quad \mathcal{F}[e^{i\mathbf{a} \cdot \mathbf{x}} T] = \widehat{T}(\boldsymbol{\xi} - \mathbf{a}).$$

In fact $(v = v(\boldsymbol{\xi}))$:

$$\langle \mathcal{F}[T(\mathbf{x} - \mathbf{a})], v \rangle = \langle T(\mathbf{x} - \mathbf{a}), \widehat{v} \rangle = \langle T, \widehat{v}(\mathbf{x} + \mathbf{a}) \rangle$$
$$= \langle T, \mathcal{F}[e^{-i\mathbf{a} \cdot \boldsymbol{\xi}} v] \rangle = \langle \widehat{T}, e^{-i\mathbf{a} \cdot \boldsymbol{\xi}} v \rangle = \langle e^{-i\mathbf{a} \cdot \boldsymbol{\xi}} \widehat{T}, v \rangle.$$

2. *Rescaling*. If $h \in \mathbb{R}$, $h \neq 0$,

$$\mathcal{F}[T(h\mathbf{x})] = \frac{1}{|h|^n} \widehat{T}\left(\frac{\boldsymbol{\xi}}{h}\right).$$

In fact:

$$\langle \mathcal{F}[T(h\mathbf{x})], v \rangle = \langle T(h\mathbf{x}), \widehat{v} \rangle = \langle T, \frac{1}{|h|^n} \widehat{v}\left(\frac{\mathbf{x}}{h}\right) \rangle$$

$$= \langle T, \mathcal{F}[v(h\boldsymbol{\xi})] \rangle = \langle \widehat{T}, v(h\boldsymbol{\xi}) \rangle = \langle \frac{1}{|h|^n} \widehat{T}\left(\frac{\boldsymbol{\xi}}{h}\right), v \rangle.$$

In particular, choosing $h = -1$, it follows that if T is *even* (*odd*) *then* $\widehat{T}$ *is even* (*odd*).

3. *Derivatives*:

$$a)\ \mathcal{F}[\partial_{x_j} T] = i\xi_j \widehat{T} \quad \text{and} \quad b)\ \mathcal{F}[x_j T] = i\partial_{\xi_j} \widehat{T}.$$

Namely:

$$\left\langle \mathcal{F}\left[\partial_{x_j}T\right], v\right\rangle = \left\langle \partial_{x_j}T, \widehat{v}\right\rangle = -\left\langle T, \partial_{x_j}\widehat{v}\right\rangle$$
$$= \left\langle T, \mathcal{F}\left[i\xi_j v\right]\right\rangle = \left\langle i\xi_j\widehat{T}, v\right\rangle.$$

For the second formula, we have:

$$\left\langle \mathcal{F}\left[x_j T\right], v\right\rangle = \left\langle x_j T, \widehat{v}\right\rangle = \left\langle T, x_j\widehat{v}\right\rangle$$
$$= \left\langle T, -i\mathcal{F}[\partial_{\xi_j} v]\right\rangle = \left\langle -i\widehat{T}, \partial_{\xi_j} v\right\rangle = \left\langle i\partial_{\xi_j}\widehat{T}, v\right\rangle.$$

4. *Convolution*[12]. If $T \in \mathcal{S}'(\mathbb{R}^n)$ and $v \in \mathcal{S}(\mathbb{R}^n)$,

$$\mathcal{F}[T * v] = \widehat{T} \cdot \widehat{v}.$$

Example 7.17. We know that $\delta \in \mathcal{S}'(\mathbb{R}^n)$. We have:

$$\widehat{\delta} = 1, \qquad \widehat{1} = (2\pi)^n \delta.$$

In fact

$$\langle \widehat{\delta}, v\rangle = \langle \delta, \widehat{v}\rangle = \int_{\mathbb{R}^n} v(\mathbf{x})\, d\mathbf{x} = \langle 1, v\rangle.$$

For the second formula, using (7.24) we have:

$$\langle \widehat{1}, v\rangle = \langle 1, \widehat{v}\rangle = \int_{\mathbb{R}^n} \widehat{v}(\boldsymbol{\xi})\, d\boldsymbol{\xi} = (2\pi)^n v(\mathbf{0})$$
$$= \langle (2\pi)^n \delta, v\rangle.$$

Example 7.18. Transform of x_j :

$$\widehat{x}_j = i(2\pi)^n \partial_{\xi_j}\delta.$$

Indeed, from 3, *b*) and Example 7.20, we may write

$$\widehat{x}_j = \mathcal{F}[x_j \cdot 1] = i\partial_{\xi_j}\widehat{1} = i(2\pi)^n \partial_{\xi_j}\delta.$$

7.6.3 Fourier transform in L^2

Since $L^2(\mathbb{R}^n) \subset \mathcal{S}'(\mathbb{R}^n)$, the Fourier transform is well defined for all functions in $L^2(\mathbb{R}^n)$. The following theorem holds, where $\overline{v}$ denotes the complex conjugate of v.

Theorem 7.3. $u \in L^2(\mathbb{R}^n)$ *if and only if* $\widehat{u} \in L^2(\mathbb{R}^n)$. *Moreover, if* $u, v \in L^2(\mathbb{R}^n)$, *the following strong Parseval identity holds:*

$$\int_{\mathbb{R}^n} \widehat{u} \cdot \overline{\widehat{v}} = (2\pi)^n \int_{\mathbb{R}^n} u \cdot \overline{v}. \tag{7.26}$$

In particular

$$\|\widehat{u}\|^2_{L^2(\mathbb{R}^n)} = (2\pi)^n \|u\|^2_{L^2(\mathbb{R}^n)}. \tag{7.27}$$

---

[12] We omit the proof.

Formula (7.27) shows that *the Fourier transform is an isometry in $L^2(\mathbb{R}^n)$* (but for the factor $(2\pi)^n$).

Proof. Since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$, it is enough to prove (7.26) for $u, v \in \mathcal{S}(\mathbb{R}^n)$. Let $w = \widehat{\overline{v}}$. From (7.25) we have

$$\int_{\mathbb{R}^n} \widehat{u} \cdot w = \int_{\mathbb{R}^n} u \cdot \widehat{w}.$$

On the other hand,

$$\widehat{w}(\mathbf{x}) = \int_{\mathbb{R}^n} e^{-i\mathbf{x}\cdot\mathbf{y}} \overline{\widehat{v}}(\mathbf{y})\, d\mathbf{y} = (2\pi)^n \overline{\mathcal{F}^{-1}[\widehat{v}]}(\mathbf{x}) = (2\pi)^n \overline{v}(\mathbf{x})$$

and (7.26) follows. $\square$

Example 7.19. Let us compute

$$\int_{\mathbb{R}} \left(\frac{\sin x}{x}\right)^2 dx.$$

We know that the Fourier transform of $p_1 = \chi_{[-1,1]}$ is $\hat{p}_1(\xi) = 2\sin\xi/\xi$, which belongs to $L^2(\mathbb{R})$. Thus, (7.27) yields

$$4\int_{\mathbb{R}} \left(\frac{\sin\xi}{\xi}\right)^2 d\xi = 2\pi \int_{\mathbb{R}} \left(\chi_{[-1,1]}(x)\right)^2 dx = 4\pi$$

whence

$$\int_{\mathbb{R}} \left(\frac{\sin x}{x}\right)^2 dx = \pi.$$

7.7 Sobolev Spaces

7.7.1 An abstract construction

Sobolev spaces constitute one of the most relevant functional settings for the treatment of boundary value problems. Here, we will be mainly concerned with Sobolev spaces based on $L^2(\Omega)$, developing only the theoretical elements we will need in the sequel[13].

The following abstract theorem is a flexible tool for generating Sobolev Spaces. The ingredients of the construction are:

- The space $\mathcal{D}'(\Omega; \mathbb{R}^n)$, in particular, for $n = 1$, $\mathcal{D}'(\Omega)$.

---

[13] We omit the most technical proofs, that can be found, for instance, in the classical books of *Adams*, 1975, or *Mazja*, 1985.

- Two Hilbert spaces H and Z with $Z \hookrightarrow \mathcal{D}'(\Omega; \mathbb{R}^n)$ for some $n \geq 1$. In particular

$$v_k \to v \text{ in } Z \quad \textbf{implies} \quad v_k \to v \text{ in } \mathcal{D}'(\Omega; \mathbb{R}^n). \qquad (7.28)$$

- A linear continuous operator $L : H \to \mathcal{D}'(\Omega; \mathbb{R}^n)$ (such as a gradient or a divergence).

We have*:*

Theorem 7.4. *Define*

$$W = \{v \in H : Lv \in Z\}$$

and

$$(u, v)_W = (u, v)_H + (Lu, Lv)_Z. \qquad (7.29)$$

Then W is a Hilbert space with inner product given by (7.29). The embedding of W in H is continuous and the restriction of L to W is continuous from W into Z.

Proof. It is easy to check that (7.29) has all the properties of an inner product, with induced norm

$$\|u\|_W^2 = \|u\|_H^2 + \|Lu\|_Z^2.$$

Thus W is an inner-product space. It remains to check its completeness. Let $\{v_k\}$ a Cauchy sequence in W. We must show that there exists $v \in H$ such that

$$v_k \to v \text{ in } H \quad \text{and} \quad Lv_k \to Lv \text{ in } Z.$$

Observe that $\{v_k\}$ and $\{Lv_k\}$ are Cauchy sequences in H and Z, respectively. Thus, there exist $v \in H$ and $z \in Z$ such that

$$v_k \to v \quad \text{in } H \quad \text{and} \quad Lv_k \to z \quad \text{in } Z.$$

The continuity of L and (7.28) yield

$$Lv_k \to Lv \text{ in } \mathcal{D}'(\Omega; \mathbb{R}^n) \quad \text{and} \quad Lv_k \to z \text{ in } \mathcal{D}'(\Omega; \mathbb{R}^n).$$

Since the limit of a sequence in $\mathcal{D}'(\Omega; \mathbb{R}^n)$ is unique, we infer that

$$Lv = z.$$

Therefore $Lv_k \to Lv$ in Z and W is a Hilbert space.

The continuity of the embedding $W \subset H$ follows from

$$\|u\|_H \leq \|u\|_W$$

while the continuity of $L_{|W} : W \to Z$ follows from

$$\|Lu\|_Z \leq \|u\|_W.$$

Thus, the proof is complete. $\square$

Remark 7.6. The norm induced by the inner product (7.29) is

$$\|u\|_W = \sqrt{\|u\|_H^2 + \|Lu\|_Z^2}$$

which is called the *graph norm of L*.

7.7.2 The space $H^1(\Omega)$

Let $\Omega \subseteq \mathbb{R}^n$ be a domain. Choose in Theorem 7.4:

$$H = L^2(\Omega), \quad Z = L^2(\Omega; \mathbb{R}^n) \hookrightarrow \mathcal{D}'(\Omega; \mathbb{R}^n)$$

and $L : H \to \mathcal{D}'(\Omega; \mathbb{R}^n)$ given by

$$L = \nabla$$

where the gradient is considered in the sense of distributions. Then, W is the **Sobolev space** of the functions in $L^2(\Omega)$, whose *first derivatives in the sense of distributions are functions in $L^2(\Omega)$*. For this space we use the symbol[14] $H^1(\Omega)$. Thus:

$$H^1(\Omega) = \left\{ v \in L^2(\Omega) : \nabla v \in L^2(\Omega; \mathbb{R}^n) \right\}.$$

In other words, if $v \in H^1(\Omega)$, every partial derivative $\partial_{x_i} v$ is a function $v_i \in L^2(\Omega)$. This means that

$$\langle \partial_{x_i} v, \varphi \rangle = -(v, \partial_{x_i} \varphi)_0 = (v_i, \varphi)_0, \qquad \forall \varphi \in \mathcal{D}(\Omega)$$

or, more explicitly,

$$\int_\Omega v(\mathbf{x}) \, \partial_{x_i} \varphi(\mathbf{x}) \, d\mathbf{x} = -\int_\Omega v_i(\mathbf{x}) \, \varphi(\mathbf{x}) \, d\mathbf{x}, \qquad \forall \varphi \in \mathcal{D}(\Omega).$$

In many applied situations, the Dirichlet integral

$$\int_\Omega |\nabla v|^2$$

represents an energy. The functions in $H^1(\Omega)$ are therefore associated with *configurations having finite energy*. From Theorem 7.4 and the separability[15] of $L^2(\Omega)$, we have:

Proposition 7.8. $H^1(\Omega)$ *is a separable Hilbert space, continuously embedded in* $L^2(\Omega)$. *The gradient operator is continuous from* $H^1(\Omega)$ *into* $L^2(\Omega; \mathbb{R}^n)$.

The inner product and the norm in $H^1(\Omega)$ are given, respectively, by

$$(u, v)_{H^1(\Omega)} = \int_\Omega uv \, d\mathbf{x} + \int_\Omega \nabla u \cdot \nabla v \, d\mathbf{x}.$$

---

[14] Also $H^{1,2}(\Omega)$ or $W^{1,2}(\Omega)$ are used.

[15] If we associate with each element u of $H^1(\Omega)$ the vector $u, u_{x_1}, ..., u_{x_n}$, we see that $H^1(\Omega)$ can be identified with a subspace of

$$L^2(\Omega) \times L^2(\Omega) \times ... \times L^2(\Omega) = L^2(\Omega; \mathbb{R}^{n+1})$$

which is separable because $L^2(\Omega)$ is separable.

and

$$\|u\|^2_{H^1(\Omega)} = \int_\Omega u^2 d\mathbf{x} + \int_\Omega |\nabla u|^2 \, d\mathbf{x}.$$

If no confusion arises, we will use the symbols

$$(u,v)_{1,2} \quad \text{instead of} \quad (u,v)_{H^1(\Omega)}$$

and[16]

$$\|u\|_{1,2} \quad \text{instead of} \quad \|u\|_{H^1(\Omega)}.$$

Example 7.20. Let $\Omega = B_1(\mathbf{0}) = \left\{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| < 1\right\}$ and

$$u(\mathbf{x}) = (-\log|\mathbf{x}|)^a, \qquad \mathbf{x} \neq \mathbf{0}.$$

We have, using polar coordinates,

$$\int_{B_1(\mathbf{0})} u^2 = 2\pi \int_0^1 (-\log r)^{2a}\, r dr < \infty, \qquad \text{for every } a \in \mathbb{R},$$

so that $u \in L^2(B_1(\mathbf{0}))$ for every $a \in \mathbb{R}$. Moreover:

$$u_{x_i} = -a x_i |\mathbf{x}|^{-2} (-\log|\mathbf{x}|)^{a-1}, \, i = 1,2,$$

and therefore

$$|\nabla u| = \left| a(-\log|\mathbf{x}|)^{a-1}\right| |\mathbf{x}|^{-1}.$$

Thus, using polar coordinates, we get

$$\int_{B_1(\mathbf{0})} |\nabla u|^2 = 2\pi a^2 \int_0^1 |\log r|^{2a-2}\, r^{-1} dr.$$

This integral is finite only if $2-2a > 1$ or $a < 1/2$. In particular, ∇u represents the gradient of u in the sense of distribution as well. We conclude that $u \in H^1(B_1(\mathbf{0}))$ only if $a < 1/2$.

We point out that when $a > 0$, u is **unbounded** near $\mathbf{0}$.

We have affirmed that the Sobolev spaces constitute an adequate functional setting to solve boundary value problems. This point requires that we go more deeply into the arguments in Section 6.1 and make some necessary observations. When we write $f \in L^2(\Omega)$, we may think of a single function

$$f : \Omega \to \mathbb{R} \text{ (or } \mathbb{C}),$$

square summable in the Lebesgue sense. However, if we want to exploit the Hilbert space structure of $L^2(\Omega)$, we need to identify two functions when they are equal a.e. in Ω. Adopting this point of view, each element in $L^2(\Omega)$ is actually an

---

[16] The numbers $1,2$ in the symbol $\|\cdot\|_{1,2}$ stay for "*first* derivatives in L^2".

equivalence class of which f is a *representative*. The drawback here is that it does not make sense anymore to compute the *value of f at a single point*, since a point is a set with measure zero!

The same considerations hold for "functions" in $H^1(\Omega)$, since

$$H^1(\Omega) \subset L^2(\Omega).$$

On the other hand, if we deal with a boundary value problem, it is clear that *we would like to compute the solution at any point in Ω*!

Even more important is the question of the *trace of a function on the boundary of a domain*. By *trace* of f on $\partial\Omega$ we mean the restriction of f to $\partial\Omega$. In a Dirichlet or Neumann problem we assign precisely the trace of the solution or of its normal derivative on $\partial\Omega$, which is a set with measure zero. Does this make any sense if $u \in H^1(\Omega)$?

It could be objected that, after all, one always works with a single representative and that the numerical approximation of the solution, only involves a finite number of points, making meaningless the distinction between functions in $L^2(\Omega)$ or in $H^1(\Omega)$ or continuous. Then, why do we have to struggle to give a precise meaning to the trace of a function in $H^1(\Omega)$?

One reason comes from numerical analysis, in particular from the need to keep under control the approximation errors and to give stability estimates.

Let us ask, for instance: if a Dirichlet data is known within an error of order ε in L^2-norm on $\partial\Omega$, can we estimate in terms of ε the corresponding error in the solution?

If we are satisfied with an L^2 or an L^∞ norm *in the interior of* the domain, this kind of estimate may be available. But often, an energy estimate is required, involving the norm in $L^2(\Omega)$ of the gradient of the solution. In this case, the L^2 norm of the boundary data is not sufficient and it turns out that the exact information on the data, necessary to restore an energy estimate, is encoded in the trace characterization of Section 7.9.

We shall introduce the notion of *trace on $\partial\Omega$* for a function in $H^1(\Omega)$, using an approximation procedure with smooth functions. However, there are two cases, in which the trace problem may be solved quite simply: the one-dimensional case and the case of functions with zero trace. We start with the first case.

• *Characterization of $H^1(a,b)$*. As Example 7.26 shows, a function in $H^1(\Omega)$ may be unbounded. In dimension $n = 1$ this cannot occur. In fact, the elements in $H^1(a,b)$ are continuous functions[17] in $[a,b]$.

Proposition 7.9. *Let $u \in L^2(a,b)$. Then $u \in H^1(a,b)$ if and only if u is continuous in $[a,b]$ and there exists $w \in L^2(a,b)$ such that*

$$u(y) = u(x) + \int_x^y w(s)\,ds, \qquad \forall x, y \in [a,b]. \tag{7.30}$$

Moreover $u' = w$ (both a.e. and in the sense of distribution).

---

[17] Rigorously: every *equivalence class in $H^1(a,b)$ has a representative continuous in $[a,b]$.*

Proof. Assume that u is continuous in $[a, b]$ and that (7.30) holds with $w \in L^2(a, b)$. Choose $x = a$. Replacing, if necessary, u by $u - u(a)$, we may assume $u(a) = 0$, so that

$$u(y) = \int_a^y w(s) \, ds, \qquad \forall x, y \in [a, b].$$

Let $v \in \mathcal{D}(a, b)$. We have:

$$\langle u', v \rangle = -\langle u, v' \rangle = -\int_a^b u(s) \, v'(s) \, ds = -\int_a^b \left[\int_a^s w(t) \, dt \right] v'(s) \, ds =$$

(exchanging the order of integration)

$$= -\int_a^b \left[\int_t^b v'(s) \, ds \right] w(t) \, dt = \int_a^b v(t) \, w(t) \, dt = \langle w, v \rangle.$$

Thus $u' = w$ in $\mathcal{D}'(a, b)$ and therefore $u \in H^1(a, b)$. From the Lebesgue Differentiation Theorem[18] we deduce that $u' = w$ a.e. as well.

Viceversa, let $u \in H^1(a, b)$. Define

$$v(x) = \int_c^x u'(s) \, ds, \qquad x \in [a, b]. \tag{7.31}$$

The function v is continuous in $[a, b]$ and the above proof shows that $v' = u'$ in $\mathcal{D}'(a, b)$. Then (Proposition 7.3) $u = v + C$, $C \in \mathbb{R}$ and therefore u is continuous in $[a, b]$ as well. Moreover, (7.31) yields

$$u(y) - u(x) = v(y) - v(x) = \int_x^y u'(s) \, ds$$

which is (7.30). $\square$

Since a function $u \in H^1(a, b)$ is continuous in $[a, b]$, the value $u(x_0)$ at every point $x_0 \in [a, b]$ makes perfect sense. In particular *the trace of u at the end points* of the interval is given by the values $u(a)$ and $u(b)$.

7.7.3 The space $H_0^1(\Omega)$

Let $\Omega \subseteq \mathbb{R}^n$. We introduce an important subspace of $H^1(\Omega)$.

Definition 7.10. *We denote by $H_0^1(\Omega)$ the closure of $\mathcal{D}(\Omega)$ in $H^1(\Omega)$.*

Thus $u \in H_0^1(\Omega)$ if and only if there exists a sequence $\{\varphi_k\} \subset \mathcal{D}(\Omega)$ such that $\varphi_k \to u$ in $H^1(\Omega)$, i.e. both $\|\varphi_k - u\|_0 \to 0$ and $\|\nabla \varphi_k - \nabla u\|_0 \to 0$ as $k \to \infty$.

Since the test functions in $\mathcal{D}(\Omega)$ have zero trace on $\partial \Omega$, every $u \in H_0^1(\Omega)$ "inherits" this property and it is reasonable to consider the elements $H_0^1(\Omega)$ as the functions in $H^1(\Omega)$ with *zero trace on $\partial \Omega$*. Clearly, $H_0^1(\Omega)$ is a Hilbert subspace of $H^1(\Omega)$.

An important property that holds in $H_0^1(\Omega)$, particularly useful in the solution of boundary value problems, is expressed by the following inequality of *Poincaré*.

---

[18] Appendix B.

Theorem 7.5. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. There exists a positive constant C_P (Poincaré's constant) such that, for every $u \in H_0^1(\Omega)$,*

$$\|u\|_0 \le C_P \|\nabla u\|_0. \tag{7.32}$$

Proof. We use a strategy which is rather common for proving formulas in $H_0^1(\Omega)$. First, we prove the formula for $v \in \mathcal{D}(\Omega)$; then, if $u \in H_0^1(\Omega)$, select a sequence $\{v_k\} \subset \mathcal{D}(\Omega)$ converging to u in norm $\|\cdot\|_{1,2}$ as $k \to \infty$, that is

$$\|v_k - u\|_0 \to 0, \qquad \|\nabla v_k - \nabla u\|_0 \to 0.$$

In particular

$$\|v_k\|_0 \to \|u\|_0, \qquad \|\nabla v_k\|_0 \to \|\nabla u\|_0.$$

Since (7.32) holds for every v_k, we have

$$\|v_k\|_0 \le C_P \|\nabla v_k\|_0.$$

Letting $k \to \infty$ we obtain (7.32) for u. Thus, it is enough to prove (7.32) for $v \in \mathcal{D}(\Omega)$. To this purpose, from the Gauss Divergence Theorem, we may write

$$\int_\Omega \operatorname{div}\left(v^2\mathbf{x}\right) d\mathbf{x} = 0 \tag{7.33}$$

since $v = 0$ on $\partial\Omega$. Now,

$$\operatorname{div}\left(v^2\mathbf{x}\right) = 2v\nabla v \cdot \mathbf{x} + nv^2$$

so that (7.33) yields

$$\int_\Omega v^2 d\mathbf{x} = -\frac{2}{n} \int_\Omega v\nabla v \cdot \mathbf{x}\, d\mathbf{x}.$$

Since Ω is bounded, we have $\max_{\mathbf{x} \in \Omega} |\mathbf{x}| = M < \infty$; therefore, using Schwarz's inequality, we get

$$\int_\Omega v^2 d\mathbf{x} = \frac{2}{n}\left|\int_\Omega v\nabla v \cdot \mathbf{x}\, d\mathbf{x}\right| \le \frac{2M}{n}\left(\int_\Omega v^2 d\mathbf{x}\right)^{1/2}\left(\int_\Omega |\nabla v|^2 d\mathbf{x}\right)^{1/2}.$$

Simplyfying, it follows that

$$\|v\|_0 \le C_P \|\nabla v\|_0$$

with $C_P = 2M/n$. $\square$

Inequality (7.32) implies that in $H_0^1(\Omega)$ the norm $\|u\|_{1,2}$ is equivalent to $\|\nabla u\|_0$. Indeed

$$\|u\|_{1,2} = \sqrt{\|u\|_0^2 + \|\nabla u\|_0^2}$$

and from (7.32),

$$\|\nabla u\|_0 \le \|u\|_{1,2} \le \sqrt{C_P^2 + 1}\, \|\nabla u\|_0.$$

Unless explicitly stated, **we will choose in $H_0^1(\Omega)$**

$$(u, v)_1 = (\nabla u, \nabla v)_0 \qquad \text{and} \qquad \|u\|_1 = \|\nabla u\|_0$$

as inner product and norm, respectively.

7.7.4 The dual of $H_0^1(\Omega)$

In the applications of the Lax-Milgram theorem to boundary value problems, the dual of $H_0^1(\Omega)$ plays an important role. In fact it deserves a special symbol.

Definition 7.11. *We denote by $H^{-1}(\Omega)$ the dual of $H_0^1(\Omega)$ with the norm*

$$\|F\|_{-1} = \sup\left\{|Fv| : v \in H_0^1(\Omega), \ \|v\|_1 \le 1\right\}.$$

The first thing to observe is that, since $\mathcal{D}(\Omega)$ is dense (by definition) and continuously embedded in $H_0^1(\Omega)$, $H^{-1}(\Omega)$ is a *space of distributions*. This means two things:

a) if $F \in H^{-1}(\Omega)$, its restriction to $\mathcal{D}(\Omega)$ is a distribution;

b) if $F, G \in H^{-1}(\Omega)$ and $F\varphi = G\varphi$ for every $\varphi \in \mathcal{D}(\Omega)$, then $F = G$.

To prove a) it is enough to note that if $\varphi_k \to \varphi$ in $\mathcal{D}(\Omega)$, then $\varphi_k \to \varphi$ in $H_0^1(\Omega)$ as well, and therefore $F\varphi_k \to F\varphi$. Thus $F \in \mathcal{D}'(\Omega)$.

To prove b), let $u \in H_0^1(\Omega)$ and $\varphi_k \to u$ in $H_0^1(\Omega)$, with $\varphi_k \in \mathcal{D}(\Omega)$. Then, since $F\varphi_k = G\varphi_k$ we may write

$$Fu = \lim_{k \to +\infty} F\varphi_k = \lim_{k \to +\infty} G\varphi_k = Gu$$

whence $F = G$.

Thus, $H^{-1}(\Omega)$ is in *one-to-one* correspondence with a subspace of $\mathcal{D}'(\Omega)$ and in this sense we will write

$$H^{-1}(\Omega) \subset \mathcal{D}'(\Omega).$$

Which distributions belong to $H^{-1}(\Omega)$? The following theorem gives a satisfactory answer.

Theorem 7.6. *$H^{-1}(\Omega)$ is the set of distributions of the form*

$$F = f_0 + \mathrm{div}\mathbf{f} \tag{7.34}$$

where $f_0 \in L^2(\Omega)$ and $\mathbf{f} = (f_1, ..., f_n) \in L^2(\Omega; \mathbb{R}^n)$. Moreover:

$$\|F\|_{-1} \le (1 + C_P)\left\{\|f_0\|_0 + \|\mathbf{f}\|_0\right\}. \tag{7.35}$$

Proof. Let $F \in H^{-1}(\Omega)$. From Riesz's Representation Theorem, there exists a unique $u \in H_0^1(\Omega)$ such that

$$(u, v)_1 = Fv \qquad \forall v \in H_0^1(\Omega).$$

Since

$$(u, v)_1 = (\nabla u, \nabla v) = -\langle \mathrm{div}\nabla u, v\rangle$$

in $\mathcal{D}'(\Omega)$, it follows that (7.34) holds with $f_0 = 0$ and $\mathbf{f} = -\nabla u$. Moreover, $\|u\|_1 = \|F\|_{-1}$.

Viceversa, let $F = f_0 + \operatorname{div} \mathbf{f}$, with $f_0 \in L^2(\Omega)$ and $\mathbf{f} = (f_1, ..., f_n) \in L^2(\Omega; \mathbb{R}^n)$. Then $F \in \mathcal{D}'(\Omega)$ and, letting $Fv = \langle F, v \rangle$, we have;

$$Fv = \int_\Omega f_0 v \; d\mathbf{x} + \int_\Omega \mathbf{f} \cdot \nabla v \; d\mathbf{x} \qquad \forall v \in \mathcal{D}(\Omega).$$

From the Schwarz and Poincaré inequalities, we have

$$|Fv| \leq (C_P + 1) \{\|f_0\|_0 + \|\mathbf{f}\|_0\} \|v\|_1. \tag{7.36}$$

Thus, F is continuous in the H_0^1- norm. It remains to show that F has a unique continuous extension to all $H_0^1(\Omega)$. Take $u \in H_0^1(\Omega)$ and $\{v_k\} \subset \mathcal{D}(\Omega)$ such that $\|v_k - u\|_1 \to 0$. Then, (7.36) yields

$$|Fv_k - Fv_h| \leq (1 + C_P) \{\|f_0\|_0 + \|\mathbf{f}\|_0\} \|v_k - v_h\|_1.$$

Therefore $\{Fv_k\}$ is a Cauchy sequence in $\mathbb{R}$ and converges to a limit which is independent of the sequence approximating u (why?) and which we may denote by Fu. Finally, since

$$|Fu| = \lim_{k \to \infty} |Fv_k| \quad \text{and} \quad \|u\|_1 = \lim_{k \to \infty} \|v_k\|_1,$$

from (7.36) we get:

$$|Fu| \leq (1 + C_P) \{\|f_0\|_0 + \|\mathbf{f}\|_0\} \|u\|_1$$

showing that $F \in H^{-1}(\Omega)$. $\square$

Theorem 7.6 says that the elements of $H^{-1}(\Omega)$ are represented by a linear combination of functions in $L^2(\Omega)$ and their first derivatives (in the sense of distributions). In particular, $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$.

Example 7.21. If $n = 1$, the Dirac δ belongs to $H^{-1}(-a, a)$. Indeed, we have $\delta = \mathcal{H}'$ where $\mathcal{H}$ is the Heaviside function, and $\mathcal{H} \in L^2(-a, a)$.

However, if $n \geq 2$ and $\mathbf{0} \in \Omega$, $\delta \notin H^{-1}(\Omega)$. For instance, let $n = 2$ and $\Omega = B_1(\mathbf{0})$. Assume $\delta \in H^{-1}(\Omega)$. Then we may write

$$\delta = f_0 + \operatorname{div} \mathbf{f}$$

for some $f_0 \in L^2(\Omega)$ and $\mathbf{f} \in L^2(\Omega; \mathbb{R}^2)$. Thus, for every $\varphi \in \mathcal{D}(\Omega)$,

$$\varphi(\mathbf{0}) = \langle \delta, \varphi \rangle = \langle f_0 + \operatorname{div} \mathbf{f}, \varphi \rangle = \int_\Omega [f_0 \varphi - \mathbf{f} \cdot \nabla \varphi] \, d\mathbf{x}.$$

From Schwarz's inequality, it follows that

$$|\varphi(\mathbf{0})|^2 \leq \left\{ \|f_0\|_0^2 + \|\mathbf{f}\|_0^2 \right\} \|\varphi\|_{1,2}^2$$

and, using the density of $\mathcal{D}(\Omega)$ in $H_0^1(\Omega)$, this estimate should hold for any $\varphi \in H_0^1(\Omega)$ as well. But this is impossible, since in $H_0^1(\Omega)$ there are functions which are unbounded near the origin, as we have seen in Example 7.26.

Example 7.22. Let Ω be a smooth, bounded domain in $\mathbb{R}^n$. Let $u = \chi_\Omega$ be its characteristic function. Since $\chi_\Omega \in L^2(\mathbb{R}^n)$, the distribution $\mathbf{F} = \nabla\chi_\Omega$ belongs to $H^{-1}(\mathbb{R}^n; \mathbb{R}^n)$. The support of $\mathbf{F} = \nabla\chi_\Omega$ coincides with $\partial\Omega$ and its action on a test $\boldsymbol{\varphi} \in \mathcal{D}(\mathbb{R}^n; \mathbb{R}^n)$ is described by the following formula:

$$\langle \nabla\chi_\Omega, \boldsymbol{\varphi} \rangle = -\int_{\mathbb{R}^n} \chi_\Omega \operatorname{div} \boldsymbol{\varphi} \, d\mathbf{x} = -\int_{\partial\Omega} \boldsymbol{\varphi} \cdot \boldsymbol{\nu} \, d\sigma.$$

We may regard $\mathbf{F}$ as a "delta uniformly distributed on $\partial\Omega$".

Remark 7.7. It is important to avoid confusion between $H^{-1}(\Omega)$ and $H^1(\Omega)^*$, the dual of $H^1(\Omega)$. Since, in general, $\mathcal{D}(\Omega)$ is **not dense** in $H^1(\Omega)$, the space $H^1(\Omega)^*$ is **not** a space of distributions. Indeed, although the restriction to $\mathcal{D}(\Omega)$ of every $T \in H^1(\Omega)^*$ is a distribution, this restriction does not identifies T. As a simple example, take $\mathbf{f} \in L^2(\Omega; \mathbb{R}^n)$ with $|\mathbf{f}| \geq c > 0$ a.e. and $\operatorname{div} \mathbf{f} = 0$. Define

$$T\varphi = \int_\Omega \mathbf{f} \cdot \nabla\varphi \, d\mathbf{x}.$$

Since $|T\varphi| \leq \|\mathbf{f}\|_0 \|\nabla\varphi\|_0$, we infer that $T \in H^1(\Omega)^*$. However, the restriction of T to $\mathcal{D}(\Omega)$ is the *null operator*, since in $\mathcal{D}'(\Omega)$ we have

$$\langle T, \varphi \rangle = -\langle \operatorname{div} \mathbf{f}, \varphi \rangle = 0 \qquad \forall \varphi \in \mathcal{D}(\Omega).$$

7.7.5 The spaces $H^m(\Omega)$, $m > 1$

Involving higher order derivatives, we may construct new Sobolev spaces. Let N be the number of multi-indexes $\alpha = (\alpha_1, ..., \alpha_n)$ such that $|\alpha| = \sum_{i=1}^n \alpha_i \leq m$. Choose in Theorem 7.4

$$H = L^2(\Omega), \quad Z = L^2(\Omega; \mathbb{R}^N) \subset \mathcal{D}'(\Omega; \mathbb{R}^N),$$

and $L : L^2(\Omega) \to \mathcal{D}'(\Omega; \mathbb{R}^N)$ given by

$$Lv = \{D^\alpha v\}_{|\alpha| \leq m}.$$

Then W is the **Sobolev space** of the functions in $L^2(\Omega)$, whose *derivatives (in the sense of distributions) up to order m included, are functions in $L^2(\Omega)$*. For this space we use the symbol $H^m(\Omega)$. Thus:

$$H^m(\Omega) = \{v \in L^2(\Omega) : D^\alpha v \in L^2(\Omega), \quad \forall \alpha : |\alpha| \leq m\}.$$

From Theorem 7.4 and the separability of $L^2(\Omega)$, we deduce:

Proposition 7.10. *$H^m(\Omega)$ is a separable Hilbert space, continuously embedded in $L^2(\Omega)$. The operators D^α, $|\alpha| \leq m$, are continuous from $H^m(\Omega)$ into $L^2(\Omega)$.*

The inner product and the norm in H^m are given, respectively, by

$$(u, v)_{H^m(\Omega)} = (u, v)_{m,2} = \sum_{|\alpha| \le m} \int_\Omega D^\alpha u D^\alpha v \, d\mathbf{x}$$

and

$$\|u\|_{H^m(\Omega)}^2 = \|u\|_{m,2}^2 = \sum_{|\alpha| \le m} \int_\Omega |D^\alpha u|^2 \, d\mathbf{x}.$$

If $u \in H^m(\Omega)$, any derivative of u of order k belongs to $H^{m-k}(\Omega)$; more generally, if $|\alpha| = k \le m$, then

$$D^\alpha u \in H^{m-k}(\Omega)$$

and $H^m(\Omega) \hookrightarrow H^{m-k}(\Omega)$, $k \ge 1$.

Example 7.23. Let $B_1(\mathbf{0}) \subset \mathbb{R}^3$ and consider $u(\mathbf{x}) = |\mathbf{x}|^{-a}$. It is easy to check (see Problem 7.15) that $u \in H^1(B_1(\mathbf{0}))$ if $a < 1/2$. The second order derivatives of u are given by:

$$u_{x_i x_j} = a(a+2) x_i x_j |\mathbf{x}|^{-a-4} - a\delta_{ij} |\mathbf{x}|^{-a-2}.$$

Then

$$\left| u_{x_i x_j} \right| \le |a(a+2)| \, |\mathbf{x}|^{-a-2}$$

so that $u_{x_i x_j} \in L^2(B_1(0))$ if $2a + 4 < 3$, or $a < -\frac{1}{2}$. Thus $u \in H^2(B_1(\mathbf{0}))$ if $a < -1/2$.

7.7.6 Calculus rules

Most calculus rules in H^m are formally similar to the classical ones, although their proofs are not so trivial. We list here a few of them:

Derivative of a product. Let $u \in H^1(\Omega)$ and $v \in \mathcal{D}(\Omega)$. Then $uv \in H^1(\Omega)$ and

$$\nabla(uv) = u\nabla v + v\nabla u. \tag{7.37}$$

Formula (7.37) holds if both $u, v \in H^1(\Omega)$ as well. In this case, however,

$$uv \in L^1(\Omega) \quad \text{and} \quad \nabla(uv) \in L^1(\Omega; \mathbb{R}^n).$$

Composition I. Let $u \in H^1(\Omega)$ and $g : \Omega' \to \Omega$ be one-to-one and Lipschitz. Then, the composition

$$u \circ g : \Omega' \to \mathbb{R}$$

belongs to $H^1(\Omega')$ and

$$\partial_{x_i}[u \circ g](\mathbf{x}) = \sum_{k=1}^n \partial_{x_k} u(g(\mathbf{x})) \partial_{x_i} g_k(\mathbf{x}) \tag{7.38}$$

both a.e. in Ω and in $\mathcal{D}'(\Omega)$. In particular, the Lipschitz change of variables $\mathbf{y} = g(\mathbf{x})$ transforms $H^1(\Omega)$ into $H^1(\Omega')$.

Composition *II*. Let $u \in H^1(\Omega)$ and $f : \mathbb{R} \to \mathbb{R}$ be Lipschitz. Then, the composition

$$f \circ u : \Omega \to \mathbb{R}$$

belongs to $H^1(\Omega)$ and

$$\partial_{x_i}[f \circ u](\mathbf{x}) = f'(u(\mathbf{x}))\,\partial_{x_i}u(\mathbf{x}) \tag{7.39}$$

both a.e. in Ω and in $\mathcal{D}'(\Omega)$.

In particular, choosing respectively

$$f(t) = |t|, \quad f(t) = \max\{t, 0\} \quad \text{and} \quad f(t) = -\min\{t, 0\},$$

it follows that the following functions:

$$|u|, \quad u^+ = \max\{u, 0\}, \quad \text{and} \quad u^- = -\min\{u, 0\}$$

all belong to $H^1(\Omega)$. For these functions, (7.39) yields

$$\nabla u^+ = \begin{cases} \nabla u & \text{if } u > 0 \\ 0 & \text{if } u \leq 0 \end{cases}, \quad \nabla u^- = \begin{cases} 0 & \text{if } u \geq 0 \\ -\nabla u & \text{if } u < 0 \end{cases}$$

and $\nabla(|u|) = \nabla u^+ + \nabla u^-$, $\nabla u = \nabla u^+ - \nabla u^-$. As a consequence, if $u \in H^1(\Omega)$ is constant in a set $K \subseteq \Omega$, then $\nabla u = 0$ a.e. in K.

7.7.7 Fourier Transform and Sobolev Spaces

The spaces $H^m(\mathbb{R}^n)$, $m \geq 1$, may be defined in terms of the Fourier transform. In fact, by Theorem 7.3,

$$u \in L^2(\mathbb{R}^n) \qquad \text{if and only if} \qquad \widehat{u} \in L^2(\mathbb{R}^n)$$

and

$$\|u\|^2_{L^2(\mathbb{R}^n)} = (2\pi)^{-n}\|\widehat{u}\|^2_{L^2(\mathbb{R}^n)}.$$

It follows that, for every multi-index α with $|\alpha| \leq m$,

$$D^\alpha u \in L^2(\mathbb{R}^n) \qquad \text{if and only if} \qquad \boldsymbol{\xi}^\alpha\widehat{u} \in L^2(\mathbb{R}^n)$$

and

$$\|D^\alpha u\|^2_{L^2(\mathbb{R}^n)} = (2\pi)^{-n}\|\boldsymbol{\xi}^\alpha\widehat{u}\|^2_{L^2(\mathbb{R}^n)}.$$

Finally, observe that

$$|\boldsymbol{\xi}^\alpha|^2 \leq |\boldsymbol{\xi}|^{2|\alpha|} \leq C(1 + |\boldsymbol{\xi}|^2)^m$$

whence we obtain the following result.

Proposition 7.11. *Let $u \in L^2(\mathbb{R}^n)$. Then:*

i) $u \in H^m(\mathbb{R}^n)$ *if and only if* $(1 + |\boldsymbol{\xi}|^2)^{m/2}\widehat{u} \in L^2(\mathbb{R}^n)$.

ii) *The norms*

$$\|u\|_{H^m(\mathbb{R}^n)} \qquad and \qquad \left\|(1 + |\boldsymbol{\xi}|^2)^{m/2}\widehat{u}\right\|_{L^2(\mathbb{R}^n)}$$

are equivalent.

- *Sobolev spaces of real order.* The norm

$$\|u\|_{H^m(\mathbb{R}^n)} = \left\|(1 + |\boldsymbol{\xi}|^2)^{m/2}\widehat{u}\right\|_{L^2(\mathbb{R}^n)}$$

makes perfect sense even if m is not an integer and we are led to the following definition.

Definition 7.12. *Let $s \in \mathbb{R}$, $0 < s < \infty$. We denote by $H^s(\mathbb{R}^n)$ the space of functions $u \in L^2(\mathbb{R}^n)$ such that $|\boldsymbol{\xi}|^s\,\widehat{u} \in L^2(\mathbb{R}^n)$.*

Intuitively, the function

$$\mathcal{F}^{-1}\left[\left(i\xi_j\right)^s \widehat{u}\right]$$

represents a "derivative of order s" of u. Then,

$$u \in H^s(\mathbb{R}^n)$$

if the "derivatives of order s" of u belong to $L^2(\mathbb{R}^n)$. We have:

Proposition 7.12. *$H^s(\mathbb{R}^n)$ is a Hilbert space with inner product and norm given by*

$$(u, v)_{H^s(\mathbb{R}^n)} = \int_{\mathbb{R}^n} \left(1 + |\boldsymbol{\xi}|^2\right)^s \widehat{u}\,\overline{\widehat{v}}\,d\boldsymbol{\xi}$$

and

$$\|u\|_{H^s(\mathbb{R}^n)} = \left\|\left(1 + |\boldsymbol{\xi}|^2\right)^{s/2}\widehat{u}\right\|_{L^2(\mathbb{R}^n)}.$$

The space $H^{1/2}(\mathbb{R}^n)$ of the L^2-functions possessing "half derivatives" in $L^2(\mathbb{R}^n)$ plays an important role in Section 7.9.

7.8 Approximations by Smooth Functions and Extensions

7.8.1 Local approximations

The functions in $H^1(\Omega)$ may be quite irregular. However, using mollifiers, any $u \in H^1(\Omega)$ may be approximated *locally* by smooth functions, in the sense that the approximation holds in every compact subset of Ω.

Denote by $\eta_\varepsilon = \frac{1}{\varepsilon^n}\eta\left(\frac{|\mathbf{x}|}{\varepsilon}\right)$ the mollifier introduced in section 7.2 and by Ω_ε the set of points $\varepsilon-$away from $\partial\Omega$, i.e. (see Remark 7.2):

$$\Omega_\varepsilon = \{\mathbf{x}\in\Omega\colon \text{dist}\,(\mathbf{x},\partial\Omega) > \varepsilon\}\,.$$

We have:

Theorem 7.7. *Let* $u\in H^1(\Omega)$ *and, for* $\varepsilon > 0$, *small, define*

$$u_\varepsilon = u * \eta_\varepsilon.$$

Then

  1. $u_\varepsilon\in C^\infty\left(\Omega_\varepsilon\right)$,
  2. *if* $\varepsilon\to 0$, $u_\varepsilon\to u$ *in* $H^1\left(\Omega'\right)$ *for every* $\Omega'\subset\subset\Omega$.

  Proof. Property 1 follows from Remark 7.2. To prove 2, recall that, for every $i = 1, 2, ..., n$, we have

$$\partial_{x_i}u_\varepsilon = \partial_{x_i}u * \eta_\varepsilon. \tag{7.40}$$

Then, 2 follows from property **d** of Lemma 7.1, applied to any $\Omega'\subset\subset\Omega$. $\square$

7.8.2 Estensions and global approximations

By Theorem 7.7, we may approximate a function in $H^1\left(\Omega\right)$ by smooth functions, as long as we stay at positive distance from $\partial\Omega$. We wonder whether an approximation is possible in all $\overline{\Omega}$. First we give the following definition.

Definition 7.13. *Denote by* $\mathcal{D}\left(\overline{\Omega}\right)$ *the set of restrictions to* $\overline{\Omega}$ *of functions in* $\mathcal{D}\left(\mathbb{R}^n\right)$.

  Thus, $\varphi\in\mathcal{D}\left(\overline{\Omega}\right)$ if there is $\psi\in\mathcal{D}\left(\mathbb{R}^n\right)$ such that $\varphi = \psi$ in $\overline{\Omega}$. Clearly, $\mathcal{D}\left(\overline{\Omega}\right)\subset C^\infty\left(\overline{\Omega}\right)$. We want to establish whether

$$\mathcal{D}\left(\overline{\Omega}\right) \ \textbf{is dense in}\ H^1\left(\Omega\right). \tag{7.41}$$

  The case $\Omega = \mathbb{R}^n$ is special, since $\mathcal{D}\left(\Omega\right)$ coincides with $\mathcal{D}\left(\overline{\Omega}\right)$. We have:

Theorem 7.8. $\mathcal{D}\left(\mathbb{R}^n\right)$ *is dense in* $H^1\left(\mathbb{R}^n\right)$. *In particular* $H^1\left(\mathbb{R}^n\right) = H_0^1\left(\mathbb{R}^n\right)$.

  Proof. First observe that $H_c^1\left(\mathbb{R}^n\right)$, the subspace of functions with compact (essential) support in $\mathbb{R}^n$, is dense in $H^1(\mathbb{R}^n)$. In fact, let $u\in H^1\left(\mathbb{R}^n\right)$ and $v\in\mathcal{D}\left(\mathbb{R}^n\right)$, such that $0\leq v\leq 1$ and $v\equiv 1$ if $|\mathbf{x}|\leq 1$. Define

$$u_s\left(\mathbf{x}\right) = v\left(\frac{\mathbf{x}}{s}\right)u\left(\mathbf{x}\right).$$

Then $u_s\in H_c^1\left(\mathbb{R}^n\right)$ and

$$\nabla u_s\left(\mathbf{x}\right) = v\left(\frac{\mathbf{x}}{s}\right)\nabla u\left(\mathbf{x}\right) + \frac{1}{s}u\left(\mathbf{x}\right)\nabla v\left(\frac{\mathbf{x}}{s}\right).$$

From the Dominated Convergence Theorem[19], it follows that

$$u_s \to u \qquad \text{in } H^1(\mathbb{R}^n) \qquad \text{as } s \to \infty.$$

On the other hand $\mathcal{D}(\mathbb{R}^n)$ is dense in $H^1_c(\mathbb{R}^n)$. In fact, if $u \in H^1_c(\mathbb{R}^n)$, we have

$$u_\varepsilon = u * \eta_\varepsilon \in \mathcal{D}(\mathbb{R}^n)$$

and $u_\varepsilon \to u$ in $H^1(\mathbb{R}^n)$. $\square$

However, in general (7.41) is not true, as the following example shows.

Example 7.24. Consider, for instance,

$$\Omega = \{(\rho, \theta) : 0 < \rho < 1, 0 < \theta < 2\pi\}.$$

The domain Ω coincides with the open unit circle, centered at the origin, without the radius

$$\{(\rho, \theta) : 0 < \rho < 1, \theta = 0\}.$$

The closure $\overline{\Omega}$ is given by the full closed circle. Let

$$u(\rho, \theta) = \rho^{1/2} \cos(\theta/2).$$

Then $u \in L^2(\Omega)$, since u is bounded. Moreover[20],

$$|\nabla u|^2 = u_\rho^2 + \frac{1}{\rho^2} u_\theta^2 = \frac{1}{4\rho} \qquad \text{in } \Omega,$$

so that $u \in H^1(\Omega)$. However, $u(\rho, 0+) = \rho^{1/2}$ while $u(\rho, 2\pi-) = -\rho^{1/2}$. Thus, u has a jump discontinuity across $\theta = 0$ and no sequence of smooth functions can converge to u in $H^1(\Omega)$.

The difficulty in Example 7.24 is that the domain Ω lies on both sides of part of its boundary (the radius $0 < \rho < 1, \theta = 0$). Thus, to have a hope that (7.41) is true we have to avoid domains with this anomaly and consider domains with some degree of regularity.

Thus, assume Ω is a C^1 or even a Lipschitz domain. Theorem 7.8 suggests a strategy to prove (7.41): given $u \in H^1(\Omega)$, extend the definition of u to all $\mathbb{R}^n$ in order to obtain a function in $H^1(\mathbb{R}^n)$ and then apply Theorem 7.8. The first thing to do is to introduce an *extension operator*:

Definition 7.14. *We say that a linear operator $E : H^1(\Omega) \to H^1(\mathbb{R}^n)$ is an extension operator if, $\forall u \in H^1(\Omega)$:*

1. *$E(u) = u$ in Ω,*
2. *if Ω is bounded, $E(u)$ is compactly supported,*
3. *E is continuous:*

$$\|Eu\|_{H^1(\mathbb{R}^n)} \le c(n, \Omega) \|u\|_{H^1(\Omega)}.$$

---

[19] Observe that $|u_s| \le |u|$ and $|\nabla u_s| \le |\nabla u| + M|u|$ where $M = \max |\nabla v|$.
[20] Appendix C.

How do we construct E? The first thing that comes into mind is to define $Eu = 0$ outside Ω (*trivial extension*). This certainly works if $u \in H_0^1(\Omega)$. In fact: $u \in H_0^1(\Omega)$ *if and only if its trivial extension belongs to* $H^1(\mathbb{R}^n)$.

However, the trivial extension works in this case only. For instance, let $u \in H^1(0, \infty)$ with $u(0) = a \neq 0$. Let Eu be the trivial extension of u. Then, in $\mathcal{D}'(\mathbb{R})$, $(Eu)' = u' + a\delta$ which is not even in $L^2(\mathbb{R})$.

Thus, we have to use another method. If Ω is a half space. i.e.

$$\Omega = \mathbb{R}_+^n = \{(x_1, ..., x_n) : x_n > 0\}$$

an extension operator can be defined by a reflection method as follows:

• *Reflection method.* Let $H^1(\mathbb{R}_+^n)$. Write $\mathbf{x} = (\mathbf{x}', x_n)$. We reflect in an even way with respect to the hyperplane $x_n = 0$, by setting $Eu = \tilde{u}$ where

$$\tilde{u}(\mathbf{x}) = u(\mathbf{x}', |x_n|).$$

Then, it is possible to prove that, in $\mathcal{D}'(\mathbb{R}^n)$:

$$\tilde{u}_{x_j}(\mathbf{x}) = \begin{cases} u_{x_j}(\mathbf{x}', |x_n|) & j < n \\ u_{x_n}(\mathbf{x}', |x_n|) \operatorname{sign} x_n & j = n. \end{cases} \tag{7.42}$$

It is now easy to check that E has the properties 1,2,3 listed above. In particular,

$$\|Eu\|_{H^1(\mathbb{R}^n)}^2 = 2 \|u\|_{H^1(\mathbb{R}_+^n)}^2.$$

• *Extension operator for Lipschitz domains.* Suppose now that Ω is a bounded Lipschitz domain. To construct an extension operator we use two rather general ideas, which may be applied in several different contexts: *localization* and *reduction to the half space.*

Localization. It is based on the following lemma. Given a set K, by *open covering of K* we mean a collection $\mathcal{U}$ of open sets, such that $K \subset \cup_{U \in \mathcal{U}} U$.

Lemma 7.3. (Partition of unity). *Let $K \subset \mathbb{R}^n$ be a compact set and $U_1, ..., U_N$ be an open covering of K. There exist functions $\psi_1, ..., \psi_N$ with the following properties:*

1. *For every $j = 1, ..., N$, $\psi_j \in C_0^\infty(U_j)$ and $0 \leq \psi_j \leq 1$.*

2. *For every $\mathbf{x} \in K$, $\sum_{j=1}^N \psi_j(\mathbf{x}) = 1$.*

Proof (sketch). Since $K \subset \cup_{j=1}^N U_j$ and each U_j is open, we can find open sets $K_j \subset\subset U_j$ such that

$$K \subset \cup_{j=1}^N K_j.$$

Let χ_{K_j} be the characteristic function of K_j and η_ε the mollifier (7.3). Define $\varphi_{j,\varepsilon} = \eta_\varepsilon * \chi_{K_j}$. According to Example 7.2, we may fix ε so small in order to have $\varphi_{j,\varepsilon} \in C_0^\infty(U_j)$ and $\varphi_{j,\varepsilon} > 0$ on K_j. Then the functions

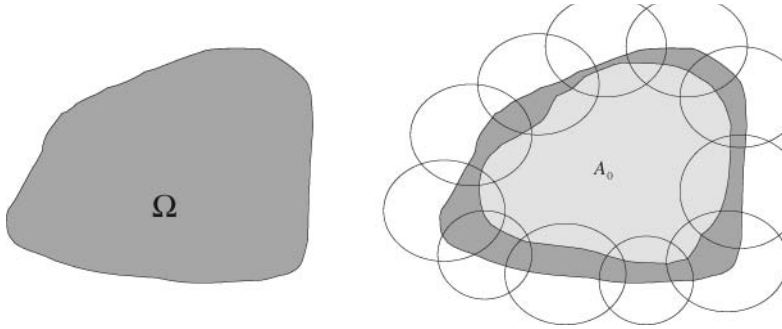$$\psi_j = \frac{\varphi_{j,\varepsilon}}{\sum_{s=1}^N \varphi_{s,\varepsilon}}$$

Fig. 7.5. A set Ω and an open covering of its closure

satisfy conditions **1** and **2**. $\square$

The set of functions $\psi_1, ..., \psi_N$ is called *a partition of unity for K, associated with the covering $U_1, ..., U_N$*. Now, if $u : K \to \mathbb{R}$, the localization procedure consists in writing

$$u = \sum_{j=1}^{N} \psi_j u \qquad (7.43)$$

i.e. as a sum of functions $u_j = \psi_j u$ supported in U_j.

Reduction to a half space. Take an open covering of $\partial\Omega$ by N balls $B_j = B(\mathbf{x}_j)$, centered at $\mathbf{x}_j \in \partial\Omega$ and such that $\partial\Omega \cap B_j$ is locally a graph of a Lipschitz function $y_n = \varphi_j(\mathbf{y}')$. This is possible, since $\partial\Omega$ is compact. Moreover, let $A_0 \subset \Omega$ be an open set containing $\Omega \backslash \cup_{j=1}^{N} B_j$ (Fig. 7.5).

Then, $A_0, B_1, ..., B_N$ is an open covering of $\overline{\Omega}$. Let $\psi_0, \psi_1, ..., \psi_N$ be a partition of unity for $\overline{\Omega}$, associated with $A_0, B_1, ..., B_N$.
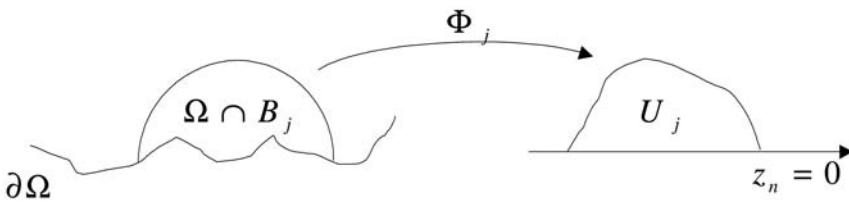
Fig. 7.6. The bi-Lipschitz transformation Φ_j *flattens* $B_j \cap \partial\Omega$

By the definition of Lipschitz domain (see Section 1.4), for each B_j, $1 \le j \le N$, there is a bi-Lipschitz transformation $\mathbf{z} = \mathbf{\Phi}_j(\mathbf{x})$ such that

$$\mathbf{\Phi}_j(B_j \cap \Omega) \equiv U_j \subset \mathbb{R}^n_+$$

and (Fig. 7.6)

$$\mathbf{\Phi}_j(B_j \cap \partial\Omega) \subset \partial\mathbb{R}^n_+ = \{z_n = 0\}.$$

Let $u \in H^1(\Omega)$ and $u_j = \psi_j u$. Then, $w_j = u_j \circ \mathbf{\Phi}_j^{-1}$ is supported in U_j, so that, extending it to zero in $\mathbb{R}^n_+ \backslash U_j$, we have $w_j \in H^1(\mathbb{R}^n_+)$.

The function $Ew_j = \tilde{w}_j$, obtained by the reflection method, belongs to $H^1(\mathbb{R}^n)$. Now we go back defining

$$Eu_j = \tilde{w}_j \circ \mathbf{\Phi}_j, \qquad 1 \leq j \leq N,$$

in B_j and $Eu_j = 0$ outside B_j. Finally, let $u_0 = \psi_0 u_0$ and let Eu_0 be the trivial extension of u_0. Set

$$Eu = \sum_{j=0}^{N} Eu_j.$$

At this point, it is not difficult to show that E satisfies the requirements 1, 2, 3 of Definition 7.14. We have proved the following

Theorem 7.9. *Let Ω be either $\mathbb{R}^n_+$ or a bounded, Lipschitz domain. Then, there exists an extension operator $E : H^1(\Omega) \to H^1(\mathbb{R}^n)$.*

An immediate consequence of Theorems 7.8 and 7.9 is the following global approximation result:

Theorem 7.10. *Let Ω be either $\mathbb{R}^n_+$ or a bounded, Lipschitz domain. Then $\mathcal{D}(\overline{\Omega})$* **is dense in** *$H^1(\Omega)$. In other words, if $u \in H^1(\Omega)$, there exists a sequence $\{u_m\} \subset \mathcal{D}(\overline{\Omega})$ such that*

$$\|u_m - u\|_{1,2} \to 0 \quad as \ m \to +\infty.$$

7.9 Traces

7.9.1 Traces of functions in $H^1(\Omega)$

The possibility of approximating any element $u \in H^1(\Omega)$ by smooth functions in $\overline{\Omega}$ represents a key tool for introducing the notion of *restriction of u on $\Gamma = \partial\Omega$*. Such restriction is called the **trace of** u on Γ and it will be an element of $L^2(\Gamma)$.

Observe that if $\Omega = \mathbb{R}^n_+$, then $\Gamma = \partial\mathbb{R}^n_+ = \mathbb{R}^{n-1}$ and $L^2(\Gamma)$ is well defined. If Ω is a Lipschitz domain, we define $L^2(\Gamma)$ by localization. More precisely, let $B_1, ..., B_N$ be an open covering of Γ by balls centered at points on Γ, as in subsection 7.8.2. If $g : \Gamma \to \mathbb{R}$, write

$$g = \sum_{j=1}^{N} \psi_j g$$

where $\psi_1, ..., \psi_N$ is a partition of unity for Γ, associated with $B_1, ..., B_N$. Since $\Gamma \cap B_j$ is the graph of a Lipschitz function $y_n = \varphi_j(\mathbf{y}')$, on $\Gamma \cap B_j$ there is a natural notion of "area element", given by

$$d\sigma = \sqrt{1 + |\nabla\varphi_j|^2} d\mathbf{y}'.$$

Thus, we say that $g \in L^2(\Gamma)$ if[21]

$$\|g\|_{L^2(\Gamma)}^2 = \sum_{j=1}^{N} \int_{\Gamma \cap B_j} \psi_j \, |g|^2 \, d\sigma < \infty. \tag{7.44}$$

$L^2(\Gamma)$ is a Hilbert space with respect to the inner product

$$(g, h)_{L^2(\Gamma)} = \sum_{j=1}^{N} \int_{\Gamma \cap B_j} \psi_j \, gh \, d\sigma.$$

Let us go back to our trace problem. We may consider $n > 1$, since there is no problem if $n = 1$. The strategy consists in the following two steps.

Let $\boldsymbol{\tau}_0 : \mathcal{D}(\overline{\Omega}) \to L^2(\Gamma)$ be the operator that associates to every function v its restriction $v_{|\Gamma}$ to Γ: $\boldsymbol{\tau}_0 v = v_{|\Gamma}$. This makes perfect sense, since each $v \in \mathcal{D}(\overline{\Omega})$ is continuous on Γ.

First step: show that $\|\boldsymbol{\tau}_0 u\|_{L^2(\Gamma)} \le c(\Omega, n) \|u\|_{1,2}$. Thus, $\boldsymbol{\tau}_0$ is continuous from $\mathcal{D}(\overline{\Omega}) \subset H^1(\Omega)$ into $L^2(\Gamma)$.

Second step: extend $\boldsymbol{\tau}_0$ to all $H^1(\Omega)$ using the density of $\mathcal{D}(\overline{\Omega})$ in $H^1(\Omega)$.

An elementary analogy may be useful. Suppose we have a function $f : \mathbb{Q} \to \mathbb{R}$ and we want to define the value of f at an irrational point x. What do we do? Since $\mathbb{Q}$ is dense in $\mathbb{R}$, we select a sequence $\{r_k\} \subset \mathbb{Q}$ such that $r_k \to x$. Then we compute $f(r_k)$ and set $f(x) = \lim_{k \to \infty} f(r_k)$. Of course, we have to prove that the limit exists, by showing, for example, that $\{f(r_n)\}$ is a Cauchy sequence and that the limit does not depend on the approximating sequence $\{r_n\}$.

Theorem 7.11. *Let Ω be either $\mathbb{R}_+^n$ or a bounded, Lipschitz domain. Then there exists a linear operator (trace operator) $\boldsymbol{\tau}_0 : H^1(\Omega) \to L^2(\Gamma)$ such that:*

1. *$\boldsymbol{\tau}_0 u = u_{|\Gamma}$ if $u \in \mathcal{D}(\overline{\Omega})$,*
2. *$\|\boldsymbol{\tau}_0 u\|_{L^2(\Gamma)} \le c(\Omega, n) \|u\|_{1,2}$.*

Proof. Let $\Omega = \mathbb{R}_+^n$. First, we prove inequality 2 for $u \in \mathcal{D}(\overline{\Omega})$. In this case $\boldsymbol{\tau}_0 u = u(\mathbf{x}', 0)$ and we must show that there is a constant c such that

$$\int_{\mathbb{R}^{n-1}} |u(\mathbf{x}', 0)|^2 \, d\mathbf{x}' \le c \|u\|_{H^1(\mathbb{R}_+^n)}^2 \qquad \forall u \in \mathcal{D}(\overline{\Omega}). \tag{7.45}$$

For every $x_n \in (0, 1)$ we may write:

$$u(\mathbf{x}', 0) = u(\mathbf{x}', x_n) - \int_0^{x_n} u_{x_n}(\mathbf{x}', t) \, dt.$$

Since by Schwarz's inequality

$$\left(\int_0^1 |u_{x_n}(\mathbf{x}', t)| \, dt \right)^2 \le \int_0^1 |u_{x_n}(\mathbf{x}', t)|^2 \, dt,$$

---

[21] Observe that the norm (7.44) depends on the particular covering and partition of unity. However, norms corresponding to different coverings and partitions of unity are all equivalent and induce the same topology on $L^2(\Gamma)$.

we deduce that (recalling the elementary inequality $(a+b)^2 \le 2a^2 + 2b^2$)

$$|u(\mathbf{x}',0)|^2 \le 2\,|u(\mathbf{x}',x_n)|^2 + 2\left(\int_0^1 |u_{x_n}(\mathbf{x}',t)|\,dt\right)^2$$

$$\le 2\,|u(\mathbf{x}',x_n)|^2 + 2\int_0^1 |u_{x_n}(\mathbf{x}',t)|^2\,dt$$

$$\le 2\,|u(\mathbf{x}',x_n)|^2 + 2\int_0^1 |\nabla u(\mathbf{x}',t)|^2\,dt$$

Integrating both sides in $\mathbb{R}^{n-1}$ with respect to $\mathbf{x}'$ and in $(0,1)$ with respect to x_n we easily obtain (7.45) with $c=2$.

Assume now $u \in H^1\left(\mathbb{R}^n_+\right)$. Since $\mathcal{D}\left(\overline{\Omega}\right)$ is dense in $H^1\left(\mathbb{R}^n_+\right)$, we can select $\{u_k\} \subset \mathcal{D}\left(\overline{\Omega}\right)$ such that $u_k \to u$ in $H^1\left(\mathbb{R}^n_+\right)$.

The linearity of $\boldsymbol{\tau}_0$ and estimate (7.45) yield

$$\|\boldsymbol{\tau}_0 u_h - \boldsymbol{\tau}_0 u_k\|_{L^2(\mathbb{R}^{n-1})} \le \sqrt{2}\,\|u_h - u_k\|_{H^1\left(\mathbb{R}^n_+\right)}.$$

Since $\{u_k\}$ is a Cauchy sequence in $H^1\left(\mathbb{R}^n_+\right)$, we infer that $\{\boldsymbol{\tau}_0 u_k\}$ is a Cauchy sequence in $L^2\left(\mathbb{R}^{n-1}\right)$. Therefore, there exists $u_0 \in L^2(\mathbb{R}^{n-1})$ such that

$$\boldsymbol{\tau}_0 u_k \to u_0 \qquad \text{in } L^2(\mathbb{R}^{n-1}).$$

The limiting element u_0 does not depend on the approximating sequence $\{u_k\}$. In fact, if $\{v_k\} \subset \mathcal{D}\left(\overline{\Omega}\right)$ and $v_k \to u$ in $H^1\left(\mathbb{R}^n_+\right)$, then

$$\|v_k - u_k\|_{H^1\left(\mathbb{R}^n_+\right)} \to 0.$$

From

$$\|\boldsymbol{\tau}_0 v_k - \boldsymbol{\tau}_0 u_k\|_{L^2(\mathbb{R}^{n-1})} \le \sqrt{2}\,\|v_k - u_k\|_{H^1\left(\mathbb{R}^n_+\right)}$$

it follows that $\boldsymbol{\tau}_0 v_k \to u_0$ in $L^2(\mathbb{R}^{n-1})$ as well.

Thus, if $u \in H^1\left(\mathbb{R}^n_+\right)$, it makes sense to define $\boldsymbol{\tau}_0 u = u_0$. It should be clear that $\boldsymbol{\tau}_0$ has the properties 1, 2.

If Ω is a bounded Lipschitz domain, the theorem can be proved once more by localization and reduction to a half space. We omit the details. $\square$

Definition 7.15. *The function $\boldsymbol{\tau}_0 u$, also denoted by $u_{|\Gamma}$, is called the* trace of u on Γ.

The following integration by parts formula for functions in $H^1\left(\Omega\right)$ is a consequence of the trace theorem 7.11.

Corollary 7.1. *Assume Ω is either $\mathbb{R}^n_+$ or a bounded, Lipschitz domain. Let $u \in H^1\left(\Omega\right)$ and $\mathbf{v} \in H^1\left(\Omega; \mathbb{R}^n\right)$. Then*

$$\int_\Omega \nabla u \cdot \mathbf{v}\,d\mathbf{x} = -\int_\Omega u\,\mathrm{div}\,\mathbf{v}\,d\mathbf{x} + \int_\Gamma (\boldsymbol{\tau}_0 u)\,(\boldsymbol{\tau}_0 \mathbf{v}) \cdot \boldsymbol{\nu}\,d\boldsymbol{\sigma}. \qquad (7.46)$$

where $\boldsymbol{\nu}$ is the outward unit normal to Γ and $\boldsymbol{\tau}_0 \mathbf{v} = (\boldsymbol{\tau}_0 v_1, ..., \boldsymbol{\tau}_0 v_n)$.

Proof. Formula (7.46) holds if $u \in \mathcal{D}\left(\overline{\Omega}\right)$ and $\mathbf{v} \in \mathcal{D}\left(\overline{\Omega}; \mathbb{R}^n\right)$. Let $u \in H^1(\Omega)$ and $\mathbf{v} \in H^1(\Omega; \mathbb{R}^n)$. Select $\{u_k\} \subset \mathcal{D}\left(\overline{\Omega}\right)$, $\{\mathbf{v}_k\} \subset \mathcal{D}\left(\overline{\Omega}; \mathbb{R}^n\right)$ such that $u_k \to u$ in $H^1(\Omega)$ and $\mathbf{v}_k \to \mathbf{v}$ in $H^1(\Omega; \mathbb{R}^n)$. Then:

$$\int_\Omega \nabla u_k \cdot \mathbf{v}_k \, d\mathbf{x} = -\int_\Omega u_k \operatorname{div} \mathbf{v}_k \, d\mathbf{x} + \int_\Gamma (\boldsymbol{\tau}_0 u_k)(\boldsymbol{\tau}_0 \mathbf{v}_k) \cdot \boldsymbol{\nu} \, d\sigma.$$

Letting $k \to \infty$, by the continuity of $\boldsymbol{\tau}_0$, we obtain (7.46). $\square$

It is not surprising that the kernel of $\boldsymbol{\tau}_0$ is precisely[22] $H_0^1(\Omega)$:

$$\boldsymbol{\tau}_0 u = 0 \iff u \in H_0^1(\Omega).$$

In similar way, we may define the trace of $u \in H^1(\Omega)$ on a relatively open subset $\Gamma_0 \subset \Gamma$.

Theorem 7.12. *Assume Ω is either $\mathbb{R}_+^n$ or a bounded, Lipschitz domain. Let Γ_0 be an open subset of Γ.*

Then there exists a trace operator $\boldsymbol{\tau}_{\Gamma_0} : H^1(\Omega) \to L^2(\Gamma_0)$ such that:

1. $\boldsymbol{\tau}_{\Gamma_0} u = u_{|\Gamma_0}$ if $u \in \mathcal{D}\left(\overline{\Omega}\right)$,

2. $\|\boldsymbol{\tau}_{\Gamma_0} u\|_{L^2(\Gamma_0)} \leq c(\Omega, n) \|u\|_{1,2}$.

The function $\boldsymbol{\tau}_{\Gamma_0} u$ is called the *trace of u on Γ_0*, often denoted by $u_{|\Gamma_0}$. The kernel of $\boldsymbol{\tau}_{\Gamma_0}$ is denoted by $H_{0,\Gamma_0}^1(\Omega)$:

$$\boldsymbol{\tau}_{\Gamma_0} u = 0 \iff u \in H_{0,\Gamma_0}^1(\Omega).$$

This space can be characterized in another way. Let V_{0,Γ_0} be the set of functions in $\mathcal{D}\left(\overline{\Omega}\right)$ vanishing in a neighborhood of $\overline{\Gamma}_0$. Then:

Proposition 7.13. *$H_{0,\Gamma_0}^1(\Omega)$ is the closure of V_{0,Γ_0} in $H^1(\Omega)$.*

7.9.2 Traces of functions in $H^m(\Omega)$

We have seen that $u \in H^m(\mathbb{R}_+^n)$, $m \geq 1$, has a trace on $\Gamma = \partial \mathbb{R}_+^n$. However, if $m = 2$, every derivative of u belongs to $H^1(\mathbb{R}_+^n)$, so that it has a trace on Γ. In particular, we may define the trace of $\partial_{x_n} u$ on Γ. Let

$$\boldsymbol{\tau}_1 u = (\partial_{x_n} u)_{|\Gamma}.$$

In general, for $m \geq 2$, we may define the trace on Γ of the derivatives $\partial_{x_n}^j u = \frac{\partial^j u}{\partial x_n^j}$ for $j = 0, 1, ..., m-1$ and set

$$\boldsymbol{\tau}_j u = (\partial_{x_n}^j u)_{|\Gamma}.$$

In this way, we construct a linear operator $\boldsymbol{\tau} : H^m(\mathbb{R}_+^n) \to L^2(\Gamma; \mathbb{R}^m)$, given by

$$\boldsymbol{\tau} u = (\boldsymbol{\tau}_0 u, ..., \boldsymbol{\tau}_{m-1} u).$$

---

[22] However, only the proof of the " $\Longleftarrow$ " part is trivial. The proof of the " $\Longrightarrow$ " part is rather tecnical and we omit it.

From Theorem 7.11, $\boldsymbol{\tau}$ satisfies the following conditions:

1. $\boldsymbol{\tau}u = (u_{|\Gamma}\ ,(\partial_{x_n}u)_{|\Gamma}\ ,...,(\partial_{x_n}^{m-1}u)_{|\Gamma}\ )$, if $u \in \mathcal{D}\left(\overline{\mathbb{R}_+^n}\right)$,

2. $\left\|\boldsymbol{\tau}u\right\|_{L^2(\Gamma;\mathbb{R}^n)} \leq c\left\|u\right\|_{H^m\left(\mathbb{R}_+^n\right)}$.

The operator $\boldsymbol{\tau}$ associates to $u \in H^m\left(\mathbb{R}_+^n\right)$ the trace on Γ of u and its derivatives up to the order $m-1$, in the direction x_n. This direction corresponds to the *interior normal* to $\Gamma = \partial\mathbb{R}_+^n$.

Analogously, for a bounded domain Ω we may define the trace on Γ of the (interior or exterior) normal derivatives of u, up to order $m-1$. This requires Ω to be at least a C^m-domain. The following theorem holds, where $\boldsymbol{\nu}$ denotes the exterior unit normal to $\partial\Omega$.

Theorem 7.13. *Assume Ω is either $\mathbb{R}_+^n$ or a bounded, C^m-domain, $m \geq 2$. Then there exists a trace operator $\boldsymbol{\tau} : H^m(\Omega) \to L^2\left(\Gamma;\mathbb{R}^m\right)$ such that:*

1. $\boldsymbol{\tau}u = (u_{|\Gamma}\ ,\frac{\partial u}{\partial\nu}_{|\Gamma}\ ,...,\frac{\partial^{m-1}u}{\partial\nu^{m-1}}_{|\Gamma}\ )$ if $u \in \mathcal{D}\left(\overline{\Omega}\right)$,

2. $\left\|\boldsymbol{\tau}u\right\|_{L^2(\Gamma;\mathbb{R}^m)} \leq c\left(\Omega,n\right)\left\|u\right\|_{H^m(\Omega)}$.

Similarly, we may define a trace of the (interior or exterior) normal derivatives of u, up to order $m-1$, on an open subset $\Gamma_0 \subset \Gamma$.

It turns out that the kernel of the operator $\boldsymbol{\tau}$ is given by the *closure of $\mathcal{D}(\Omega)$ in $H^m(\Omega)$*, denoted by $H_0^m(\Omega)$. Precisely:

$$\boldsymbol{\tau}u = (0,...,0) \quad \Longleftrightarrow \quad u \in H_0^m(\Omega).$$

Clearly, $H_0^m(\Omega)$ is a Hilbert subspace of $H^m(\Omega)$. If $u \in H_0^m(\Omega)$, u and its normal derivatives up to order $m-1$ have zero trace on Γ.

7.9.3 Trace spaces

The operator $\boldsymbol{\tau}_0 : H^1(\Omega) \to L^2(\Gamma)$ is **not** surjective. In fact the image of $\boldsymbol{\tau}_0$ is *strictly contained in $L^2(\Gamma)$*. In other words, there are functions in $L^2(\Gamma)$ which are *not* traces of functions in $H^1(\Omega)$. So, the natural question is: which functions $g \in L^2(\Gamma)$ are traces of functions in $H^1(\Omega)$? The answer is not elementary: roughly speaking, we could characterize them as *functions possessing half derivatives in $L^2(\Gamma)$*. It is as if in the restriction to the boundary, a function of $H^1(\Omega)$ loses "one half of each derivative". To give an idea of what this means, let us consider the case $\Omega = \mathbb{R}_+^n$. We have:

Theorem 7.14. *Let $u \in H^1\left(\mathbb{R}_+^n\right)$. Then $\operatorname{Im}\boldsymbol{\tau}_0 = H^{1/2}\left(\mathbb{R}^{n-1}\right)$.*

Proof (sketch). First we show that $\operatorname{Im}\boldsymbol{\tau}_0 \subseteq H^{1/2}\left(\mathbb{R}^{n-1}\right)$. Let $u \in H^1\left(\mathbb{R}_+^n\right)$ and extend it to all $\mathbb{R}^n$ by even reflection with respect to the plane $x_n = 0$. We write

the points in $\mathbb{R}^n$ as $\mathbf{x} = (\mathbf{x}', x_n)$, with $\mathbf{x}' = (x_1, ..., x_n)$. Define $g(\mathbf{x}') = u(\mathbf{x}', 0)$. We show that $g \in H^{1/2}(\mathbb{R}^{n-1})$, that is

$$\|g\|^2_{H^{1/2}(\mathbb{R}^{n-1})} = \int_{\mathbb{R}^{n-1}} (1 + |\boldsymbol{\xi}'|^2)^{1/2} |\widehat{g}(\boldsymbol{\xi}')|^2 d\boldsymbol{\xi}' < \infty.$$

First, we consider $u \in \mathcal{D}(\mathbb{R}^n)$ and express $\widehat{g}$ in terms of $\widehat{u}$. By the Fourier inversion formula, we may write

$$u(\mathbf{x}', x_n) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^{n-1}} e^{i\mathbf{x}' \cdot \boldsymbol{\xi}'} \left(\int_{\mathbb{R}} \widehat{u}(\boldsymbol{\xi}', \xi_n) e^{ix_n \xi_n} d\xi_n \right) d\boldsymbol{\xi}'$$

so that

$$g(\mathbf{x}') = \frac{1}{(2\pi)^{n-1}} \int_{\mathbb{R}^{n-1}} e^{i\mathbf{x}' \cdot \boldsymbol{\xi}'} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \widehat{u}(\boldsymbol{\xi}', \xi_n) d\xi_n \right) d\boldsymbol{\xi}'.$$

This shows that

$$\widehat{g}(\boldsymbol{\xi}') = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{u}(\boldsymbol{\xi}', \xi_n) d\xi_n.$$

Thus:

$$\|g\|^2_{H^{1/2}(\mathbb{R}^{n-1})} = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^{n-1}} (1 + |\boldsymbol{\xi}'|^2)^{1/2} \left| \int_{\mathbb{R}} \widehat{u}(\boldsymbol{\xi}', \xi_n) d\xi_n \right|^2 d\boldsymbol{\xi}'.$$

Note now the following two facts. First, from Schwarz's inequality, we may write

$$\left| \int_{\mathbb{R}} \widehat{u}(\boldsymbol{\xi}', \xi_n) d\xi_n \right| \leq \int_{\mathbb{R}} (1 + |\boldsymbol{\xi}|^2)^{-1/2}(1 + |\boldsymbol{\xi}|^2)^{1/2} |\widehat{u}(\boldsymbol{\xi}', \xi_n)| d\xi_n$$

$$\leq \left(\int_{\mathbb{R}} (1 + |\boldsymbol{\xi}|^2) |\widehat{u}(\boldsymbol{\xi}', \xi_n)|^2 d\xi_n \right)^{1/2} \left(\int_{\mathbb{R}} (1 + |\boldsymbol{\xi}|^2)^{-1} d\xi_n \right)^{1/2}.$$

Second,[23]

$$\int_{\mathbb{R}} (1 + |\boldsymbol{\xi}|^2)^{-1} d\xi_n = \int_{\mathbb{R}} (1 + |\boldsymbol{\xi}'|^2 + \xi_n^2)^{-1} d\xi_n = \frac{\pi}{(1 + |\boldsymbol{\xi}'|^2)^{1/2}}.$$

Thus,

$$\|g\|^2_{H^{1/2}(\mathbb{R}^{n-1})} \leq \frac{1}{4\pi} \int_{\mathbb{R}^n} (1 + |\boldsymbol{\xi}|^2) |\widehat{u}(\boldsymbol{\xi})|^2 d\boldsymbol{\xi} = \frac{1}{4\pi} \|u\|^2_{H^1(\mathbb{R}^n)} = \frac{1}{2\pi} \|u\|^2_{H^1(\mathbb{R}^n_+)} < \infty.$$

Therefore, $g \in H^{1/2}(\mathbb{R}^{n-1})$. By the usual density argument, this is true for every $u \in H^1(\mathbb{R}^n)$ and shows that $\text{Im}\,\tau_0 \subseteq H^{1/2}(\mathbb{R}^{n-1})$.

To prove the opposite inclusion, take any $g \in H^{1/2}(\mathbb{R}^{n-1})$ and define

$$u(\mathbf{x}', x_n) = \frac{1}{(2\pi)^{n-1}} \int_{\mathbb{R}^{n-1}} e^{-(1+|\boldsymbol{\xi}'|)x_n} \widehat{g}(\boldsymbol{\xi}') e^{i\mathbf{x}' \cdot \boldsymbol{\xi}'} d\boldsymbol{\xi}', \qquad x_n \geq 0.$$

---

[23] $\int_{\mathbb{R}} (a^2 + t^2)^{-1} dt = \left[\frac{1}{a} \arctan\left(\frac{t}{a} \right) \right]_{-\infty}^{+\infty} = \frac{\pi}{a} \quad (a > 0).$

Then, clearly $u(\mathbf{x}', 0) = g(\mathbf{x}')$ and it can be proved that $u \in H^1(\mathbb{R}_+^n)$. Therefore, $g \in \operatorname{Im} \boldsymbol{\tau}_0$ so that $H^{1/2}(\mathbb{R}^{n-1}) \subseteq \operatorname{Im} \boldsymbol{\tau}_0$. $\square$

If Ω is a bounded, Lipschitz domain, it is possible to define $H^{1/2}(\Gamma)$ by localization and reduction to the half space, as we did for $L^2(\Gamma)$. In this way we can endow $H^{1/2}(\Gamma)$ with an inner product that *makes it a Hilbert space, continuously embedded in* $L^2(\Gamma)$. It turn out that $H^{1/2}(\Gamma)$ coincides with $\operatorname{Im} \boldsymbol{\tau}_0$:

$$H^{1/2}(\Gamma) = \left\{ u_{|\Gamma} : u \in H^1(\Omega) \right\}. \tag{7.47}$$

Actually, changing slightly our point of view, we could take (7.47) as a definition of $H^{1/2}(\Gamma)$ and endow $H^{1/2}(\Gamma)$ with the equivalent norm

$$\|g\|_{H^{1/2}(\Gamma)} = \inf \left\{ \|u\|_{H^1(\Omega)} : u \in H^1(\Omega), \ u_{|\Gamma} = g \right\}. \tag{7.48}$$

This norm is equal to the smallest among the norms of all elements in $H^1(\Omega)$ sharing the same trace g on Γ and takes into account that the trace operator $\boldsymbol{\tau}_0$ is not injective, since we know that $\operatorname{Ker} \boldsymbol{\tau}_0 = H_0^1(\Omega)$. In particular, the following *trace inequality holds:*

$$\left\| u_{|\Gamma} \right\|_{H^{1/2}(\Gamma)} \leq \|u\|_{1,2}, \tag{7.49}$$

which means that *the trace operator* $\boldsymbol{\tau}_0$ *is continuous from* $H^1(\Omega)$ *onto* $H^{1/2}(\Gamma)$.

In similar way, if Γ_0 is a relatively open subset of Γ, by localization and reduction to the half space, we may define the space $H^{1/2}(\Gamma_0)$ and *make it a Hilbert space, continuously embedded in* $L^2(\Gamma_0)$. $H^{1/2}(\Gamma_0)$ coincides with $\operatorname{Im} \boldsymbol{\tau}_{\Gamma_0}$, that is

$$H^{1/2}(\Gamma_0) = \left\{ u_{|\Gamma_0} : u \in H^1(\Omega) \right\},$$

and can be endowed with the norm

$$\|g\|_{H^{1/2}(\Gamma_0)} = \inf \left\{ \|u\|_{H^1(\Omega)} : u \in H^1(\Omega), \ u_{|\Gamma_0} = g \right\}.$$

In particular, the following *trace inequality holds:*

$$\left\| u_{|\Gamma_0} \right\|_{H^{1/2}(\Gamma_0)} \leq \|u\|_{H^1(\Omega)}. \tag{7.50}$$

which means that *the trace operator* $\boldsymbol{\tau}_{\Gamma_0}$ *is continuous from* $H^1(\Omega)$ *in* $H^{1/2}(\Gamma_0)$.

Finally, if Ω is $\mathbb{R}_+^n$ or a bounded C^m- domain, $m \geq 2$, the space of the traces of functions in $H^m(\Omega)$ is the fractional order Sobolev space $H^{m-1/2}(\Gamma)$, still showing a loss of "half derivative". Coherently, the trace of a normal derivative undergoes a loss of one more derivative and belongs to $H^{m-3/2}(\Gamma)$; the derivatives of order $m-1$ have traces in $H^{1/2}(\Gamma)$. Thus we obtain:

$$\boldsymbol{\tau} : H^m(\Omega) \to \left(H^{m-1/2}(\Gamma), H^{m-3/2}(\Gamma), ..., H^{1/2}(\Gamma) \right).$$

The kernel of $\boldsymbol{\tau}$ is $H_0^m(\Omega)$.

7.10 Compactness and Embeddings

7.10.1 Rellich's theorem

Since

$$\|u\|_0 \le \|u\|_{1,2},$$

$H^1(\Omega)$ is *continuously embedded* in $L^2(\Omega)$ i.e., if a sequence $\{u_k\}$ converges to u in $H^1(\Omega)$ it converges to u in $L^2(\Omega)$ as well.

If we assume that Ω is a **bounded, Lipschitz domain,** then the embedding of $H^1(\Omega)$ in $L^2(\Omega)$ is also **compact**. Thus, a bounded sequence $\{u_k\} \subset H^1(\Omega)$ has the following important property:

There exists a subsequence $\{u_{k_j}\}$ and $u \in H^1(\Omega)$, such that

a. $u_{k_j} \to u$ in $L^2(\Omega)$,

b. $u_{k_j} \rightharpoonup u$ in $H^1(\Omega)$ (i.e. u_{k_j} converges weakly[24] to u in $H^1(\Omega)$).

Actually, only property **a** follows from the compactness of the embedding. Property **b** expresses a general fact in every Hilbert space H: every bounded subset $E \subset H$ is *sequentially weakly compact* (Theorem 6.11).

Theorem 7.15. *Let Ω be a bounded, Lipschitz domain. Then $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$.*

Proof. We use the compactness criterion expressed in Theorem 6.9. First, observe that, for every $v \in \mathcal{D}(\mathbb{R}^n)$ we may write

$$v(\mathbf{x}+\mathbf{h}) - v(\mathbf{x}) = \int_0^1 \frac{d}{dt} v(\mathbf{x}+t\mathbf{h})\, dt = \int_0^1 \nabla v(\mathbf{x}+t\mathbf{h}) \cdot \mathbf{h}\, dt$$

whence

$$|v(\mathbf{x}+\mathbf{h}) - v(\mathbf{x})|^2 = \left| \int_0^1 \nabla v(\mathbf{x}+t\mathbf{h}) \cdot \mathbf{h}\, dt \right|^2 \le |\mathbf{h}|^2 \int_0^1 |\nabla v(\mathbf{x}+t\mathbf{h})|^2\, dt.$$

Integrating on $\mathbb{R}^n$ we find

$$\int_{\mathbb{R}^n} |v(\mathbf{x}+\mathbf{h}) - v(\mathbf{x})|^2\, d\mathbf{x} \le |\mathbf{h}|^2 \int_{\mathbb{R}^n} d\mathbf{x} \int_0^1 |\nabla v(\mathbf{x}+t\mathbf{h})|^2\, dt \le |\mathbf{h}|^2 \|\nabla v\|_{L^2(\mathbb{R}^n)}^2$$

so that

$$\int_{\mathbb{R}^n} |v(\mathbf{x}+\mathbf{h}) - v(\mathbf{x})|^2\, d\mathbf{x} \le |\mathbf{h}|^2 \|\nabla v\|_{L^2(\mathbb{R}^n)}^2. \tag{7.51}$$

Since $\mathcal{D}(\mathbb{R}^n)$ is dense in $H^1(\mathbb{R}^n)$, we infer that (7.51) holds for every $u \in H^1(\mathbb{R}^n)$ as well.

---

[24] Section 6.7.

Let now $S \subset H^1(\Omega)$ be bounded, i.e. there exists a number M such that:

$$\|u\|_{H^1(\Omega)} \leq M, \qquad \forall u \in S.$$

By Theorem 7.9, every $u \in S$ has an extension $\widetilde{u} \in H^1(\mathbb{R}^n)$, with support cotained in an open set $\Omega' \supset\supset \Omega$. Thus, $\widetilde{u} \in H_0^1(\Omega')$ and moreover,

$$\|\nabla\widetilde{u}\|_{L^2(\Omega')} \leq c \|\nabla u\|_{L^2(\Omega)} \leq cM.$$

Denote by $\widetilde{S}$ the set of such extensions. Then (7.51) holds for every $\widetilde{u} \in \widetilde{S}$:

$$\int_{\Omega'} |\widetilde{u}(\mathbf{x}+\mathbf{h}) - \widetilde{u}(\mathbf{x})|^2 \, d\mathbf{x} \leq |\mathbf{h}|^2 \|\nabla\widetilde{u}\|_{L^2(\mathbb{R}^n)}^2 \leq c^2 M^2 |\mathbf{h}|^2.$$

Theorem 6.9 implies that $\widetilde{S}$ is precompact in $L^2(\Omega')$, which implies that S is precompact in $L^2(\Omega)$. $\square$

7.10.2 Poincaré's inequalities

Under suitable hypotheses, the norm $\|u\|_{1,2}$ is equivalent to $\|\nabla u\|_0$. This means that there exists a constant C_P, depending only on n and Ω, such that

$$\|u\|_0 \leq C_P \|\nabla u\|_0. \tag{7.52}$$

Inequalities like (7.52) are called **Poincaré's inequalities** and play a big role in the variational treatment of boundary value problems, as we shall realize in the next chapter. We have already proved (Theorem 7.5) that (7.52) holds if $u \in H_0^1(\Omega)$, i.e. for functions vanishing on $\partial\Omega$.

On the other hand, (7.52) cannot hold if $u = \text{constant} \neq 0$. Roughly speaking, the hypotheses that guarantee the validity of (7.52) require that u vanishes in some "non trivial set". For instance, under each one of the following conditions, (7.52) holds:

i) $u \in H_{0,\Gamma_0}^1(\Omega)$ (u vanishes on a non empty relatively open subset $\Gamma_0 \subset \partial\Omega$);

ii) $u \in H^1(\Omega)$ and $u = 0$ on a set $E \subset \Omega$ with positive measure: $|E| = a > 0$;

iii) $u \in H^1(\Omega)$ and $\int_\Omega u = 0$ (u has mean value zero in Ω).

Theorem 7.16. *Let Ω be a bounded Lipschitz domain. Assume that u satisfies one among the hypotheses $i), ii), iii)$ above. Then, there exists C_P such that*

$$\|u\|_0 \leq C_P \|\nabla u\|_0. \tag{7.53}$$

Proof. Assume $i)$ holds. By contradiction suppose (7.53) is not true. This means that for every integer $j \geq 1$, there exists $u_j \in H_{0,\Gamma_0}^1(\Omega)$ such that

$$\|u_j\|_0 > j \|\nabla u_j\|_0. \tag{7.54}$$

Normalize u_j in $L^2(\Omega)$ by setting

$$w_j = \frac{u_j}{\|u_j\|_0}.$$

Then, from (7.54),

$$\|w_j\|_0 = 1 \quad \text{and} \quad \|\nabla w_j\|_0 < \frac{1}{j} \leq 1.$$

Thus $\{w_j\}$ is bounded in $H^1(\Omega)$ and by Rellich's Theorem there exists a subsequence $\{w_{j_k}\}$ and $w \in H^1_{0,\Gamma_0}(\Omega)$ such that

- $w_{j_k} \to w$ in $L^2(\Omega)$,
- $\nabla w_{j_k} \rightharpoonup \nabla w$ in $L^2(\Omega)$.

The continuity of the norm gives

$$\|w\|_0 = \lim_{j \to \infty} \|w_j\|_0 = 1.$$

On the other hand, the weak semicontinuity of the norm (Theorem 6.10) yields,

$$\|\nabla w\|_0 \leq \liminf_{j \to \infty} \|\nabla w_j\|_0 = 0$$

so that $\nabla w = \mathbf{0}$. Since Ω is connected, w is constant and since $w \in H^1_{0,\Gamma_0}(\Omega)$, we infer $w = 0$, in contradiction to $\|w\|_0 = 1$.

The proof in the other cases is identical. $\square$

Remark 7.8. If $u \in H^1(\Omega)$, let

$$\frac{1}{|\Omega|} \int_\Omega u = u_\Omega. \tag{7.55}$$

Then $w = u - u_\Omega$ has zero mean value and (7.53) holds for w. Thus, in general, the Poincaré inequality takes the form:

$$\|u - u_\Omega\|_0 \leq C_P \|\nabla u\|_0.$$

7.10.3 Sobolev inequality in $\mathbb{R}^n$

From Proposition 7.9 we know that the elements of $H^1(\mathbb{R})$ are continuous and (Problem 7.19) vanish as $x \to \pm\infty$. Moreover, if $u \in \mathcal{D}(\mathbb{R})$, we may write

$$u^2(x) = \int_{-\infty}^x \frac{d}{ds} u^2(s)\, ds = 2\int_{-\infty}^x u(s)\, u'(s)\, ds.$$

Using Schwarz's inequality and $2ab \leq 2a^2 + 2b^2$, we get

$$u(x)^2 \leq 2\|u\|_{L^2(\mathbb{R})}\|u'\|_{L^2(\mathbb{R})} \leq \|u\|^2_{L^2(\mathbb{R})} + \|u'\|^2_{L^2(\mathbb{R})} = \|u\|^2_{H^1(\mathbb{R})}.$$

Since $\mathcal{D}(\mathbb{R})$ is dense in $H^1(\mathbb{R})$, this inequality holds if $u \in H^1(\mathbb{R})$ as well. We have proved:

Proposition 7.14. *Let* $u \in H^1(\mathbb{R})$. *Then* $u \in L^\infty(\mathbb{R})$ *and*

$$\|u\|_{L^\infty(\mathbb{R})} \leq \|u\|_{H^1(\mathbb{R})}.$$

On the other hand, Example 7.26 implies that, when $\Omega \subseteq \mathbb{R}^n$, $n \geq 2$,

$$u \in H^1(\Omega) \nRightarrow u \in L^\infty(\Omega).$$

However, it is possible to prove that u is actually p−summable with a suitable $p > 2$. Moreover, the L^p−norm of u can be estimated by the H^1−norm of u.

To realize which exponent p is the correct one, assume that the inequality

$$\|u\|_{L^p(\mathbb{R}^n)} \leq c \|\nabla u\|_{L^2(\mathbb{R}^n)} \tag{7.56}$$

is valid **for every** $u \in \mathcal{D}(\mathbb{R}^n)$, *where c may depend on p and n* **but not on** u. We now use a typical "dimensional analysis" argument.

Inequality (7.56) must be *invariant under dilations* in the following sense. Let $u \in \mathcal{D}(\mathbb{R}^n)$ and for $\lambda > 0$ set

$$u_\lambda(\mathbf{x}) = u(\lambda \mathbf{x}).$$

Then $u_\lambda \in \mathcal{D}(\mathbb{R}^n)$ so that inequality (7.56) must be true for u_λ, with c **independent of** λ:

$$\|u_\lambda\|_{L^p(\mathbb{R}^n)} \leq c \|\nabla u_\lambda\|_{L^2(\mathbb{R}^n)}. \tag{7.57}$$

Now,

$$\int_{\mathbb{R}^n} |u_\lambda|^p \, d\mathbf{x} = \int_{\mathbb{R}^n} |u(\lambda \mathbf{x})|^p \, d\mathbf{x} = \frac{1}{\lambda^n} \int_{\mathbb{R}^n} |u(\mathbf{y})|^p \, d\mathbf{y}$$

while

$$\int_{\mathbb{R}^n} |\nabla u_\lambda|^2 \, d\mathbf{x} = \int_{\mathbb{R}^n} |\nabla u(\lambda \mathbf{x})|^2 \, d\mathbf{x} = \frac{1}{\lambda^{n-2}} \int_{\mathbb{R}^n} |\nabla u(\mathbf{y})|^2 \, d\mathbf{y}.$$

Therefore, (7.57) becomes

$$\frac{1}{\lambda^{n/p}} \left(\int_{\mathbb{R}^n} |u|^p \, d\mathbf{y} \right)^{1/p} \leq c(n, p) \frac{1}{\lambda^{(n-2)/2}} \left(\int_{\mathbb{R}^n} |\nabla u|^2 \, d\mathbf{y} \right)^{1/2}$$

or

$$\|u\|_{L^p(\mathbb{R}^n)} \leq c \lambda^{1 - \frac{n}{2} + \frac{n}{p}} \|\nabla u\|_{L^2(\mathbb{R}^n)}.$$

The only way to get a constant independent of λ is to choose p such that

$$1 - \frac{n}{2} + \frac{n}{p} = 0.$$

Hence, if $n > 2$, the correct p is given by

$$p^* = \frac{2n}{n-2}$$

which is called the *Sobolev exponent* for $H^1(\mathbb{R}^n)$. The following theorem of *Sobolev, Gagliardo, Nirenberg* holds:

Theorem 7.17. *Let* $u \in H^1(\mathbb{R}^n), n \geq 3$. *Then* $u \in L^{p^*}(\mathbb{R}^n)$ *with* $p^* = \frac{2n}{n-2}$, *and the following inequality holds.*

$$\|u\|_{L^{p^*}(\mathbb{R}^n)} \leq c \|\nabla u\|_{L^2(\mathbb{R}^n)} \tag{7.58}$$

where $c = c(n)$.

In the case $n = 2$ the correct statement is:

Proposition 7.15. *Let* $u \in H^1(\mathbb{R}^2)$. *Then* $u \in L^p(\mathbb{R})$ *for* $2 \leq p < \infty$, *and*

$$\|u\|_{L^p(\mathbb{R}^2)} \leq c(p) \|u\|_{H^1(\mathbb{R}^2)}.$$

7.10.4 Bounded domains

We now consider bounded domains. In dimension $n = 1$, the elements of $H^1(a, b)$ are continuous in $[a, b]$ and therefore bounded as well. Furthermore, the following inequality holds:

$$\|v\|_{L^\infty(a,b)} \leq C^* \|v\|_{H^1(a,b)} \tag{7.59}$$

with

$$C^* = \sqrt{2} \max \left\{ (b-a)^{-1/2}, (b-a)^{1/2} \right\}.$$

Indeed, by Schwarz's inequality we have, for any $x, y \in [a, b]$, $y > x$:

$$u(y) = u(x) + \int_x^y u'(s)\, ds$$

$$\leq u(x) + \sqrt{b-a} \|u'\|_{L^2(a,b)}$$

whence, using the elementary inequality $(A + B)^2 \leq 2A^2 + 2B^2$,

$$u(y)^2 \leq 2u(x)^2 + 2(b-a)\|u'\|_{L^2(a,b)}^2.$$

Integrating over (a, b) with respect to x we get

$$(b-a)u(y)^2 \leq 2\|u\|_{L^2(a,b)}^2 + 2(b-a)^2\|u'\|_{L^2(a,b)}^2$$

from which (7.59) follows easily.

In dimension $n \geq 2$, the improvement in summability is indicated in the following theorem.

Theorem 7.18. *Let* Ω *be a bounded, Lipschitz domain. Then:*

1. *If* $n > 2$, $H^1(\Omega) \hookrightarrow L^p(\Omega)$ *for* $2 \leq p \leq \frac{2n}{n-2}$. *Moreover, if* $2 \leq p < \frac{2n}{n-2}$, *the embedding of* $H^1(\Omega)$ *in* $L^p(\Omega)$ *is compact.*

2. *If* $n = 2$, $H^1(\Omega) \hookrightarrow L^p(\Omega)$ *for* $2 \leq p < \infty$, *with compact embedding.*

In the above cases

$$\|u\|_{L^p(\Omega)} \le c(n, p, \Omega) \|u\|_{H^1(\Omega)}.$$

For instance, in the important case $n = 3$, we have

$$p^* = \frac{2n}{n-2} = 6.$$

Hence

$$H^1(\Omega) \hookrightarrow L^6(\Omega)$$

and

$$\|u\|_{L^6(\Omega)} \le c(\Omega) \|u\|_{H^1(\Omega)}.$$

When the embedding of $H^1(\Omega)$ in $L^p(\Omega)$ is compact, the Poincaré inequality in Theorem 7.16 may be replaced by (see Problem 7.20)

$$\|u\|_{L^p(\Omega)} \le c(n, p, \Omega) \|\nabla u\|_{L^2(\Omega)}. \tag{7.60}$$

Theorem 7.18 shows what we can conclude about a H^1-function with regards to further regularity. It is natural to expect something more for H^m-functions, with $m > 1$. In fact:

Theorem 7.19. *Let Ω be a bounded, Lipschitz domain, and $m > n/2$. Then*

$$H^m(\Omega) \hookrightarrow C^k(\overline{\Omega}), \qquad for \ 0 \le k < m - \frac{n}{2}$$

with compact embedding. In particular,

$$\|u\|_{C^k(\overline{\Omega})} \le c(n, k, \Omega) \|u\|_{H^m(\Omega)}.$$

Theorem 7.19 implies that, in dimension $n = 2$,

$$H^2(\Omega) \subset C^0(\overline{\Omega}).$$

In fact, if $m = 2$, $n = 2$ then $m - n/2 = 1$, so that $k = 0$. Similarly

$$H^3(\Omega) \subset C^1(\overline{\Omega}),$$

since $m - n/2 = 3 - 1 = 2$.

Also in dimension $n = 3$, we have

$$H^2(\Omega) \subset C^0(\overline{\Omega}) \quad \text{and} \quad H^3(\Omega) \subset C^1(\overline{\Omega}),$$

as it is easy to check.

Remark 7.9. If $u \in H^m(\Omega)$ for any $m \ge 1$, then $u \in C^\infty(\overline{\Omega})$. This kind of results is very useful in the regularity theory for boundary value problems.

7.11 Spaces Involving Time

7.11.1 Functions with values in Hilbert spaces

The natural functional setting for evolution problems requires spaces which involve time. Given a function $u = u(\mathbf{x},t)$, it is often convenient to separate the roles of space and time adopting the following point of view. Assume that $t \in [0, T]$ and that for every t, or at least for a.e. t, the function $u(\cdot, t)$ belongs to a Hilbert space V (e.g. $L^2(\Omega)$ or $H^1(\Omega)$).

Then, we may consider u as a function of the real variable t with values in V:

$$u \colon [0, T] \to V.$$

When we adopt this convention, we write $u(t)$ and $\dot{u}(t)$ instead of $u(\mathbf{x},t)$ and $u_t(\mathbf{x},t)$.

We can extend to these types of functions the notions of *measurability* and *integral*, without too much effort, following more or less the procedure outlined in Appendix B. First, we introduce the set of functions $s \colon [0, T] \to V$ which assume only a finite number of values. These functions are called *simple* and are of the form

$$s(t) = \sum_{j=1}^{N} \chi_{E_j}(t)\, u_j \qquad (0 \leq t \leq T) \tag{7.61}$$

where, $u_1, ..., u_N \in V$ and $E_1, ..., E_N$ are Lebesgue measurable, mutually disjoint subsets of $[0, T]$.

We say that $f \colon [0, T] \to V$ is *measurable* if there exists a sequence of simple functions $s_k \colon [0, T] \to V$ such that, as $k \to \infty$,

$$\|s_k(t) - f(t)\|_V \to 0, \qquad \text{a.e. in } [0, T].$$

It is not difficult to prove that, if f is *measurable* and $v \in V$, the (real) function $t \longmapsto (f(t), v)_V$ is Lebesgue measurable in $[0, T]$.

The notion of integral is defined first for simple functions. If s is given by (7.61), we define

$$\int_0^T s(t)\, dt = \sum_{j=1}^{N} |E_j|\, u_j.$$

Then:

Definition 7.16. *We say that $f \colon [0, T] \to V$ is summable in $[0, T]$ if there exists a sequence $s_k \colon [0, T] \to V$ of simple functions such that*

$$\int_0^T \|s_k(t) - f(t)\|_V \; dt \to 0 \qquad \text{as } k \to +\infty. \tag{7.62}$$

If f is summable in $[0, T]$, we define the integral of f as follows:

$$\int_0^T f(t)\, dt = \lim_{k \to +\infty} \int_0^T s_k(t)\, dt \qquad \text{as } k \to +\infty. \tag{7.63}$$

Since (check it)

$$\left\| \int_0^T [s_h(t) - s_k(t)] \, dt \right\|_V \leq \int_0^T \|s_h(t) - s_k(t)\|_V \, dt$$

$$\leq \int_0^T \|s_k(t) - f(t)\|_V \, dt + \int_0^T \|s_k(t) - f(t)\|_V \, dt$$

it follows from (7.62) that the real sequence

$$\left\{ \int_0^T s_k(t) \, dt \right\}$$

is a Cauchy sequence so that the limit (7.63) is well defined and does not depend on the choice of the approximating sequence $\{s_k\}$. Moreover, the following important theorem holds:

Theorem 7.20. (Bochner). *A measurable function $f \colon [0, T] \to V$ is summable in $[0, T]$ if and only if the real function $t \longmapsto \|f(t)\|_V$ is summable in $[0, T]$. Moreover*

$$\left\| \int_0^T f(t) \, dt \right\|_V \leq \int_0^T \|f(t)\|_V \, dt \tag{7.64}$$

and

$$\left(u, \int_0^T f(t) \, dt \right)_V = \int_0^T (u, f(t))_V \, dt, \qquad \forall u \in V. \tag{7.65}$$

The inequality (7.64) is well known in the case of real or complex functions. By Riesz's Representation Theorem, (7.65) shows that the action of any element of V^* commutes with the integrals.

7.11.2 Sobolev spaces involving time

Once the definition of integral has been given, we can introduce the spaces $C([0, T]; V)$ and $L^p(0, T; V)$, $1 \leq p \leq \infty$. The symbol $C([0, T]; V)$ denotes the set of continuous functions $u \colon [0, T] \to V$. Endowed with the norm

$$\|u\|_{L^\infty(0,T;V)} = \max_{0 \leq t \leq T} \|u(t)\|_V \,,$$

$C([0, T]; V)$ is a Banach space.

We define $L^p(0, T; V)$ as the set of measurable functions $u \colon [0, T] \to V$ such that:

if $1 \leq p < \infty$

$$\|u\|_{L^p(0,T;V)} = \left(\int_0^T \|u(t)\|_V^p \, dt \right)^{1/p} < \infty \tag{7.66}$$

while if $p = \infty$

$$\|u\|_{L^\infty(0,T;V)} = \operatorname*{ess\,sup}_{0 \le t \le T} \|u(t)\|_V < \infty.$$

Endowed with the above norms, $L^p(0,T;V)$ becomes a Banach space for $1 \le p \le \infty$. If $p = 2$, the norm (7.66) is induced by the inner product

$$(u,v)_{L^2(0,T;V)} = \int_0^T (u(t), v(t))_V \, dt$$

that makes $L^2(0,T;V)$ a Hilbert space.

To define *Sobolev spaces*, we need to give the notion of derivative in the sense of distributions for functions $u \in L^1_{loc}(0,T;V)$.

We say that $\dot{u} \in L^1_{loc}(0,T;V)$ is the *derivative in the sense of distribution* (or the *weak derivative*) of u if

$$\int_0^T \varphi(t) \, \dot{u}(t) \, dt = - \int_0^T \dot{\varphi}(t) \, u(t) \, dt$$

for every $\varphi \in \mathcal{D}(0,T)$ or, equivalently, if

$$\int_0^T \varphi(t) \, (\dot{u}(t), v)_V \, dt = - \int_0^T \dot{\varphi}(t) \, (u(t), v)_V \, dt \qquad \forall v \in V. \tag{7.67}$$

Then, we can introduce the following spaces:

a) *We denote by* $W^{1,p}(0,T;V)$ *the Sobolev space of the functions* $u \in L^p(0,T;V)$ *whose weak derivative*

$$\dot{u} \in L^p(0,T;V).$$

With the norm

$$\|u\|_{W^{1,p}(0,T;V)} = \left(\int_0^T \|u(t)\|_V^p \, dt + \int_0^T \|\dot{u}(t)\|_V^p \, dt \right)^{1/p}, \qquad \text{if } 1 \le p < \infty$$

and

$$\|u\|_{W^{1,\infty}(0,T;V)} = \sup_{0 \le t \le T} \|u(t)\|_V + \sup_{0 \le t \le T} \|\dot{u}(t)\|_V, \qquad \text{if } p = \infty$$

these spaces are all Banach spaces.

b) *If* $p = 2$, *we may write* $H^1(0,T;V)$ *instead of* $W^{1,2}(0,T;V)$. This is a Hilbert space with inner product

$$(u,v)_{H^1(0,T;V)} = \int_0^T \{(u(t), v(t))_V + (\dot{u}(t), \dot{u}(t))_V\} \, dt.$$

Since functions in $H^1(a,b)$ are continuous in $[a,b]$, it makes sense to consider the value of u at any point of $[a,b]$. In a certain way, the functions in $W^{1,p}(0,T;V)$ depends only on the real variable t, so that the following theorem is not surprising.

Theorem 7.21. *Let $u \in H^1(0, T; V)$. Then, $u \in C([0, T]; V)$ and*

$$\max_{0 \leq t \leq T} \|u(t)\|_V \leq C(T) \|u\|_{H^1(0,T;V)}.$$

Moreover, the fundamental theorem of calculus holds:

$$u(t) = u(s) + \int_s^t \dot{u}(r) \, dr \qquad 0 \leq s \leq t \leq T.$$

The typical functional setting for the applications to initial-boundary value problems is a Hilbert triplet (V, H, V^*),

$$V \hookrightarrow H \hookrightarrow V^*,$$

with V separable. It is necessary to deal with functions $u \in L^2(0, T; V)$ whose derivative $\dot{u}$ belongs to $L^2(0, T; V^*)$. This means that in the left hand side of (7.67), the inner product $(\dot{u}(t), v)_V$ has to be replaced by the duality $\langle \dot{u}(t), v \rangle_*$. The following result is fundamental[25].

Theorem 7.22. *Let $u \in L^2(0, T; V)$, with $\dot{u} \in L^2(0, T; V^*)$. Then :*

a) $u \in C([0, T]; H)$ and

$$\max_{0 \leq t \leq T} \|u(t)\|_H \leq C \left\{ \|u\|_{L^2(0,T;V)} + \|\dot{u}\|_{L^2(0,T;V^*)} \right\}. \tag{7.68}$$

b) If also $v \in L^2(0, T; V)$ and $\dot{v} \in L^2(0, T; V^*)$, the following integration by parts formula holds:

$$\int_s^t \left\{ \langle \dot{u}(r), v(r) \rangle_* + \langle u(r), \dot{v}(r) \rangle_* \right\} dr = (u(t), v(t))_H - (u(s), v(s))_H \tag{7.69}$$

for all $s, t \in [0, T]$.

Remark 7.10. From (7.69) we infer,

$$\frac{d}{dt} (u(t), v(t))_H = \langle \dot{u}(t), v(t) \rangle_* + \langle u(t), \dot{v}(t) \rangle_*$$

a.e. $t \in [0, T]$ and (letting $u = v$)

$$\int_s^t \frac{d}{dt} \|u(r)\|_H^2 \, dt = \|u(t)\|_H^2 - \|u(s)\|_H^2. \tag{7.70}$$

We conclude this chapter with a useful result (see Problem 7.25): the weak convergence in $L^2(0, T; V)$ "preserves boundedness in $L^\infty(0, T; V)$".

Proposition 7.16. *Let $\{u_k\} \subset L^2(0, T; V)$, weakly convergent to u. Assume that*

$$\sup_{t \in [0,T]} \|u_k(t)\|_V \leq C$$

with C independent of k. Then, also,

$$\sup_{t \in [0,T]} \|u(t)\|_V \leq C.$$

---

[25] For the proof, see *Dautray-Lions*, volume 5, chapter XVIII, 1985.

Problems

7.1. *Approximations of δ.*

(a) Let $B_r = B_r(\mathbf{0}) \subset \mathbb{R}^n$. Show that, if χ_{B_r} is the characteristic function of B_r,

$$\lim_{r \to 0} \frac{1}{|B_r|} \chi_{B_r} = \delta \qquad \text{in } \mathcal{D}'(\mathbb{R}^n).$$

(b) Let η_ε be the mollifier in (7.3). Show that $\lim_{\varepsilon \to 0} \eta_\varepsilon = \delta$, in $\mathcal{D}'(\mathbb{R}^n)$.

(c) Let $\Gamma_D(\mathbf{x}, t)$ be the fundamental solution of the heat equation $u_t = D\Delta u$. Show that

$$\Gamma_D(\cdot, t) \to \delta, \quad \text{in } \mathcal{D}'(\mathbb{R}^n)$$

as $t \to 0^+$.

7.2. Let $\{x_k\} \subset \mathbb{R}$, $x_k \to +\infty$. Show that $\sum_{k=1}^\infty c_k \delta(x - x_k)$ converges in $\mathcal{D}'(\mathbb{R})$ for all $\{c_k\} \subset \mathbb{R}$.

7.3. Show that the series

$$\sum_{k=1}^\infty c_k \sin kx$$

converges in $\mathcal{D}'(\mathbb{R})$ if the numerical sequence $\{c_k\}$ is *slowly increasing*, i.e. if there exists $p \in \mathbb{R}$ such that $c_k = O(k^p)$ as $k \to \infty$.

7.4. Show that if $F \in \mathcal{D}'(\mathbb{R}^n)$, $v \in \mathcal{D}(\mathbb{R}^n)$ and v vanishes in an open set containing the support of F, then $\langle F, v \rangle = 0$. Is it true that $\langle F, v \rangle = 0$ if v vanishes *only* on the support of F?

7.5. Let $u(x) = |x|$ and $S(x) = \text{sign}(x)$. Prove that $u' = S$ in $\mathcal{D}'(\mathbb{R})$.

7.6. Prove that $x^m \delta^{(k)} = 0$ in $\mathcal{D}'(\mathbb{R})$, if $0 \le k < m$.

7.7. Let $u(x) = \ln|x|$. Then, $u' = \text{p.v.} \frac{1}{x}$, in $\mathcal{D}'(\mathbb{R})$.

[*Hint.* Write

$$\langle u', \varphi \rangle = -\langle u, \varphi' \rangle = -\int_\mathbb{R} \ln|x| \, \varphi'(x) \, dx = -\lim_{\varepsilon \to 0} \int_{\{|x| > \varepsilon\}} \ln|x| \, \varphi'(x) \, dx$$

and integrate by parts].

7.8. Let $n = 3$ and $\mathbf{F} \in \mathcal{D}'(\Omega; \mathbb{R}^3)$. Define curl $\mathbf{F} \in \mathcal{D}'(\Omega; \mathbb{R}^3)$ by the formula

$$\text{curl } \mathbf{F} = (\partial_{x_2} F_3 - \partial_{x_3} F_2, \partial_{x_3} F_1 - \partial_{x_1} F_3, \partial_{x_1} F_2 - \partial_{x_2} F_1).$$

Check that, for all $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3) \in \mathcal{D}(\Omega; \mathbb{R}^3)$,

$$\langle \text{curl } \mathbf{F}, \boldsymbol{\varphi} \rangle = \langle \mathbf{F}, \text{curl } \boldsymbol{\varphi} \rangle.$$

7.9. Show that if

$$u(x_1, x_2) = -\frac{1}{2\pi} \ln(x_1^2 + x_2^2)$$

then
$$-\Delta u = \delta, \qquad \text{in } \mathcal{D}' \left(\mathbb{R}^2 \right).$$

7.10. Show that if $u_k \to u$ in $L^p(\mathbb{R}^n)$ then $u_k \to u$ in $\mathcal{S}'(\mathbb{R}^n)$.

7.11. Solve the equation $x^2 \delta = 0$ in $\mathcal{D}' \left(\mathbb{R} \right).$

7.12. Let $u \in C^\infty (\mathbb{R})$ with compact support in $[0, 1]$. Compute $comb * u.$
[*Answer.* $\sum_{k=-\infty}^{+\infty} u \left(x - k \right)$].

7.13. Let $\mathcal{H} = \mathcal{H} (x)$ be the Heaviside function. Prove that

$$a) \; \mathcal{F} \left[\text{sign} x \right] = \frac{2}{i} \; p.v.\frac{1}{\xi}, \qquad b) \; \mathcal{F} \left[\mathcal{H} \right] = \pi\delta + \frac{1}{i} \; p.v.\frac{1}{\xi}.$$

[*Hint. a*) Let $u \left(x \right) = \text{sign}(x).$ Note that $u' = 2\delta.$ Transform this equation to obtain

$$\xi\widehat{u} \left(\xi \right) = -2i.$$

Solve this equation using formula (7.18), and recall that $\widehat{u}$ is odd while δ is even.
b) Write $\mathcal{H} \left(x \right) = \frac{1}{2} + \frac{1}{2}\text{sign} x$ and use a)].

7.14. Compute the Fourier transform of $comb$.

7.15. Let $\Omega = B_1 \left(\mathbf{0} \right) \subset \mathbb{R}^n$, $n > 2$, and $u \left(\mathbf{x} \right) = |\mathbf{x}|^{-a}$, $\mathbf{x} \neq \mathbf{0}$. Determine for which values of a, $u \in H^2 \left(\Omega \right)$.

7.16. Choose in Theorem 7.4,

$$H = L^2(\Omega; \mathbb{R}^n), \qquad Z = L^2(\Omega) \subset \mathcal{D}' \left(\Omega \right)$$

and $L : H \to \mathcal{D}' \left(\Omega \right)$ given by $L = \text{div}$. Identify the resulting space W.

7.17. Let X and Z be Banach spaces with $Z \hookrightarrow \mathcal{D}' \left(\Omega; \mathbb{R}^n \right)$ (e.g. $L^p(\Omega)$ or $L^p(\Omega; \mathbb{R}^n)$).
Let $L : X \to \mathcal{D}' \left(\Omega; \mathbb{R}^n \right)$ be a linear continuous operator (e.g. a gradient or a divergence). Define
$$W = \{v \in X : Lv \in Z\}$$
with norm
$$\|u\|_W^2 = \|u\|_X^2 + \|Lu\|_Z^2.$$
Prove that W is a Banach space, continuously embedded in X.
[*Hint*: Follow the proof of Theorem 7.4].

7.18. *The Sobolev spaces $W^{1,p}$.* Let $\Omega \subseteq \mathbb{R}^n$ be a domain. For $p \geq 1$. Define

$$W^{1,p}(\Omega) = \{v \in L^p(\Omega) : \nabla v \in L^p(\Omega; \mathbb{R}^n)\}.$$

Using the result of Problem 7.17, show that $W^{1,p}(\Omega)$ is a Banach space.

7.19. Let $u \in H^s (\mathbb{R})$. Prove that, if $s > 1/2$, $u \in C (\mathbb{R})$ and $u \left(x \right) \to 0$ as $x \to \pm\infty$.
[*Hint*: Show that $\widehat{u} \in L^1 (\mathbb{R})$].

7.20. Let u and Ω satisfy the hypotheses of Theorem 7.16. Prove that, if the embedding $H^1(\Omega) \hookrightarrow L^p(\Omega)$ is compact, then

$$\|u\|_{L^p(\Omega)} \le c(n, p, \Omega) \|\nabla u\|_{L^2(\Omega)}.$$

7.21. Let Ω be a bounded domain (not necessarily Lipschitz). Show that $H_0^1(\Omega)$ is compactly embedded in $L^2(\Omega)$.

[*Hint:* Extend u by zero outside Ω].

7.22. Let

$$H_{0,a}^1(a, b) = \left\{ u \in H^1(a, b) : u(a) = 0 \right\}.$$

Show that Poincaré's inequality holds in $H_{0,a}^1(a, b)$.

7.23. Let $n > 1$ and

$$\Omega = \left\{ (\mathbf{x}', x_n) : \mathbf{x}' \in \mathbb{R}^{n-1},\, 0 < x_n < d \right\}.$$

Show that in $H_0^1(\Omega)$ a Poincaré inequality holds.

7.24. Let Ω be a bounded, Lipschitz domain and let $\Gamma = \partial\Omega$.

a) Show that

$$(u, v)_{1,\partial} = \int_\Gamma u_{|\Gamma}\, v_{|\Gamma}\ d\sigma + \int_\Omega \nabla u \cdot \nabla v\ d\mathbf{x}$$

is an inner product in $H^1(\Omega)$.

b) Show that the norm

$$\|u\|_{1,\partial} = \left(\int_{\partial\Omega} u_{|\Gamma}^2\ d\sigma + \int_\Omega |\nabla u|^2\ d\mathbf{x} \right)^{1/2} \tag{7.71}$$

is equivalent to $\|u\|_{1,2}$.

[*Hint: b*) Let $u_{|\Gamma} = g$. By the Projection Theorem 6.2, it is possible to write in a unique way

$$u = u_0 + \tilde{g},$$

with $u_0 \in H_0^1(\Omega)$, $\tilde{g} \in H^1(\Omega)$ and $(u_0, \tilde{g})_{1,2} = 0$. Using (7.48), show that $\|g\|_{H^{1/2}(\Gamma)} = \|\tilde{g}\|_{1,2}$].

7.25. Prove Proposition 7.16.

[*Hint.* Recall that a sequence of real functions $\{g_k\}$ convergent to g in $L^1(0, T)$, has a subsequence converging a.e. to the same limit (Theorem B.4).

Apply this result to the sequence $g_k(t) = (u_k(t), v)_V$ and observe that $|g_k(t)| \le C \|v\|_V$].

Variational Formulation of Elliptic Problems

Elliptic Equations – The Poisson Problem – Diffusion, Drift and Reaction ($n = 1$) – Variational Formulation of Poisson's Problem – General Equations in Divergence Form – Regularity – Equilibrium of a plate – A Monotone Iteration Scheme for Semilinear Equations – A Control Problem

8.1 Elliptic Equations

Poisson's equation $\Delta u = f$ is the simplest among the *elliptic equations*, according to the classification in Section 5.5, at least in dimension two. This type of equations plays an important role in the modelling of a large variety of phenomena, often of stationary nature. Typically, in drift, diffusion and reaction models, like those considered in Chapter 2, a stationary condition corresponds to a steady state, with no more dependence on time.

Elliptic equations appear in the theory of electrostatic and electromagnetic potentials or in the search of vibration modes of elastic structures as well (e.g. through the method of separation of variables for the wave equation).

Let us define precisely what we mean by *elliptic equation* in dimension n.

Let $\Omega \subseteq \mathbb{R}^n$ be a domain, $\mathbf{A}\left(\mathbf{x}\right) = \left(a_{ij}\left(\mathbf{x}\right)\right)$ a square matrix of order n, $\mathbf{b}(\mathbf{x}) = \left(b_1\left(\mathbf{x}\right), ..., b_n\left(\mathbf{x}\right)\right)$, $\mathbf{c}(\mathbf{x}) = \left(c_1\left(\mathbf{x}\right), ..., c_n\left(\mathbf{x}\right)\right)$ vector fields in $\mathbb{R}^n$, $a_0 = a_0(\mathbf{x})$ and $f = f(\mathbf{x})$ real functions. An equation of the form

$$-\sum_{i,j=1}^{n} \partial_{x_i}\left(a_{ij}\left(\mathbf{x}\right) u_{x_j}\right) + \sum_{i=1}^{n} \partial_{x_i}\left(b_i(\mathbf{x})u\right) + \sum_{i=1}^{n} c_i(\mathbf{x})u_{x_i} + a_0\left(\mathbf{x}\right)u = f\left(\mathbf{x}\right) \quad (8.1)$$

or

$$-\sum_{i,j=1}^{n} a_{ij}\left(\mathbf{x}\right) u_{x_i x_j} + \sum_{i=1}^{n} b_i\left(\mathbf{x}\right) u_{x_i} + a_0\left(\mathbf{x}\right)u = f\left(\mathbf{x}\right) \quad (8.2)$$

is said to be **elliptic in** Ω if $\mathbf{A}$ is **positive** in Ω, i.e. if the following *ellipticity condition* holds:

$$\sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\, \xi_i \xi_j > 0, \qquad \forall \mathbf{x} \in \Omega, \ \forall \boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\xi} \neq \mathbf{0}.$$

We say that (8.1) is in **divergence form** since it mat be written as

$$\underbrace{-\mathrm{div}(\mathbf{A}(\mathbf{x})\,\nabla u)}_{diffusion} + \underbrace{\mathrm{div}(\mathbf{b}(\mathbf{x})u) + \mathbf{c}(\mathbf{x}) \cdot \nabla u}_{transport} + \underbrace{a_0(\mathbf{x})\,u}_{reaction} = \underbrace{f(\mathbf{x})}_{external\ source} \qquad (8.3)$$

which emphasizes the particular structure of the higher order terms. Usually, the first term models the diffusion in heterogeneous or anisotropic media, when the constitutive law for the flux function $\mathbf{q}$ is given by the Fourier or Fick law:

$$\mathbf{q} = -\mathbf{A}\nabla u.$$

Here u may represent a temperature or the concentration of a substance. Thus, the term $-\mathrm{div}(\mathbf{A}\nabla u)$ is associated with thermal or molecular diffusion. The matrix $\mathbf{A}$ is called *diffusion matrix*; the dependence of $\mathbf{A}$ on $\mathbf{x}$ denotes anisotropic diffusion.

The examples in Chapter 2 explain the meaning of the other terms in equation (8.3). In particular, $\mathrm{div}(\mathbf{b}u)$ models *convection or transport* and corresponds to a flux function given by

$$\mathbf{q} = \mathbf{b}u.$$

The vector $\mathbf{b}$ has the dimensions of a **velocity**. Think, for instance, of the fumes emitted by a factory installations, which diffuse and are transported by the wind. In this case $\mathbf{b}$ is the wind velocity. Note that, if $\mathrm{div}\mathbf{b} = 0$, then $\mathrm{div}(\mathbf{b}u)$ reduces to $\mathbf{b} \cdot \nabla u$ which is of the same form of the third term $\mathbf{c} \cdot \nabla u$.

The term $a_0 u$ models *reaction*. If u is the concentration of a substance, a_0 represents the rate of decomposition ($a_0 > 0$) or growth ($a_0 < 0$).

Finally, f represents an external action, distributed in Ω, e.g. the rate of heat per unit mass supplied by an external source.

If the entries a_{ij} of the matrix $\mathbf{A}$ and the component b_j of $\mathbf{b}$ are all differentiable, we may compute the divergence of both $\mathbf{A}\nabla u$ and $\mathbf{b}u$, and reduce (8.1) to the *non-divergence form*

$$-\sum_{i,j=1}^{n} a_{ij}(\mathbf{x})\, u_{x_i x_j} + \sum_{k=1}^{n} \tilde{b}_k(\mathbf{x})\, u_{x_k} + \tilde{c}(\mathbf{x})\, u = f(\mathbf{x})$$

where

$$\tilde{b}_k(\mathbf{x}) = \sum_{i=1}^{n} \partial_{x_i} a_{ik}(\mathbf{x}) + b_k(\mathbf{x}) + c_k(\mathbf{x}) \quad \text{and} \quad \tilde{c}(\mathbf{x}) = \mathrm{div}\mathbf{b}(\mathbf{x}) + a_0(\mathbf{x}).$$

However, when the a_{ij} or the b_j are *not differentiable*, we must keep the divergence form and interpret the differential equation (8.3) in a suitable weak sense.

A *non-divergence form equation* is also associated with diffusion phenomena through stochastic processes which generalize the Brownian motion, called *diffusion processes*. In simple cases, we may proceed as in Section 2.6. For example, considering a random walk in $h\mathbb{Z}^2$, separately symmetric along each axis, and passing to the limit in a suitable way as h and the time step τ go to zero, we obtain an equation of the form

$$u_t = D_1(x, y) u_{xx} + D_2(x, y) u_{yy}$$

with diffusion matrix

$$\mathbf{A}(x, y) = \begin{pmatrix} D_1(x, y) & 0 \\ 0 & D_2(x, y) \end{pmatrix}$$

where $D_1(x, y) > 0$, $D_2(x, y) > 0$. Thus, the steady state case is a solution of a non-divergence form equation.

In the next section we give a brief account of the various notions of solution available for these kinds of equations, using the Poisson equation as a model.

We develop the basic theory of elliptic equations in divergence form, recasting the most common boundary value problems within the functional framework of the abstract variational problems of Section 6.6.

8.2 The Poisson Problem

Assume we are given a domain $\Omega \subset \mathbb{R}^n$, a constant $\alpha > 0$ and two real functions $a_0, f : \Omega \to \mathbb{R}$. We want to determine a function u satisfying the equation

$$-\alpha \Delta u + a_0 u = f \qquad \text{in } \Omega$$

and one of the usual boundary conditions on $\partial \Omega$.

Let us examine what we mean by *solving* the above Poisson problem. The obvious part is the final goal: we have to show *existence, uniqueness and stability* of the solution; then, based on these results, we want to *compute* the solution by Numerical Analysis methods.

Less obvious is the *meaning of solution*. In fact, in principle, every problem may be formulated in several ways and a different notion of solution is associated with each way. What is important in the applications is to select the "most efficient notion" for the problem under examination, where "efficiency" may stand for the best compromise between *simplicity of both formulation and theoretical treatment, sufficient flexibility and generality, adaptability to numerical methods*.

Here is a (non exhaustive!) list of various notions of solution for the Poisson problem.

• **Classical** solutions are twice continuously differentiable functions; the differential equation and the boundary conditions are satisfied in the usual pointwise sense.

• **Strong** solutions belong to the Sobolev space $H^2(\Omega)$. Thus, they possess derivatives in $L^2(\Omega)$ up to the second order, in the sense of distributions.

The differential equation is satisfied in the pointwise sense, a.e. with respect to the Lebesgue measure in Ω, while the boundary condition is satisfied in the sense of traces.

• **Distributional** solutions belong to $L^1_{loc}(\Omega)$ and the equation holds in the sense of distributions, i.e.:

$$\int_\Omega \{-\alpha u \Delta\varphi + a_0(\mathbf{x})u\varphi\}\,d\mathbf{x} = \int_\Omega f\varphi d\mathbf{x}, \quad \forall\varphi \in \mathcal{D}(\Omega).$$

The boundary condition is satisfied in a very weak sense.

• **Weak or variational** solutions belong to the Sobolev space $H^1(\Omega)$. The boundary value problem is recast within the framework of the abstract variational theory developed in Section 6.6. Often the new formulation represents a version of the *principle of virtual work.*

Clearly, all these notions of solution must be connected by a *coherence principle*, which may be stated as follows: if all the data (domain, boundary data, forcing terms) and the solution are C^∞, *all the above notions must be equivalent.* Thus, the *non-classical* notions constitute a generalization of the classical one.

An important task, with consequences in the error control in numerical methods, is to establish the optimal degree of regularity of a non-classical solution.

More precisely, let u be a non-classical solution of the Poisson problem. The question is: *how much does the regularity of the data a_0, f and of the domain Ω affect the regularity of the solution?*

An exhaustive answer requires rather complicated tools. In the sequel we shall indicate only the most relevant results.

The theory for classical and strong solutions is well established and can be found, e.g. in the book of *Gilbarg-Trudinger,* 1998. From the numerical point of view, the *method of finite differences* best fits the differential structure of the problem and aims at approximating classical solutions.

The distributional theory is well developed, is quite general, but is not the most appropriate framework for solving boundary value problems.

Indeed, the sense in which the boundary values are attained is one of the most delicate points, when one is willing to widen the notion of solution.

For our purposes, the most convenient notion of solution is the last one: it leads to a quite flexible formulation with a sufficiently high degree of generality and a basic theory solely relying on the Lax-Milgram Theorem (Section 6.6). Moreover, the analogy (and often the coincidence) with the principle of virtual work indicates a direct connection with the physical interpretation.

Finally, the variational formulation is the most natural to implement the *Galerkin method (finite elements, spectral elements, etc...),* widely used in the numerical approximation of the solutions of boundary value problems.

To present the main ideas behind the variational formulation, we start from one-dimensional problems with an equation slightly more general than Poisson's equation.

8.3 Diffusion, Drift and Reaction ($n = 1$)

8.3.1 The problem

We shall derive the variational formulation of the following problem:

$$\begin{cases} \underbrace{-(p(x)u')'}_{diffusion} + \underbrace{q(x)\,u'}_{transport} + \underbrace{r(x)u}_{reaction} = f(x), & \text{in the interval } (a, b) \\ \qquad\qquad \text{boundary conditions} \qquad\qquad \text{at } x = a \text{ and } x = b. \end{cases} \tag{8.4}$$

We may interpret (8.4) as a stationary problem of *diffusion, drift and reaction*.

The main steps for the weak formulation are the following:

a. Select a space of *smooth test functions, adapted to the boundary conditions.*

b. Multiply the differential equation by a *test* function and integrate over (a, b) .

c. Carry one of the derivatives in the divergence term onto the test function via an integration by parts, using the boundary conditions and obtaining an *integral equation*.

d. Interpret the integral equation as an *abstract variational problem* (Section 6.6) in a suitable Hilbert space. In general, this is a Sobolev space, given by the closure of the space of test functions.

8.3.2 Dirichlet conditions

We start by analyzing *homogeneous Dirichlet conditions*:

$$\begin{cases} -(p(x)u')' + q(x)\,u' + r(x)u = f(x), & \text{in } (a, b) \\ \qquad\qquad u(a) = u(b) = 0. \end{cases} \tag{8.5}$$

Assume first that $p \in C^1([a, b])$, with $p > 0$, and $q, r, f \in C([a, b])$.

Let $u \in C^2(a, b) \cap C([a, b])$ be a classical solution of (8.5). We select $C_0^1(a, b)$ as the space of test functions. These test functions have a continuous derivative and compact support in (a, b). In particular, *they vanish at the end points.*

Now we multiply the equation by an arbitrary $v \in C_0^1(a, b)$ and integrate over (a, b). We find:

$$-\int_a^b (pu')'v\,dx + \int_a^b [qu' + ru]v\,dx = \int_a^b fv\,dx. \tag{8.6}$$

Integrating by parts the first term and using $v(a) = v(b) = 0$, we get:

$$-\int_a^b (pu')'v dx = \int_a^b pu'v' dx - [pu'v]_a^b = \int_a^b pu'v' dx.$$

From (8.6) we derive the integral equation

$$\int_a^b [pu'v' + qu'v + ruv] dx = \int_a^b fv dx, \qquad \forall v \in C_0^1(a,b). \tag{8.7}$$

Thus, (8.5) implies (8.7).

On the other hand, assume that (8.7) is true. Integrating by parts in the reverse order, we recover (8.6), which can be written in the form

$$\int_a^b \{-(pu')' + q(x)u' + r(x)u - f(x)\} v dx = 0 \qquad \forall v \in C_0^1(a,b).$$

The arbitrariness of v entails[1]

$$-(p(x)u')' + q(x)u' + r(x)u - f(x) = 0 \quad \text{in } (a,b)$$

i.e. the original differential equation.

Thus, **for classical solutions, the two formulations** (8.5) e (8.7) **are equivalent**. Observe that equation (8.7)

- involves *only one derivative* of u, instead of two,
- makes perfect sense even if p, q, r and f are merely locally integrable,
- has transformed (8.5) into an integral equation, valid on an infinite-dimensional space of test functions.

These features lead to the following functional setting:

a) we enlarge the class of test functions to $H_0^1(a,b)$, which is the closure of $C_0^1(a,b)$ in H^1−norm;

b) we look for a solution belonging to $H_0^1(a,b)$, in which the homogeneous Dirichlet conditions are already included.

Thus, the **weak** or **variational** formulation of problem (8.5) is:

Determine $u \in H_0^1(a,b)$ *such that*

$$\int_a^b \{pu'v' + qu'v + ruv\} dx = \int_a^b fv dx, \qquad \forall v \in H_0^1(a,b). \tag{8.8}$$

If we introduce the bilinear form

$$B(u,v) = \int_a^b \{pu'v' + qu'v + ruv\} dx$$

---

[1] If $g \in C([a,b])$ and $\int_a^b gv dx = 0$ for every $v \in C_0^1(a,b)$, then $g \equiv 0$ (exercise).

and the linear functional

$$Lv = \int_a^b fv \, dx,$$

equation (8.8) can be recast as

$$B(u, v) = Lv, \quad \forall v \in H_0^1(a, b).$$

Then existence, uniqueness and stability follow from the Lax-Milgram Theorem 6.5, under rather natural hypotheses on p, q, r, f. Recall that, by Poincaré's inequality (7.32) we have

$$\|u\|_0 \leq C_P \|u'\|_0,$$

so that we may choose in $H_0^1(a, b)$ the norm

$$\|u\|_1 = \|u'\|_0$$

equivalent to $\|u\|_{1,2} = \|u\|_0 + \|u'\|_0$

Proposition 8.1. *Assume that $p, q, q', r \in L^\infty(a, b)$ and $f \in L^2(a, b)$. If*

$$p(x) \geq \alpha > 0 \quad \text{and} \quad -\frac{1}{2} q'(x) + r(x) \geq 0 \quad \text{a.e. in } (a, b), \tag{8.9}$$

then (8.8) has a unique solution $u \in H_0^1(a, b)$. Moreover

$$\|u'\|_0 \leq \frac{C_P}{\alpha} \|f\|_0. \tag{8.10}$$

Proof. Let us check that the hypotheses of the Lax-Milgram Theorem hold, with $V = H_0^1(a, b)$.

Continuity of the bilinear form B. We have:

$$|B(u, v)| \leq \int_a^b \{\|p\|_{L^\infty} |u'v'| + \|q\|_{L^\infty} |u'v| + \|r\|_{L^\infty} |uv|\} \, dx.$$

Using the Schwarz and Poincaré inequalities, we obtain

$$|B(u, v)| \leq \|p\|_{L^\infty} \|u'\|_0 \|v'\|_0 + \|q\|_{L^\infty} \|u'\|_0 \|v\|_0 + \|r\|_{L^\infty} \|u\|_0 \|v\|_0$$
$$\leq \left(\|p\|_{L^\infty} + C_P \|q\|_{L^\infty} + C_P^2 \|r\|_{L^\infty}\right) \|u'\|_0 \|v'\|_0$$

so that B is continuous in V.

Coercivity of B. We may write:

$$B(u, u) = \int_a^b \{p(u')^2 + qu'u + ru^2\} \, dx$$

$$\geq \alpha \|u'\|_0^2 + \frac{1}{2} \int_a^b q \left(u^2\right)' \, dx + \int_a^b ru^2 \, dx$$

$$\text{(integrating by parts)} = \alpha \|u'\|_0^2 + \int_a^b \left\{-\frac{1}{2} q' + r\right\} u^2 \, dx$$

$$\text{(from (8.9))} \geq \alpha \|u'\|_0^2$$

and therefore B is $V-$coercive.

Continuity of L in V. The Schwarz and Poincaré inequalities yield

$$|Lv| = \left| \int_a^b fv \; dx \right| \le \|f\|_0 \|v\|_0 \le C_P \|f\|_0 \|v'\|_0 \,.$$

so that $\|L\|_{V^*} \le C_P \|f\|_0$.

Then, the Lax-Milgram Theorem gives existence, uniqueness and the stability estimate (8.10). $\square$

Remark 8.1. The hypotheses on the coefficient q are rather restrictive; a better result can be achieved using the *Fredholm Alternative* Theorem and a *weak maximum principle*, as we will show later on. This remark holds for the other types of boundary value problems as well.

Remark 8.2. If $q = 0$, the bilinear form B is symmetric. From Proposition 6.6, in this case the weak solution minimizes in $H_0^1(a, b)$ the "energy functional"

$$J(u) = \int_a^b \left\{ p(u')^2 + ru^2 - 2fu \right\} dx.$$

Then, equation (8.8) coincides with the Euler equation of J:

$$J'(u)v = 0, \qquad \forall v \in H_0^1(a, b).$$

Remark 8.3. In the case of nonhomogeneous Dirichlet conditions, e.g. $u(a) = A$, $u(b) = B$, set $w = u - y$, where $y = y(x)$ is the straight line through the points (a, A), (b, B), given by

$$y(x) = A + (x - a) \frac{B - A}{b - a}.$$

Then, the variational problem for w is

$$\int_a^b [pw'v' + qw'v + rwv]dx = \int_a^b (Fv + Gv') \; dx \qquad \forall v \in H_0^1(a, b) \qquad (8.11)$$

with

$$F(x) = f(x) + \frac{B - A}{b - a} q(x) - r(x) \left(A + (x - a) \frac{B - A}{b - a} \right)$$

and

$$G(x) = \frac{B - A}{b - a} p(x).$$

Proposition 8.1 still holds with small adjustments (see Problem 8.1).

Remark 8.4. In subsection 6.6.3. we presented the Galerkin approximation method in an abstract setting. In this case, we have to construct a sequence $\{V_k\}$ of subspaces of $H_0^1(a, b)$ such that:

a) Every V_k is *finite-dimensional* (for instance, but not necessarily, of dimension k);

b) $V_k \subset V_{k+1}$ (this is actually not strictly necessary);

c) $\overline{\cup V_k} = H_0^1(a, b)$.

Given a basis $\psi_1, \psi_2, ..., \psi_k$ for V_k, we write

$$u_k = \sum_{j=1}^{k} c_j \psi_j$$

and determine $c_1, c_2, ..., c_k$ by solving the k linear algebraic equations

$$\sum_{i=1}^{k} a_{ij} c_j = L\psi_i, \qquad i = 1, 2, ..., k,$$

where the elements a_{ij} of the stiffness matrix are given by

$$a_{ij} = B(\psi_j, \psi_i) = \int_a^b [p\psi_j'\psi_i' + q\psi_j'\psi_i + r\psi_j\psi_i]dx, \qquad i = 1, 2, ..., k.$$

Observe that, for the approximations, *Céa's Lemma* 6.1 *yields* the estimate

$$\|u - u_k\|_1 \leq \frac{\|p\|_{L^\infty} + C_P \|q\|_{L^\infty} + C_P^2 \|r\|_{L^\infty}}{\alpha} \inf_{v_k \in V_k} \|u - v_k\|_1 .$$

This inequality shows the relative influence of diffusion, transport and reaction on the approximation. In principle, the Galerkin approximation works as long as $\|q\|_{L^\infty} + \|r\|_{L^\infty}$ is not much greater than a_0 and $\|p\|_{L^\infty} /\alpha$ is not large. In the opposite case, one needs suitably stabilized numerical methods (see *Quarteroni-Valli*, 2000). Clearly, this remark extends to the other types of boundary conditions as well.

8.3.3 Neumann, Robin and mixed conditions

We now derive the weak formulation of the Neumann problem

$$\begin{cases} -(p(x)u')' + q(x) u' + r(x)u = f(x), & \text{in } (a, b) \\ -p(a) u'(a) = A, \quad p(b) u'(b) = B. \end{cases} \tag{8.12}$$

The boundary conditions prescribe the outward flux at the end points. This way of writing the Neumann conditions, with the presence of the factor p in front of the derivative, is naturally associated with the divergence structure of the diffusion term.

Again, assume first that $p \in C^1([a, b])$, with $p > 0$, and $q, r, f \in C^0([a, b])$. A classical solution u has a continuous derivative up to the end points so that $u \in C^2(a, b) \cap C^1([a, b])$.

As space of test functions, we choose $C^1([a, b])$. Multiplying the equation by an arbitrary $v \in C^1([a, b])$ and integrating over (a, b), we find again

$$-\int_a^b (pu')'vdx + \int_a^b [qu' + ru]vdx = \int_a^b fvdx. \qquad (8.13)$$

Integrating by parts the first term and using the Neumann conditions, we get

$$-\int_a^b (pu')'vdx = \int_a^b pu'v'dx - [pu'v]_a^b = \int_a^b pu'v'dx - v(b)B - v(a)A.$$

Then (8.13) becomes

$$\int_a^b [pu'v' + qu'v + ruv]dx - v(b)B - v(a)A = \int_a^b fvdx, \qquad (8.14)$$

for every $v \in C^1([a, b])$.

Thus, (8.12) implies (8.14). If the choice of the test functions is correct, we should be able to recover the classical formulation from (8.14).

Indeed, let us start recovering the differential equation. Since

$$C_0^1(a, b) \subset C^1([a, b]),$$

(8.14) clearly holds for every $v \in C_0^1(a, b)$. Then, (8.14) reduces to (8.7) and we deduce, as before,

$$-(pu')' + qu' + ru - f = 0, \qquad \text{in } (a, b). \qquad (8.15)$$

Let us now use the test functions which *do not vanish at the end points*. Integrating by parts the first term in (8.14) we have:

$$\int_a^b pu'v'dx = -\int_a^b (pu')'v \, dx + p(b)v(b)u'(b) - p(a)v(a)u'(a).$$

Inserting this expression into (8.14) and taking into account (8.15) we find:

$$v(b)[p(b)u'(b) - B] - v(a)[p(a)u'(a) + A] = 0.$$

The arbitrariness of the values $v(b)$ and $v(a)$ forces

$$p(b)u'(b) = B, \qquad -p(a)u'(a) = A,$$

recovering the Neumann conditions as well.

Thus, **for classical solutions, the two formulations** (8.12) and (8.14) **are equivalent**.

Enlarging the class of test functions to $H^1(a, b)$, which is the closure of $C^1([a, b])$ in H^1-norm, we may state the **weak** or **variational** formulation of problem (8.12) as follows:

Determine $u \in H^1(a, b)$ such that, $\forall v \in H^1(a, b)$,

$$\int_a^b \{pu'v' + qu'v + ruv\} \, dx = \int_a^b fv \, dx + v(b) B + v(b) A. \tag{8.16}$$

We point out that the Neumann conditions are encoded in equation (8.16), rather than forced by the choice of the test functions, as in the Dirichlet problem.

Introducing the bilinear form

$$B(u, v) = \int_a^b \{pu'v' + qu'v + ruv\} \, dx$$

and the linear functional

$$Lv = \int_a^b fv \, dx + v(b) B + v(a) A,$$

equation (8.16) can be recast in the abstract form

$$B(u, v) = Lv, \quad \forall v \in H^1(a, b).$$

Again, existence, uniqueness and stability of a weak solution follow from the Lax-Milgram Theorem, under rather natural hypotheses on p, q, r, f.

Recall that if $v \in H^1(a, b)$, inequality (7.59) yields

$$v(x) \leq C^* \|v\|_{1,2} \tag{8.17}$$

for every $x \in [a, b]$, with $C^* = \sqrt{2} \max \{(b - a)^{-1/2}, (b - a)^{1/2}\}$.

Proposition 8.2. *Assume that:*

i) $p, q, r \in L^\infty(a, b)$ and $f \in L^2(a, b)$

ii) $p(x) \geq \alpha_0 > 0$, $r(x) \geq c_0 > 0$ a.e. in (a, b) and

$$K_0 \equiv \min\{\alpha_0, c_0\} - \frac{1}{2}\|q\|_{L^\infty} > 0.$$

Then, (8.8) has a unique solution $u \in H^1(a, b)$. Furthermore

$$\|u\|_{1,2} \leq K_0^{-1} \{\|f\|_0 + C^*(|A| + |B|)\}. \tag{8.18}$$

Proof. Let us check that the hypotheses of the Lax-Milgram Theorem hold, with $V = H^1(a, b)$.

Continuity of the bilinear form B. We have:

$$|B(u, v)| \leq \int_a^b \{\|p\|_{L^\infty} |u'v'| + \|q\|_{L^\infty} |u'v| + \|r\|_{L^\infty} |uv|\} \, dx.$$

Using Schwarz's inequality, we easily get

$$|B(u, v)| \leq (\|p\|_{L^\infty} + \|q\|_{L^\infty} + \|r\|_{L^\infty}) \|u\|_{1,2} \|v\|_{1,2}$$

so that B is continuous in V.

Coercivity of B. We have

$$B\left(u, u\right) = \int_a^b \left\{p(u')^2 + qu'u + ru^2\right\} dx.$$

The Schwarz inequality gives

$$\left|\int_a^b qu'u\, dx\right| \leq \|q\|_{L^\infty} \|u'\|_0 \|u\|_0 \leq \frac{1}{2} \|q\|_{L^\infty} \left\{\|u'\|_0^2 + \|u\|_0^2\right\}.$$

Then, by *ii*),

$$B\left(u, u\right) \geq (\alpha_0 - \frac{1}{2} \|q\|_{L^\infty}) \|u'\|_0^2 + (c_0 - \frac{1}{2} \|q\|_{L^\infty}) \|u\|_0^2 \geq K_0 \|u\|_{1,2}^2$$

so that B is $V-$coercive.

Continuity of L in V. Schwarz's inequality and (8.17) yield

$$|Lv| \leq \|f\|_0 \|v\|_0 + |v\left(b\right) B + v\left(a\right) A| \leq$$
$$\leq \left\{\|f\|_0 + C^* \left(|A| + |B|\right)\right\} \|v\|_{1,2}$$

whence $\|L\|_{V^*} \leq \|f\|_0 + C^* \left(|A| + |B|\right)$.

Then, the Lax-Milgram Theorem gives existence, uniqueness and the stability estimate (8.18). □

Remark 8.5. Suppose, $p = 1$, $q = r = 0$. The problem reduces to

$$\begin{cases} u'' = f & \text{in } (a, b) \\ -u'\left(a\right) = A, \quad u'\left(b\right) = B. \end{cases}$$

Hypothesis *ii*) is not satisfied (since $r = 0$). If u is a solution of the problem and $k \in \mathbb{R}$, also $u + k$ is a solution of the same problem. We cannot expect uniqueness. Not even we may prescribe f, A, B arbitrarily, if we want that a solution exists. In fact, integrating the equation $u'' = f$ over (a, b), we deduce that the Neumann data and f must satisfy the *compatibility condition*

$$B + A = \int_a^b f\left(x\right) dx. \tag{8.19}$$

If (8.19) does not hold, *the problem has no solution*. Thus, to have existence and uniqueness we must require that (8.19) holds and select a solution (e.g.) with zero mean value in (a, b). We will return on this kind of solvability questions in Chapter 9.

Robin conditions. Suppose that the boundary conditions in problem (8.12) are:

$$-p\left(a\right) u'\left(a\right) = A, \qquad p\left(b\right) u'\left(b\right) + hu\left(b\right) = B \quad (h > 0, \text{constant})$$

where, for simplicity, the Robin condition is imposed at $x = b$ only. With small adjustments, we may repeat the same computations made for the Neumann conditions (see Problem 8.3). The **weak formulation** is:

Determine $u \in H^1(a, b)$ such that, $\forall v \in H^1(a, b)$,

$$\int_a^b \{pu'v' + qu'v + ruv\} \, dx + hu(b)v(b) = \int_a^b fv \, dx + v(b)B + v(a)A. \quad (8.20)$$

Introducing the bilinear form

$$\tilde{B}(u, v) = \int_a^b \{pu'v' + qu'v + ruv\} \, dx + hu(b)v(b)$$

we may write our problem in the abstract form

$$\tilde{B}(u, v) = Lv \qquad \forall v \in H^1(a, b).$$

We have:

Proposition 8.3. *Assume that i) and ii) of Proposition 8.2 hold and that $h > 0$. Then (8.20) has a unique solution $u \in H^1(a, b)$. Furthermore*

$$\|u\|_{1,2} \le K_0^{-1} \{\|f\|_0 + C^*(|A| + |B|)\}.$$

Proof. Let $V = H^1(a, b)$. Since

$$\tilde{B}(u, u) = B(u, u) + hu^2(b) \ge K_0 \|u\|_{1,2}^2$$

and

$$\left| \tilde{B}(u, v) \right| \le |B(u, v)| + h|u(b)v(b)|$$
$$\le \left(\|p\|_{L^\infty} + \|q\|_{L^\infty} + \|r\|_{L^\infty} + h(C^*)^2 \right) \|u\|_{1,2} \|v\|_{1,2},$$

$\tilde{B}$ is continuous and $V-$coercive. The conclusion follows easily. $\square$

Mixed conditions. The weak formulation of mixed problems does not present particular difficulties. Suppose, for instance, we assign at the end points the conditions

$$u(a) = 0, \qquad p(b)u'(b) = B.$$

Thus, we have a mixed Dirichlet-Neumann problem. The only relevant observation is the choice of the functional setting. Since $u(a) = 0$, we have to choose $V = H_{0,a}^1$, the space of functions $v \in H^1(a, b)$, vanishing at $x = a$. The Poincaré inequality holds in $H_{0,a}^1$ (see Problem 7.21), so that we may choose $\|u'\|_0$ as the norm in $H_{0,a}^1$. Moreover, the following inequality

$$v(x) \le C^{**} \|v'\|_0 \qquad (8.21)$$

holds[2] for every $x \in [a, b]$, with $C^{**} = (b - a)^{1/2}$.

---

[2] Since $v(a) = 0$, we have $v(x) = \int_a^x v'$ so that $|v(x)| \le \sqrt{b - a} \, \|v'\|_0$.

The **weak formulation** is: *Determine $u \in H_{0,a}^1$ such that*

$$\int_a^b \{pu'v' + qu'v + ruv\} \, dx = \int_a^b fv dx + v(b) B, \qquad \forall v \in H_{0,a}^1. \qquad (8.22)$$

We have:

Proposition 8.4. *Assume that i) and ii) of Proposition 8.2 hold. Then, (8.22) has a unique solution $u \in H_{0,a}^1$. Furthermore*

$$\|u'\|_0 \le K_0^{-1} \left\{ C_P \|f\|_0 + C^{**} |B| \right\}.$$

We leave the proof as an exercise.

8.4 Variational Formulation of Poisson's Problem

Guided by the one-dimensional case, we now analyze the variational formulation of Poisson's problem in dimension $n > 1$, starting with a Dirichlet condition.

8.4.1 Dirichlet problem

Let $\Omega \subset \mathbb{R}^n$ be a *bounded domain.* We examine the following problem:

$$\begin{cases} -\alpha \Delta u + a_0(\mathbf{x}) u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \qquad (8.23)$$

where $\alpha > 0$, constant. To achieve a weak formulation, we first assume that a_0 and f are smooth and that $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ is a classical solution of (8.23). We select $C_0^1(\Omega)$ as the space of test functions, having continuous first derivatives and compact support in Ω. In particular, *they vanish in a neighborhood of $\partial\Omega$.* Let $v \in C_0^1(\Omega)$ and multiply the Poisson equation by v. We get

$$\int_\Omega \{-\alpha \Delta u + a_0 u - f\} v \, d\mathbf{x} = 0. \qquad (8.24)$$

Integrating by parts and using the boundary condition, we obtain

$$\int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 uv\} \, d\mathbf{x} = \int_\Omega fv \, d\mathbf{x}, \qquad \forall v \in C_0^1(\Omega). \qquad (8.25)$$

Thus (8.23) implies (8.25).

On the other hand, assume (8.25) is true. Integrating by parts in the reverse order we return to (8.24), which entails $-\alpha \Delta u + a_0 u - f = 0$ in Ω.

Thus, **for classical solutions, the two formulations** (8.23) **and** (8.25) **are equivalent**.

Observe that (8.25) only involves first order derivatives of the solution and of the test function. Then, enlarging the space of test functions to $H_0^1(\Omega)$, closure of $C_0^1(\Omega)$ in the norm $\|u\|_1 = \|\nabla u\|_0$, we may state the **weak** formulation of problem (8.5) as follows:

Determine $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 uv\} \, d\mathbf{x} = \int_\Omega fv \, d\mathbf{x}, \qquad \forall v \in H_0^1(\Omega). \qquad (8.26)$$

Introducing the bilinear form

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

and the linear functional

$$Lv = \int_\Omega fv \, d\mathbf{x},$$

equation (8.26) corresponds to the abstract variational problem

$$B(u,v) = Lv, \quad \forall v \in H_0^1(\Omega).$$

Then, the well-posedness of this problem follows from the the Lax-Milgram Theorem under the hypothesis $a_0 \geq 0$. Precisely:

Theorem 8.1. *Assume that $f \in L^2(\Omega)$ and that $0 \leq a_0(\mathbf{x}) \leq \gamma_0$ a.e. in Ω. Then, problem (8.26) has a unique solution $u \in H_0^1(\Omega)$. Moreover*

$$\|\nabla u\|_0 \leq \frac{C_P}{\alpha} \|f\|_0.$$

Proof. We check that the hypotheses of the Lax-Milgram Theorem hold, with $V = H_0^1(\Omega)$.

Continuity of the bilinear form B. The Schwarz and Poincaré inequalities yield:

$$|B(u,v)| \leq \alpha \|\nabla u\|_0 \|\nabla v\|_0 + \gamma_0 \|u\|_0 \|v\|_0$$
$$\leq (\alpha + C_P^2 \gamma_0) \|\nabla u\|_0 \|\nabla v\|_0$$

so that B is continuous in $H_0^1(\Omega)$.

Coercivity of B. It follows from

$$B(u,u) = \int_\Omega \alpha |\nabla u|^2 \, d\mathbf{x} + \int_\Omega a_0 u^2 d\mathbf{x} \geq \alpha \|\nabla u\|_0^2$$

since $a_0 \geq 0$.

Continuity of L. The Schwarz and Poincaré inequalities give

$$|Lv| = \left| \int_\Omega fv \, d\mathbf{x} \right| \leq \|f\|_0 \|v\|_0 \leq C_P \|f\|_0 \|\nabla v\|_0.$$

Hence $L \in H^{-1}(\Omega)$ and $\|L\|_{H^{-1}(\Omega)} \leq C_P \|f\|_0$. The conclusions follow from the Lax-Milgram Theorem. $\square$

Remark 8.6. Suppose that $c = 0$ and that u represents the equilibrium position of an elastic membrane. Then $B(u, v)$ represents the work done by the elastic internal forces, due to a *virtual displacement* v. On the other hand Lv expresses the work done by the external forces. The weak formulation (8.26) states that these two works balance, which constitutes a version of the *principle of virtual work*.

Furthermore, due to the symmetry of B, the solution u of the problem **minimizes in $H_0^1(\Omega)$ the Dirichlet functional**

$$E(u) = \underbrace{\int_\Omega \alpha |\nabla u|^2 \, d\mathbf{x}}_{\text{internal elastic energy}} - \underbrace{\int_\Omega fu \, d\mathbf{x}}_{\text{external potential energy}}$$

which represents the **total potential energy**. Equation (8.26) constitutes the Euler equation for E.

Thus, in agreement with the principle of virtual work, *u minimizes the potential energy among all the admissible configurations.*

Similar observations can be made for the other types of boundary conditions.

• *Non homogeneous Dirichlet conditions.* Suppose that the Dirichlet condition is $u = g$ on $\partial\Omega$. If Ω is a Lipschitz domain and $g \in H^{1/2}(\partial\Omega)$, then g is the trace on $\partial\Omega$ of a (non unique) function $\tilde{g} \in H^1(\Omega)$, called *extension of g to Ω.* Then, setting

$$w = u - \tilde{g}$$

we are reduced to homogeneous boundary conditions. In fact, $w \in H_0^1(\Omega)$ and is a solution of the equation

$$\int_\Omega \{\alpha \nabla w \cdot \nabla v \, dx + a_0 wv\} \, d\mathbf{x} = \int_\Omega Fv \, d\mathbf{x}, \quad \forall v \in H_0^1(\Omega)$$

where $F = f - \alpha \nabla \tilde{g} - a_0 \tilde{g} \in L^2(\Omega)$. The Lax-Milgram Theorem yields existence, uniqueness and the stability estimate

$$\|\nabla w\|_0 \leq \frac{C_P}{\alpha}\{\|f\|_0 + (\alpha + a_0)\|\tilde{g}\|_{1,2}\} \tag{8.27}$$

for any extension $\tilde{g}$ of g. Since $\|u\|_{1,2} \leq \|w\|_{1,2} + \|\tilde{g}\|_{1,2}$ and recalling that (subsection 7.9.3)

$$\|g\|_{H^{1/2}(\partial\Omega)} = \inf\left\{\|\tilde{g}\|_{1,2} : \tilde{g} \in H^1(\Omega), \ \tilde{g}_{|\partial\Omega} = g\right\},$$

taking the lowest upper bound with respect to $\tilde{g}$, from (8.27) we deduce, in terms of u:

$$\|u\|_{1,2} \leq C(\alpha, \gamma_0, n, \Omega)\left\{\|f\|_0 + \|g\|_{H^{1/2}(\partial\Omega)}\right\}.$$

8.4.2 Neumann, Robin and mixed problems

Let $\Omega \subset \mathbb{R}^n$ be a bounded, *Lipschitz* domain. We examine the following problem:

$$\begin{cases} -\alpha \Delta u + a_0 (\mathbf{x}) u = f & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u = g & \text{on } \partial \Omega \end{cases} \qquad (8.28)$$

where $\alpha > 0$ is constant and $\boldsymbol{\nu}$ denotes the outward normal unit vector to $\partial \Omega$. As usual, to derive a weak formulation, we first assume that a_0, f and g are smooth and that $u \in C^2 (\Omega) \cap C^1 (\overline{\Omega})$ is a classical solution of (8.28). We choose $C^1 (\overline{\Omega})$ as the space of test functions, having continuous first derivatives up to $\partial \Omega$. Let $v \in C^1 (\overline{\Omega})$, arbitrary, and multiply the Poisson equation by v. Integrating over Ω, we get

$$\int_\Omega \{-\alpha \Delta u + a_0 u\} v \ d\mathbf{x} = \int_\Omega f v \ d\mathbf{x}. \qquad (8.29)$$

An integration by parts gives

$$- \int_{\partial \Omega} \alpha \partial_{\boldsymbol{\nu}} u \ v d\sigma + \int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 u v\} \ d\mathbf{x} = \int_\Omega f v \ d\mathbf{x}, \quad \forall v \in C^1 (\overline{\Omega}). \qquad (8.30)$$

Using the Neumann condition we may write

$$\int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 u v\} \ d\mathbf{x} = \int_\Omega f v \ d\mathbf{x} + \alpha \int_{\partial \Omega} g v \ d\sigma \qquad \forall v \in C^1 (\overline{\Omega}). \qquad (8.31)$$

Thus (8.28) implies (8.31).

On the other hand, suppose that (8.31) is true. Integrating by parts in the reverse order, we find

$$\int_\Omega \{-\alpha \Delta u + a_0 u - f\} v \ d\mathbf{x} + \int_{\partial \Omega} \alpha \partial_{\boldsymbol{\nu}} u \ v d\sigma = \alpha \int_{\partial \Omega} g v \ d\sigma, \qquad (8.32)$$

for every $\forall v \in C^1 (\overline{\Omega})$. Since $C_0^1 (\Omega) \subset C^1 (\overline{\Omega})$ we may insert any $v \in C_0^1 (\Omega)$ into (8.32), to get

$$\int_\Omega \{-\alpha \Delta u + a_0 u - f\} v \ d\mathbf{x} = 0.$$

The arbitrariness of $v \in C_0^1 (\Omega)$ entails $-\alpha \Delta u + a_0 u - f = 0$ in Ω. Therefore (8.32) becomes

$$\int_{\partial \Omega} \partial_{\boldsymbol{\nu}} u \ v d\sigma = \int_{\partial \Omega} g v \ d\sigma \qquad \forall v \in C^1 (\overline{\Omega})$$

and the arbitrariness of $v \in C^1 (\overline{\Omega})$ forces $\partial_{\boldsymbol{\nu}} u = g$, recovering the Neumann condition as well.

Thus, **for classical solutions, the two formulations** (8.28) and (8.31) **are equivalent**.

Recall now that, by Theorem 7.10, $C^1 (\overline{\Omega})$ is dense in $H^1 (\Omega)$, which therefore constitutes the natural Sobolev space for the Neumann problem. Then, enlarging the space of test functions to $H^1 (\Omega)$, we may give the **weak** formulation of problem (8.5) as follows:

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

$$\int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 uv\}\, d\mathbf{x} = \int_\Omega fv d\mathbf{x} + \alpha \int_{\partial\Omega} gv d\sigma, \qquad \forall v \in H^1(\Omega). \qquad (8.33)$$

Again we point out that the Neumann condition is encoded in (8.33) and not explicitly expressed as in the case of Dirichlet boundary conditions. Since we used the density of $C^1(\overline{\Omega})$ in $H^1(\Omega)$ and the trace of v on $\partial\Omega$, some regularity of the domain (Lipschitz is enough) is needed, even in the variational formulation. Introducing the bilinear form

$$B(u, v) = \int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 uv\}\, d\mathbf{x} \qquad (8.34)$$

and the linear functional

$$Lv = \int_\Omega fv\, d\mathbf{x} + \alpha \int_{\partial\Omega} gv d\sigma, \qquad (8.35)$$

(8.33) may be formulated as the abstract variational problem

$$B(u, v) = Lv, \quad \forall v \in H_0^1(\Omega).$$

The following theorem states the well-posedness of this problem under reasonable hypotheses on the data. Recall from Theorem 7.11 the *trace inequality*

$$\|v\|_{L^2(\partial\Omega)} \leq \overline{C}(n, \Omega) \|v\|_{1,2}. \qquad (8.36)$$

Theorem 8.2. *Let* $\Omega \subset \mathbb{R}^n$ *be a bounded, Lipschitz domain,* $f \in L^2(\Omega)$, $g \in L^2(\partial\Omega)$ *and* $0 < c_0 \leq a_0(\mathbf{x}) \leq \gamma_0$ *a.e. in* Ω.
  Then, problem (8.33) has a unique solution $u \in H^1(\Omega)$. *Moreover,*

$$\|u\|_{1,2} \leq \frac{1}{\min\{\alpha, c_0\}} \left\{\|f\|_0 + \overline{C}\alpha \|g\|_{L^2(\partial\Omega)}\right\}.$$

Proof. We check that the hypotheses of the Lax-Milgram Theorem hold, with $V = H^1(\Omega)$.
  Continuity of the bilinear form B. The Schwarz inequality yields:

$$|B(u, v)| \leq \alpha \|\nabla u\|_0 \|\nabla v\|_0 + \gamma_0 \|u\|_0 \|v\|_0$$
$$\leq (\alpha + \gamma_0) \|u\|_{1,2} \|v\|_{1,2}$$

so that B is continuous in $H^1(\Omega)$.
  Coercivity of B. It follows from

$$B(u, u) = \int_\Omega \alpha |\nabla u|^2\, d\mathbf{x} + \int_\Omega a_0 u^2 d\mathbf{x} \geq \min\{\alpha, c_0\} \|u\|_{1,2}^2$$

since $a_0(\mathbf{x}) \geq c_0 > 0$ a.e. in Ω.

Continuity of L. From Schwarz's inequality and (8.36) we get:

$$|Lv| \leq \left| \int_{\Omega} fv \, d\mathbf{x} \right| + \alpha \left| \int_{\partial\Omega} gv \, d\sigma \right| \leq \|f\|_0 \|v\|_0 + \alpha \|g\|_{L^2(\partial\Omega)} \|v\|_{L^2(\partial\Omega)}$$

$$\leq \left\{ \|f\|_0 + \overline{C}\alpha \|g\|_{L^2(\partial\Omega)} \right\} \|v\|_{1,2} \, .$$

Therefore L is continuous in $H^1(\Omega)$ with

$$\|L\|_{H^1(\Omega)^*} \leq \|f\|_{L^2(\Omega)} + \overline{C}\alpha \|g\|_{L^2(\partial\Omega)} \, .$$

The conclusion follows from the Lax-Milgram Theorem. $\square$

Remark 8.7. As in the one-dimensional case, without the condition $a_0(\mathbf{x}) \geq c_0 > 0$, neither the existence nor the uniqueness of a solution is guaranteed. Let, for example, $a_0 = 0$. Then two solutions of the same problem differ by a constant. A way to restore uniqueness is to select a solution with, e.g., zero mean value, that is

$$\int_{\Omega} u(\mathbf{x}) \, d\mathbf{x} = 0.$$

The existence of a solution requires the following *compatibility condition* on the data f and g:

$$\int_{\Omega} f \, d\mathbf{x} + \alpha \int_{\partial\Omega} g \, d\sigma = 0, \tag{8.37}$$

obtained by substituting $v = 1$ into the equation

$$\int_{\Omega} \alpha \nabla u \cdot \nabla v \, dx = \int_{\Omega} fv \, d\mathbf{x} + \alpha \int_{\partial\Omega} gv \, d\sigma.$$

Note that, since Ω is bounded, the function $v = 1$ belongs to $H^1(\Omega)$.

If $a_0 = 0$ and (8.37) does not hold, problem (8.28) has no solution. Viceversa, we shall see later that, if this condition is fulfilled, a solution exists.

If $g = 0$, (8.37) has a simple interpretation. Indeed problem (8.28) is a model for the equilibrium configuration of a membrane whose boundary is free to slide along a vertical guide. The compatibility condition $\int_{\Omega} f d\mathbf{x} = 0$ expresses the obvious fact that, at equilibrium, the resultant of the external loads must vanish.

Robin problem. The same arguments leading to the weak formulation of the Neumann problem (8.28) may be used for the problem

$$\begin{cases} -\alpha \Delta u + a_0(\mathbf{x}) u = f & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u + hu = g & \text{on } \partial\Omega. \end{cases} \tag{8.38}$$

The weak formulation comes from (8.30), observing that

$$\partial_{\boldsymbol{\nu}} u = -hu + g \quad \text{on } \partial\Omega.$$

We obtain the following **variational formulation**:

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

$$\int_\Omega \{\alpha \nabla u \cdot \nabla v + a_0 uv\} \ d\mathbf{x} + \alpha \int_{\partial \Omega} huv \ d\sigma = \int_\Omega fv \ d\mathbf{x} + \alpha \int_{\partial \Omega} g \ d\sigma \quad \forall v \in H^1(\Omega).$$

We have:

Theorem 8.3. *Let Ω, f, g and a_0 be as in Theorem 8.2 and $0 \leq h(\mathbf{x}) \leq h_0$ a.e. on $\partial \Omega$. Then, problem (8.38) has a unique weak solution $u \in H^1(\Omega)$. Moreover*

$$\|u\|_{1,2} \leq \frac{1}{\min\{\alpha, c_0\}} \left\{ \|f\|_0 + \overline{C}\alpha \|g\|_{L^2(\partial \Omega)} \right\}.$$

Proof. Introducing the bilinear form

$$\tilde{B}(u, v) = B(u, v) + \alpha \int_{\partial \Omega} huv \ d\sigma$$

the variational formulation becomes

$$\tilde{B}(u, v) = Lv \qquad \forall v \in H^1(\Omega)$$

where B and L are defined in (8.34) and (8.35), respectively.

From the Schwarz inequality and (8.36), we infer

$$\left| \int_{\partial \Omega} huv \ d\sigma \right| \leq h_0 \|u\|_{L^2(\partial \Omega)} \|v\|_{L^2(\partial \Omega)} \leq \overline{C}^2 h_0 \|u\|_{1,2} \|v\|_{1,2}.$$

On the other hand, the positivity of α, a_0 and h entails that

$$\tilde{B}(u, u) \geq B(u, u) \geq \min\{\alpha, c_0\} \|u\|_{1,2}^2.$$

The conclusions follow easily. $\square$

Mixed Dirichlet-Neumann problem. Let Γ_D be a non empty relatively open subset of $\partial \Omega$. Set $\Gamma_N = \partial \Omega \backslash \Gamma_D$ and consider the problem

$$\begin{cases} -\alpha \Delta u + a_0(\mathbf{x}) u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \partial_\nu u = g & \text{on } \Gamma_N. \end{cases}$$

The correct functional setting is $H^1_{0,\Gamma_D}(\Omega)$, i.e. the set of functions in $H^1(\Omega)$ with zero trace on Γ_D. From Theorem 7.16, the Poincaré inequality holds in $H^1_{0,\Gamma_D}(\Omega)$ and therefore we may choose the norm

$$\|u\|_{H^1_{0,\Gamma_D}(\Omega)} = \|\nabla u\|_0.$$

From (8.29) and the Gauss formula, we obtain, since $u = 0$ on Γ_D,

$$-\int_{\Gamma_N} \alpha \partial_{\boldsymbol{\nu}} u \, v d\sigma + \int_{\Omega} \{\alpha \nabla u \cdot \nabla v + a_0 uv\} \, d\mathbf{x} = \int_{\Omega} fv \, d\mathbf{x}, \quad \forall v \in C^1\left(\overline{\Omega}\right).$$

The Neumann condition on Γ_N, yields the following **variational formulation**:

Determine $u \in H^1_{0,\Gamma_D}(\Omega)$ such that, $\forall v \in H^1_{0,\Gamma_D}(\Omega)$,

$$\int_{\Omega} \alpha \nabla u \cdot \nabla v \, dx + \int_{\Omega} a_0 uv \, d\mathbf{x} = \int_{\Omega} fv \, d\mathbf{x} + \alpha \int_{\Gamma_N} gv \, d\sigma.$$

Using the *trace* inequality (Theorem 7.12)

$$\|v\|_{L^2(\Gamma_N)} \leq \widetilde{C} \|v\|_{1,2}, \tag{8.39}$$

the proof of the next theorem follows the usual pattern.

Theorem 8.4. *Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain. Assume $f \in L^2(\Omega)$, $g \in L^2(\Gamma_N)$ and $0 \leq a_0(\mathbf{x}) \leq \gamma_0$ a.e. in Ω. Then the mixed problem has a unique solution $u \in H^1_{0,\Gamma_D}(\Omega)$. Moreover:*

$$\|\nabla u\|_0 \leq \frac{1}{\alpha} \|f\|_0 + \widetilde{C} \|g\|_{L^2(\Gamma_N)}.$$

8.4.3 Eigenvalues of the Laplace operator

In subsection 6.9.2 we have seen how the efficacy of the separation of variables method for a given problem relies on the existence of a basis of eigenfunctions associated with that problem. The abstract results in subsection 6.9.4, concerning the spectrum of a weakly coercive bilinear form, constitute the appropriate tools for analyzing the spectral properties of uniformly elliptic operators and in particular of the Laplace operator. It is important to point out that the spectrum of a differential operator **must be associated with specific homogeneous boundary conditions**.

Thus, for instance, we may consider the *Dirichlet eigenfunctions for the Laplace operator in a domain Ω*, i.e. the non trivial solutions of the problem

$$\begin{cases} -\Delta u = \lambda u & \text{in } \Omega \\ \quad u = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.40}$$

A weak solution of problem (8.40) is a function $u \in H^1_0(\Omega)$ such that

$$a(u,v) \equiv (\nabla u, \nabla v)_0 = \lambda (u,v)_0 \qquad \forall v \in H^1_0(\Omega).$$

If Ω is bounded, the bilinear form is $H^1_0(\Omega)$−coercive so that Theorem 6.15 gives:

Theorem 8.5. *Let Ω be a bounded domain. Then, there exists in $L^2(\Omega)$ an orthonormal basis $\{u_k\}_{k \geq 1}$ consisting of Dirichlet eigenfunctions for the Laplace operator. The corresponding eigenvalues $\{\lambda_k\}_{k \geq 1}$ are all positive and may be arranged in an increasing sequence*

$$0 < \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_k \leq \cdots$$

with $\lambda_k \to +\infty$.

The sequence $\left\{ u_k / \sqrt{\lambda_k} \right\}_{k \geq 1}$ constitutes an orthonormal basis in $H_0^1(\Omega)$, with respect to the scalar product $(u, v)_1 = (\nabla u, \nabla v)_0$.

Remark 8.8. Let $u \in L^2(\Omega)$ and denote by $c_k = (u, u_k)_0$ the Fourier coefficients of u with respect to the orthonormal basis $\{u_k\}_{k \geq 1}$. Then we may write

$$u = \sum_{k=1}^{\infty} c_k u_k \qquad \text{and} \qquad \|u\|_0^2 = \sum_{k=1}^{\infty} c_k^2.$$

Note that

$$\|\nabla u_k\|_0^2 = (\nabla u_k, \nabla u_k)_0 = \lambda_k (u_k, u_k)_0 = \lambda_k.$$

Thus, $u \in H_0^1(\Omega)$ if and only if

$$\|\nabla u\|_0^2 = \sum_{k=1}^{\infty} \lambda_k c_k^2 < \infty. \tag{8.41}$$

Moreover, (8.41) implies that, for every $u \in H_0^1(\Omega)$,

$$\|\nabla u\|_0^2 \geq \lambda_1 \sum_{k=1}^{\infty} c_k^2 = \lambda_1 \|u\|_0^2.$$

We deduce the following **variational principle for the first Dirichlet eigenvalue**:

$$\lambda_1 = \min \left\{ \frac{\int_\Omega |\nabla u|^2}{\int_\Omega u^2} : u \in H_0^1(\Omega),\ u \text{ non identically zero.} \right\} \tag{8.42}$$

The quotient in (8.42) is called *Raiyeigh's quotient.*

If the domain Ω is smooth, it can be shown that λ_1 is *simple*, i.e. the corresponding eigenspace has dimension 1, and that **the corresponding normalized eigenvector u_1 is either strictly positive or strictly negative in Ω.**

Similar theorems hold for the other types of boundary value problems as well. For instance, the *Neumann eigenfunctions for the Laplace operator in Ω* are the non trivial solutions of the problem

$$\begin{cases} -\Delta u = \mu u & \text{in } \Omega \\ \partial_\nu u = 0 & \text{on } \partial\Omega. \end{cases}$$

Applying Theorem 6.15 we find:

Theorem 8.6. *If Ω is a bounded Lipschitz domain, there exists in $L^2(\Omega)$ an orthonormal basis $\{u_k\}_{k\geq1}$ consisting of Neumann eigenfunctions for the Laplace operator. The corresponding eigenvalues form a non decreasing sequence $\{\mu_k\}_{k\geq1}$, with $\mu_1 = 0$ and $\mu_k \to +\infty$.*

Moreover, the sequence $\{u_k/\sqrt{\mu_k+1}\}_{k\geq1}$ constitutes an orthonormal basis in $H^1(\Omega)$, with respect to the scalar product $(u,v)_{1,2} = (u,v)_0 + (\nabla u, \nabla v)_0$.

8.4.4 An asymptotic stability result

The results in the last subsection may be used sometimes to prove the asymptotic stability of a steady state solution of an evolution equation as time $t \to +\infty$.

As an example consider the following problem for the heat equation. Suppose that $u \in C^{2,1}\left(\overline{\Omega} \times [0,+\infty)\right)$ is the (unique) solution of

$$\begin{cases} u_t - \Delta u = f(\mathbf{x}) & \mathbf{x} \in \Omega,\ t > 0 \\ u(\mathbf{x},0) = u_0(\mathbf{x}) & \mathbf{x} \in \Omega \\ u(\boldsymbol{\sigma},t) = 0 & \boldsymbol{\sigma} \in \partial\Omega,\ t > 0 \end{cases}$$

where Ω is a smooth, bounded domain. Denote by $u_\infty = u_\infty(\mathbf{x})$ the solution of the stationary problem

$$\begin{cases} -\Delta u_\infty = f & \text{in } \Omega \\ u_\infty = 0 & \text{on } \partial\Omega. \end{cases}$$

Proposition 8.5. *For $t \geq 0$, we have*

$$\|u(\cdot,t) - u_\infty\|_0 \leq e^{-\lambda_1 t}\left\{C_P^2\|f\|_0 + \|u_0\|_0\right\} \tag{8.43}$$

where λ_1 is the first Dirichlet eigenvalue for the Laplace operator in Ω.

Proof. Set $g(\mathbf{x}) = u_0(\mathbf{x}) - u_\infty(\mathbf{x})$. The function $w(\mathbf{x},t) = u(\mathbf{x},t) - u_\infty(\mathbf{x})$ solves the problem

$$\begin{cases} w_t - \Delta w = 0 & \mathbf{x} \in \Omega,\ t > 0 \\ w(\mathbf{x},0) = g(\mathbf{x}) & \mathbf{x} \in \Omega \\ w(\boldsymbol{\sigma},t) = 0 & \boldsymbol{\sigma} \in \partial\Omega,\ t > 0. \end{cases} \tag{8.44}$$

Let us use the method of separation of variables and look for solutions of the form $w(\mathbf{x},t) = v(\mathbf{x})z(t)$. We find

$$\frac{z'(t)}{z(t)} = \frac{\Delta v(\mathbf{x})}{v(\mathbf{x})} = -\lambda$$

with λ constant. Thus we are lead to the eigenvalue problem

$$\begin{cases} -\Delta v = \lambda v & \text{in } \Omega \\ v = 0 & \text{on } \partial\Omega. \end{cases}$$

From Theorem 8.5, there exists in $L^2(\Omega)$ an orthonormal basis $\{u_k\}_{k \geq 1}$ consisting of eigenvectors, corresponding to a sequence of non decreasing eigenvalues $\{\lambda_k\}$, with $\lambda_1 > 0$ and $\lambda_k \to +\infty$. Then, if $g_k = (g, u_k)_0$, we can write

$$g = \sum_1^\infty g_k u_k \qquad \text{and} \qquad \|g\|_0^2 = \sum_{k=1}^\infty g_k^2.$$

As a consequence, we find $z_k(t) = e^{-\lambda_k t}$ and finally

$$w(\mathbf{x}, t) = \sum_1^\infty e^{-\lambda_k t} g_k u_k(\mathbf{x}).$$

Thus,

$$\|u(\cdot, t) - u_\infty\|_0^2 = \|w(\cdot, t)\|_0^2$$
$$= \sum_{k=1}^\infty e^{-2\lambda_k t} g_k^2$$

and since $\lambda_k > \lambda_1$ for every k, we deduce that

$$\|u(\cdot, t) - u_\infty\|_0^2 \leq \sum_{k=1}^\infty e^{-2\lambda_1 t} g_k^2 = e^{-2\lambda_1 t} \|g\|_0^2.$$

Theorem 8.1 yields, in particular

$$\|u_\infty\|_0 \leq C_P^2 \|f\|_0,$$

and hence

$$\|g\|_0 \leq \|u_0\|_0 + \|u_\infty\|_0$$
$$\leq \|u_0\|_0 + C_P^2 \|f\|_0$$

giving (8.43). $\square$

Proposition 8.5 implies that the steady state u_∞ is *asymptotically stable in* $L^2(\Omega)-$norm as $t \to +\infty$. The speed of convergence is exponential[3] and it is determined by the first eigenvalue λ_1.

8.5 General Equations in Divergence Form

8.5.1 Basic assumptions

In this section we consider boundary value problems for elliptic operators with general diffusion and transport terms. Let $\Omega \subset \mathbb{R}^n$ be a **bounded domain** and set

$$\mathcal{E}u = -\text{div}\,(\mathbf{A}(\mathbf{x})\nabla u - \mathbf{b}(\mathbf{x})u) + \mathbf{c}(\mathbf{x}) \cdot \nabla u + a_0(\mathbf{x})u \tag{8.45}$$

where $\mathbf{A} = (a_{ij})_{i,j=1,\ldots,n}$, $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{c} = (c_1, \ldots, c_n)$ and a_0 is a real function.

---

[3] Compare with subsection 2.1.4.

Throughout this section, we will assume that the following hypotheses hold.

1. The differential operator $\mathcal{E}$ is **uniformly elliptic**, i.e. there exist **positive numbers** α and M such that:

$$\alpha \left|\boldsymbol{\xi}\right|^2 \leq \mathbf{A}\left(\mathbf{x}\right) \boldsymbol{\xi} \cdot \boldsymbol{\xi} \leq M \left|\boldsymbol{\xi}\right|^2, \quad \forall \boldsymbol{\xi} \in \mathbb{R}^n, \text{ a.e. in } \Omega. \qquad (8.46)$$

2. The coefficients $\mathbf{b}$, $\mathbf{c}$ and a_0 are all **bounded**:

$$\left|\mathbf{b}\left(\mathbf{x}\right)\right| \leq \beta, \quad \left|\mathbf{c}\left(\mathbf{x}\right)\right| \leq \gamma, \quad \left|a_0\left(\mathbf{x}\right)\right| \leq \gamma_0, \quad \text{a.e. in } \Omega. \qquad (8.47)$$

The uniform ellipticity condition (8.46) states that $\mathbf{A}$ is *positive*[4] in Ω with the minimum eigenvalue bounded from below by α, called *ellipticity constant*, and the maximum eigenvalue bounded from above by M. We point out that at this level of generality, we allow discontinuities also of the diffusion matrix $\mathbf{A}$, of the transport coefficients $\mathbf{b}$ and $\mathbf{c}$, in addition to the reaction coefficient a_0.

We want to extend to these type of operators the theory developed so far. The uniform ellipticity is a necessary requirement. In this section, we first indicate some sufficient conditions assuring the well-posedness of the usual boundary value problems, based on the use of the Lax-Milgram Theorem.

On the other hand, these conditions may be sometimes considered rather restrictive. When they are not satisfied, precise information on solvability and well-posedness can be obtained from Theorem 6.12.

As in the preceding sections, we start from the Dirichlet problem.

8.5.2 Dirichlet problem

Consider the problem

$$\begin{cases} \mathcal{E}u = f + \operatorname{div} \mathbf{f} & \text{in } \Omega \\ \quad u = 0 & \text{on } \partial\Omega \end{cases} \qquad (8.48)$$

where $f \in L^2\left(\Omega\right)$ e $\mathbf{f} \in L^2\left(\Omega; \mathbb{R}^n\right)$.

A comment on the right hand side of (8.48) is in order. We have denoted by $H^{-1}\left(\Omega\right)$ the dual of $H_0^1\left(\Omega\right)$. We know (Theorem 7.6) that every element $F \in H^{-1}\left(\Omega\right)$ can be identified with an element in $\mathcal{D}'\left(\Omega\right)$ of the form

$$F = f + \operatorname{div} \mathbf{f}.$$

Moreover

$$\|F\|_{H^{-1}(\Omega)} \leq \|f\|_0 + \|\mathbf{f}\|_0. \qquad (8.49)$$

Thus, the right hand side of (8.48) represents a generic element of $H^{-1}\left(\Omega\right)$.

---

[4] If $\mathbf{A}$ is only *nonnegative*, the equation is *degenerate elliptic* and things get too complicated for this introductory book.

As in Section 8.4, to derive a variational formulation of (8.48), we first assume that all the coefficients and the data f, $\mathbf{f}$ are smooth. Then, we multiply the equation by a test function $v \in C_0^1(\Omega)$ and integrate over Ω:

$$\int_\Omega [-\mathrm{div}(\mathbf{A}\nabla u - \mathbf{b}u)\, v]\, d\mathbf{x} + \int_\Omega [\mathbf{c}{\cdot}\nabla u + a_0 u]\, v\, d\mathbf{x} = \int_\Omega [f + \mathrm{div}\mathbf{f}]\, v d\mathbf{x}.$$

Integrating by parts, we find, since $v = 0$ on $\partial\Omega$:

$$\int_\Omega [-\mathrm{div}(\mathbf{A}\nabla u - \mathbf{b}u)\, v]\, d\mathbf{x} = \int_\Omega [\mathbf{A}\nabla u \cdot \nabla v\, - \mathbf{b}u \cdot \nabla v]\, d\mathbf{x}$$

and

$$\int_\Omega v\, \mathrm{div}\, \mathbf{f}\, d\mathbf{x} = -\int_\Omega \mathbf{f}{\cdot}\nabla v\, d\mathbf{x}.$$

Thus, the resulting equation is:

$$\int_\Omega \{\mathbf{A}\nabla u \cdot \nabla v - \mathbf{b}u \cdot \nabla v + \mathbf{c}{\cdot}\nabla u\, + a_0 uv\}\, d\mathbf{x} = \int_\Omega \{fv - \mathbf{f}{\cdot}\nabla v\, \}\, d\mathbf{x} \qquad (8.50)$$

for every $v \in C_0^1(\Omega)$.

It is not difficult to check that **for classical solutions, the two formulations** (8.48) and (8.50) **are equivalent**.

We now enlarge the space of test functions to $H_0^1(\Omega)$ and introduce the bilinear form

$$B(u, v) = \int_\Omega \{\mathbf{A}\nabla u \cdot \nabla v - \mathbf{b}u \cdot \nabla v + \mathbf{c}{\cdot}\nabla u\, + a_0 uv\}\, d\mathbf{x}$$

and the linear functional

$$Fv = \int_\Omega \{fv - \mathbf{f}{\cdot}\nabla v\, \}\, d\mathbf{x}.$$

Then, the **weak formulation** of problem (8.48) is the following:

Determine $u \in H_0^1(\Omega)$ such that

$$B(u, v) = Fv, \qquad \forall v \in H_0^1(\Omega). \qquad (8.51)$$

A set of hypotheses that ensure the well-posedness of the problem is indicated in the following proposition.

Proposition 8.6. *Assume that hypotheses (8.46) and (8.47) hold and that $f \in L^2(\Omega)$, $\mathbf{f} \in L^2(\Omega; \mathbb{R}^n)$. Then if $\mathbf{b}$ and $\mathbf{c}$ have Lipschitz components and*

$$\frac{1}{2}\mathrm{div}\,(\mathbf{b} - \mathbf{c}) + a_0 \geq 0, \text{ a.e. in } \Omega, \qquad (8.52)$$

problem (8.51) has a unique solution. Moreover, the following stability estimate holds:

$$\|u\|_1 \leq \frac{1}{\alpha}\{\|f\|_0 + \|\mathbf{f}\|_0\}. \qquad (8.53)$$

Proof. We apply the Lax-Milgram Theorem with $V = H_0^1(\Omega)$. The continuity of B in V follows easily. In fact, the Schwarz inequality and the bound in (8.46) give:

$$\left| \int_\Omega \mathbf{A}\nabla u \cdot \nabla v \, d\mathbf{x} \right| \leq \int_\Omega \sum_{i,j=1}^n \left| a_{ij} \partial_{x_i} u \, \partial_{x_j} v \right| d\mathbf{x}$$

$$\leq M \int_\Omega |\nabla u| \, |\nabla v| \, d\mathbf{x} \leq M \|\nabla u\|_0 \|\nabla v\|_0 .$$

Moreover, using (8.46) and Poincaré's inequality as well, we get

$$\left| \int_\Omega [\mathbf{b}u \cdot \nabla v - \mathbf{c}v \cdot \nabla u] \, d\mathbf{x} \right| \leq (\beta + \gamma) C_P \|\nabla u\|_0 \|\nabla v\|_0$$

and

$$\left| \int_\Omega a_0 uv \, d\mathbf{x} \right| \leq \gamma_0 \int_\Omega |u| \, |v| \, d\mathbf{x} \leq \gamma_0 C_P^2 \|\nabla u\|_0 \|\nabla v\|_0 .$$

Thus, we can write

$$|B(u,v)| \leq \left(M + (\beta + \gamma)C_p + \gamma C_p^2 \right) \|\nabla u\|_0 \|\nabla v\|_0$$

which shows the continuity of B. Let us analyze the coercivity of B. We have:

$$B(u,u) = \int_\Omega \left\{ \mathbf{A}\nabla u \cdot \nabla u - (\mathbf{b} - \mathbf{c})u \cdot \nabla u + a_0 u^2 \right\} d\mathbf{x}.$$

Observe that, since $u = 0$ on $\partial\Omega$, integrating by parts we obtain

$$\int_\Omega (\mathbf{b} - \mathbf{c})u \cdot \nabla u \, d\mathbf{x} = \frac{1}{2} \int_\Omega (\mathbf{b} - \mathbf{c}) \cdot \nabla u^2 d\mathbf{x} = -\frac{1}{2} \int_\Omega \operatorname{div}(\mathbf{b} - \mathbf{c}) \, u^2 d\mathbf{x}.$$

Therefore, from (8.46) and (8.52), it follows that

$$B(u,u) \geq \alpha \int_\Omega |\nabla u|^2 \, d\mathbf{x} + \int_\Omega \left[\frac{1}{2}\operatorname{div}(\mathbf{b} - \mathbf{c}) + a_0 \right] u^2 d\mathbf{x} \geq \alpha \|\nabla u\|_0^2$$

so that B is $V-$coercive. Since we already know that $F \in H^{-1}(\Omega)$, the Lax-Milgram Theorem and (8.49) give existence, uniqueness and the stability estimate (8.53). $\square$

Remark 8.9. If $\mathbf{A}$ is symmetric and $\mathbf{b} = \mathbf{c} = \mathbf{0}$, the solution u is a minimizer in $H_0^1(\Omega)$ for the "energy" functional

$$E(u) = \int_\Omega \left\{ \mathbf{A}\nabla u \cdot \nabla u + cu^2 - fu \right\} d\mathbf{x}.$$

As in Remark 8.6, equation (8.51) constitutes the Euler equation for E.

Remark 8.10. If the Dirichlet condition is nonhomogeneous, i.e.

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

with $g \in H^{1/2}(\partial\Omega)$, we consider an *extension* $\tilde{g}$ of g in $H^1(\Omega)$ and set $w = u - \tilde{g}$. In this case we require that Ω is at least a Lipschitz domain, to ensure the existence of $\tilde{g}$. Then $w \in H_0^1(\Omega)$ and solves the equation
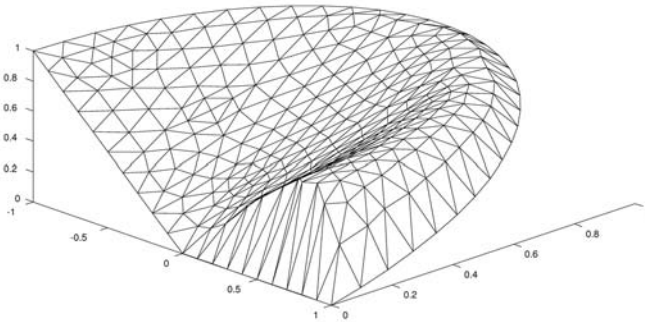
$$\mathcal{E}w = f + \text{div}(\mathbf{f} + \mathbf{A}\nabla\tilde{g} - \mathbf{b}\tilde{g}) - \mathbf{c} \cdot \nabla\tilde{g} - c\tilde{g}.$$

From (8.46) and (8.47) we have

$$\mathbf{c} \cdot \nabla\widehat{g} + c\widehat{g} \in L^2(\Omega) \text{ and } \mathbf{A}\nabla\widehat{g} - \mathbf{b}\widehat{g} \in L^2(\Omega; \mathbb{R}^n).$$

Therefore, the Lax-Milgram Theorem yields existence, uniqueness and the estimate

$$\|u\|_{1,2} \leq C(\alpha, n, M, \beta, \gamma, \gamma_0, \Omega)\left\{\|f\|_0 + \|\mathbf{f}\|_0 + \|g\|_{H^{1/2}(\partial\Omega)}\right\}.$$

Fig. 8.1. The solution of problem (8.54)

Example 8.1. Figure 8.1 shows the solution of the following Dirichlet problem in the upper half circle:

$$\begin{cases} -u_{\rho\rho} - \rho^{-1}u_\rho - \rho u_\theta = 0 & \rho < 1, 0 < \theta < \pi \\ u(1, \theta) = \sin(\theta/2) & 0 \leq \theta \leq \pi \\ u(\rho, 0) = 0, \ u(\rho, \pi) = \rho & \rho \leq 1 \end{cases} \tag{8.54}$$

where (ρ, θ) denotes polar coordinates. Note that, in rectangular coordinates,

$$-\rho u_\theta = y u_x - x u_y \tag{8.55}$$

so that it represents a transport term of the type $\mathbf{c} \cdot \nabla u$ with $\mathbf{c} = (y, -x)$. Since div$\mathbf{b} = 0$, Proposition 8.6 ensures the well posedness of the problem.

Alternative for the Dirichlet problem. We will see later that problem (8.48) is actually well posed under the condition

$$\text{div}\mathbf{b}+a_0 \geq 0,$$

which does not involve the coefficient $\mathbf{c}$.

In particular, this condition is fulfilled if $a_0\left(\mathbf{x}\right) \geq 0$ and $\mathbf{b}\left(\mathbf{x}\right) = \mathbf{0}$ a.e. in Ω. In general however, we cannot prove that the bilinear form B is coercive. What we may affirm is that B is **weakly coercive**, i.e. *there exists $\lambda_0 \in \mathbb{R}$ such that*:

$$\tilde{B}\left(u,v\right) = B\left(u,v\right) + \lambda_0 \left(u,v\right)_0 \equiv B\left(u,v\right) + \lambda_0 \int_\Omega uv \; d\mathbf{x}$$

is coercive. In fact, from the elementary inequality

$$|ab| \leq \varepsilon a^2 + \frac{1}{4\varepsilon}b^2, \quad \forall \varepsilon > 0,$$

we get

$$\left| \int_\Omega (\mathbf{b} - \mathbf{c})u\cdot\nabla u \; d\mathbf{x} \right| \leq (\beta + \gamma) \int_\Omega |u\cdot\nabla u| \, d\mathbf{x} \leq \varepsilon \left\| \nabla u \right\|_0^2 + \frac{(\beta + \gamma)^2}{4\varepsilon} \left\| u \right\|_0^2.$$

Therefore:

$$\tilde{B}\left(u,u\right) \geq \alpha \left\| \nabla u \right\|_0^2 + \lambda_0 \left\| u \right\|_0^2 - \varepsilon \left\| \nabla u \right\|_0^2 - \left(\frac{(\beta + \gamma)^2}{4\varepsilon} + \gamma \right) \left\| u \right\|_0^2. \qquad (8.56)$$

If we choose $\varepsilon = \alpha/2$ and $\lambda_0 = (\beta + \gamma)^2/4\varepsilon + \gamma$, we obtain

$$\tilde{B}\left(u,u\right) \geq \frac{\alpha}{2} \left\| \nabla u \right\|_0^2$$

which shows the coercivity of $\tilde{B}$. Introduce now the Hilbert triplet

$$V = H_0^1\left(\Omega\right), \; H = L^2\left(\Omega\right), \; V^* = H^{-1}\left(\Omega\right)$$

and recall that, since Ω is a **bounded, Lipschitz** domain, $H_0^1\left(\Omega\right)$ is dense and compactly embedded in $L^2\left(\Omega\right)$. Finally, define the *adjoint bilinear form of B* by

$$B^*\left(u,v\right) \equiv \int_\Omega \left\{ (\mathbf{A}^\top \nabla u + \mathbf{c}u) \cdot \nabla v - \mathbf{b}v \cdot \nabla u + a_0 uv \right\} \; d\mathbf{x} = B\left(v,u\right),$$

associated with the *formal adjoint* of $\mathcal{E}$

$$\mathcal{E}^* u = -\text{div}\left(\mathbf{A}^\top \nabla u + \mathbf{c}u\right) - \mathbf{b} \cdot \nabla u + a_0 u.$$

We are now in position to apply Theorem 6.12 to our variational problem. The conclusions are:

1) The subspaces $\mathcal{N}_B$ and $\mathcal{N}_{B^*}$ of the solutions of the homogeneous problems

$$B(u, v) = 0, \quad \forall v \in H_0^1(\Omega)$$

and

$$B^*(w, v) = 0, \quad \forall v \in H_0^1(\Omega)$$

share the same dimension d, $0 \leq d < \infty$.

2) The problem

$$B(u, v) = Fv, \qquad \forall v \in H_0^1(\Omega)$$

has a solution if and only if $Fw = 0$ for every $w \in \mathcal{N}_{B^*}$.

Let us translate the statements 1) and 2) into a less abstract language:

Theorem 8.7. *Let Ω be a bounded, Lipschitz domain, $f \in L^2(\Omega)$ and $\mathbf{f} \in L^2(\Omega; \mathbb{R}^n)$. Assume (8.46) and (8.47) hold. Then, we have the following alternative:*

a) Either $\mathcal{E}$ is an isomorphism between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ and therefore problem (8.48) has a unique weak solution, with

$$\|\nabla u\|_0 \leq C(n, a, K, \beta, \gamma)\{\|f\|_0 + \|\mathbf{f}\|_0\}$$

or the homogeneous and the adjoint homogeneous problems

$$\begin{cases} \mathcal{E}u = 0 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \end{cases} \quad \text{and} \quad \begin{cases} \mathcal{E}^*w = 0 & \text{in } \Omega \\ w = 0 & \text{on } \partial\Omega \end{cases}$$

have each d linearly independent solutions.

b) Moreover, problem (8.48) has a solution if and only if

$$\int_\Omega \{fw - \mathbf{f}\cdot\nabla w\}\, d\mathbf{x} = 0 \tag{8.57}$$

for every solution w of the adjoint homogeneous problem.

Theorem 8.7 implies that if we can show the uniqueness of the solution of problem (8.48), then automatically we infer both the existence and the stability estimate.

To show uniqueness, the weak maximum principles in subsection 8.5.5 are quite useful. We will be back to this argument there.

The conditions (8.57) constitute d *compatibility conditions* that the data have to satisfy in order for a solution to exists.

8.5.3 Neumann problem

Let Ω be a bounded, *Lipschitz* domain. The Neumann condition for an operator in the divergence form (8.45) assigns on $\partial\Omega$ the flux naturally associated with the operator. This flux is composed by two terms: $\mathbf{A}\nabla u \cdot \boldsymbol{\nu}$, due to the diffusion term $-\mathrm{div}\,\mathbf{A}\nabla u$, and $-\mathbf{b}u\cdot\boldsymbol{\nu}$, due to the convective term $\mathrm{div}(\mathbf{b}u)$, where $\boldsymbol{\nu}$ is the outward unit normal on $\partial\Omega$. We set

$$\partial_\nu^{\mathcal{E}} u \equiv (\mathbf{A}\nabla u - \mathbf{b}u)\cdot\boldsymbol{\nu} = \sum_{i,j=1}^n a_{ij}\partial_{x_j} u\, \nu_i - u\sum_j b_j \nu_j.$$

We call $\partial_\nu^{\mathcal{E}} u$ *conormal derivative of* u. Thus, the correct Neumann problem is:

$$\begin{cases} \mathcal{E}u = f & \text{in } \Omega \\ \partial_\nu^{\mathcal{E}} u = g & \text{on } \partial\Omega. \end{cases} \tag{8.58}$$

with $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$. The variational formulation of problem (8.58) may be obtained by the usual integration by parts technique. It is enough to note, that, multiplying the differential equation $\mathcal{E}u = f$ by a test function $v \in H^1(\Omega)$ and using the Neumann condition, we get, formally:

$$\int_\Omega \{(\mathbf{A}\nabla u - \mathbf{b}u)\nabla v + (\mathbf{c}\cdot\nabla u)v + a_0 uv\}\,d\mathbf{x} = \int_\Omega fv\,d\mathbf{x} + \int_{\partial\Omega} gv\,d\sigma.$$

Introducing the bilinear form

$$B(u,v) = \int_\Omega \{(\mathbf{A}\nabla u - \mathbf{b}u)\nabla v + (\mathbf{c}\cdot\nabla u)v + a_0 uv\}\,d\mathbf{x} \tag{8.59}$$

and the linear functional

$$Fv = \int_\Omega fv\,d\mathbf{x} + \int_{\partial\Omega} gv\,d\sigma,$$

we are led to the following **weak formulation**, that can be easily checked to be equivalent to the original problem, when all the data are smooth:

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

$$B(u,v) = Fv, \quad \forall v \in H^1(\Omega). \tag{8.60}$$

If the size of $\mathbf{b} - \mathbf{c}$ is small enough, problem (8.60) is well-posed, as the following proposition shows.

Proposition 8.7. *Assume that hypotheses (8.46) and (8.47) hold and that $f \in L^2(\Omega)$, $g \in L^2(\partial\Omega)$. If $a_0(\mathbf{x}) \geq c_0 > 0$ a.e. in Ω and*

$$\alpha_0 \equiv \min\{\alpha - (\beta+\gamma)/2, c_0 - (\beta+\gamma)/2\} > 0, \tag{8.61}$$

then, problem (8.60) has a unique solution. Moreover, the following stability estimate holds:

$$\|u\|_{1,2} \leq \frac{1}{\alpha_0}\left\{\|f\|_0 + \overline{C}(n,\Omega)\|g\|_{L^2(\partial\Omega)}\right\}.$$

Proof (sketch). Since

$$|B(u,v)| \leq (M + \beta + \gamma + \gamma_0) \|u\|_{1,2} \|v\|_{1,2}$$

B is continuous in $H^1(\Omega)$. Moreover, we may write

$$B(u,u) \geq \alpha \int_\Omega |\nabla u|^2 \, d\mathbf{x} - \left| \int_\Omega [(\mathbf{b} - \mathbf{c}) \cdot \nabla u] \ u \ d\mathbf{x} \right| + \int_\Omega a_0 u^2 d\mathbf{x}.$$

From Schwarz's inequality and the inequality $2ab \leq a^2 + b^2$, we obtain

$$\left| \int_\Omega [(\mathbf{b} - \mathbf{c}) \cdot \nabla u] \ u \ d\mathbf{x} \right| \leq (\beta + \gamma) \|\nabla u\|_0 \|u\|_0 \leq \frac{(\beta + \gamma)}{2} \|u\|_{1,2}^2.$$

Thus, if (8.61) holds, we get $B(u,u) \geq \alpha_0 \|u\|_{1,2}^2$ and therefore B is coercive. Finally, using (8.36), it is not difficult to check that $F \in H^1(\Omega)^*$, with

$$\|F\|_{H^1(\Omega)^*} \leq \|f\|_0 + \overline{C}(n, \Omega) \|g\|_{L^2(\partial\Omega)}.$$

$\square$

Alternative for the Neumann problem. The bilinear form B is coercive also under the conditions

$$(\mathbf{b} - \mathbf{c}) \cdot \boldsymbol{\nu} \leq 0 \ \text{ a.e. on } \partial\Omega \quad \text{and} \quad \frac{1}{2} \text{div}(\mathbf{b} - \mathbf{c}) + a_0 \geq c_0 > 0 \ \text{ a.e in } \Omega,$$

as it can be checked following the proof of Proposition 8.6.

However, in general the bilinear form B is only *weakly coercive*. In fact, choosing in (8.56) $\varepsilon = \alpha/2$ and $\lambda_0 = (\beta + \gamma)^2 / 2\varepsilon + 2\gamma + 2\gamma_0$, we easily get

$$\tilde{B}(u,u) = B(u,u) + \lambda_0 \|u\|_0^2 \geq \frac{\alpha}{2} \|\nabla u\|_0^2 + \left(\frac{(\beta + \gamma)^2}{4\varepsilon} + \gamma + \gamma_0 \right) \|u\|_0^2$$

and therefore $\tilde{B}$ is coercive. Applying theorem 6.12, we obtain the following alternative:

Theorem 8.8. *Let Ω be a bounded, Lipschitz domain. Assume that (8.46) and (8.47) hold. Then, if $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$:*

a) Either problem (8.58) has a unique solution $u \in H^1(\Omega)$ and

$$\|u\|_{1,2} \leq C(n, \alpha, M, \beta, \gamma, \gamma_0) \left\{ \|f\|_0 + \|g\|_{L^2(\partial\Omega)} \right\}$$

or the homogeneous and the adjoint homogeneous problems

$$\begin{cases} \mathcal{E}u = 0 & \text{in } \Omega \\ (\mathbf{A}\nabla u - \mathbf{b}u) \cdot \boldsymbol{\nu} = 0 & \text{on } \partial\Omega \end{cases} \quad \text{and} \quad \begin{cases} \mathcal{E}^* w = 0 & \text{in } \Omega \\ (\mathbf{A}^\top \nabla w + \mathbf{c}w) \cdot \boldsymbol{\nu} = 0 & \text{on } \partial\Omega \end{cases}$$

have each d linearly independent solutions.

b) Moreover, problem (8.60) has a solution if and only if

$$Fw = \int_\Omega fw \; d\mathbf{x} + \int_{\partial\Omega} gw \; d\sigma = 0 \tag{8.62}$$

for every solution w of the adjoint homogeneous problem.

Remark 8.11. Again, uniqueness implies existence. Note that if $\mathbf{b} = \mathbf{c} = \mathbf{0}$ and $a_0 = 0$, then the solutions of the adjoint homogeneous problem are the constant functions. Therefore $d = 1$ and the compatibility condition (8.62) reduces to the well known equation

$$\int_\Omega f \; d\mathbf{x} + \int_{\partial\Omega} g \; d\sigma = 0.$$

Remark 8.12. Note that in the right hand side of (8.58) there is no term of the form div $\mathbf{f}$, as was the case in the Dirichlet problem. Indeed, it is better to avoid terms of that form for the following reason. Consider, for instance, the problem $-\Delta u = \text{div} \mathbf{f}$, $\partial_\nu u = 0$. A weak formulation would be, after the usual integration by parts,

$$\int_\Omega \nabla u \cdot \nabla v \; d\mathbf{x} = \int_{\partial\Omega} (\mathbf{f} \cdot \boldsymbol{\nu}) \, v d\sigma - \int_\Omega \mathbf{f} \cdot \nabla v \; d\mathbf{x} \qquad \forall v \in H^1(\Omega). \tag{8.63}$$

However, even if $\mathbf{f}$ is smooth, (8.63) is equivalent to $\text{div}(\nabla u + \mathbf{f}) = 0$ in the sense of distributions, but with

$$(\nabla u + \mathbf{f}) \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega,$$

giving rise to a different problem.

8.5.4 Robin and mixed problems

Robin problem. The variational formulation of the problem

$$\begin{cases} \mathcal{E}u = f & \text{in } \Omega \\ \partial_\nu^\mathcal{E} u + hu = g & \text{on } \partial\Omega. \end{cases} \tag{8.64}$$

is obtained by replacing the bilinear form B in problem (8.60), by

$$\tilde{B}(u,v) = B(u,v) + \int_{\partial\Omega} huv \; d\sigma$$

If $0 \le h(\mathbf{x}) \le h_0$ a.e. on $\partial\Omega$, Proposition 8.7 still holds for problem (8.64).

As for the Neumann problem, in general the bilinear form B is only *weakly coercive* and a theorem perfectly analogous to Theorem 8.8 holds. We leave the details as an exercise.

Mixed Dirichlet-Neumann problem. Let Γ_D be a non empty relatively open subset of $\partial\Omega$ and $\Gamma_N = \partial\Omega\backslash\Gamma_D$. Consider the mixed problem

$$\begin{cases} \mathcal{E}u = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \partial_\nu^{\mathcal{E}}u = g & \text{on } \Gamma_N \end{cases}$$

As in subsection 8.4.2, the correct functional setting is $H_{0,\Gamma_D}^1(\Omega)$ with the norm $\|u\|_{H_{0,\Gamma_D}^1(\Omega)} = \|\nabla u\|_0$. Introducing the linear functional

$$Fv = \int_\Omega fv \, d\mathbf{x} + \int_{\Gamma_N} gv \, d\sigma,$$

the **variational formulation** is the following: *Determine $u \in H_{0,\Gamma_D}^1(\Omega)$ such that*

$$B(u, v) = Fv, \quad \forall v \in H_{0,\Gamma_D}^1(\Omega). \tag{8.65}$$

Proceeding as in Proposition 8.6, we may prove the following result:

Proposition 8.8. *Assume that hypotheses (8.46) and (8.47) hold and that $f \in L^2(\Omega)$, $g \in L^2(\Gamma_N)$. If $\mathbf{b}$ and $\mathbf{c}$ have Lipschitz components and*

$$(\mathbf{b} - \mathbf{c}) \cdot \boldsymbol{\nu} \le 0 \quad \text{a.e. on } \Gamma_N, \quad \frac{1}{2}\text{div}(\mathbf{b} - \mathbf{c}) + a_0 \ge 0, \quad \text{a.e. in } \Omega,$$

then problem (8.65) has a unique solution $u \in H_{0,\Gamma_D}^1(\Omega)$. Moreover, the following stability estimate holds:

$$\|u\|_1 \le \frac{1}{\alpha}\left\{\|f\|_0 + \overline{C}\|g\|_{L^2(\Gamma_N)}\right\}.$$

Remark 8.13. If $u = g_0$ on Γ_D, i.e. if the Dirichlet data are nonhomogeneous, set $w = u - \widetilde{g}_0$, where $\widetilde{g}_0 \in H^1(\Omega)$ is an extension of g_0. Then $w \in H_{0,\Gamma_D}^1(\Omega)$ and solves

$$B(w, v) = B(\tilde{g}_0, v) + \int_\Omega fv \, d\mathbf{x} + \int_{\Gamma_N} gv \, d\sigma \quad \forall v \in H_{0,\Gamma_D}^1(\Omega).$$

For the mixed problem as well, in general the bilinear form is only weakly coercive and we may resort to the alternative theorem, achieving a result similar to Theorems 8.7. Only note that the compatibility conditions (8.60) take the form

$$Fw = \int_\Omega fw \, d\mathbf{x} + \int_{\Gamma_N} gw \, d\sigma = 0$$

for every solution w of the adjoint problem

$$\begin{cases} \mathcal{E}^*w = 0 & \text{in } \Omega \\ w = 0 & \text{on } \Gamma_D \\ (\mathbf{A}^\top\nabla w + \mathbf{c}w) \cdot \boldsymbol{\nu} = 0 & \text{on } \Gamma_N. \end{cases}$$

8.5.5 Weak Maximum Principles

In Chapter 2 we have given a version of the maximum principle for the Laplace equation. This principle has an extension valid for general divergence form operators. First, some remarks.

Let Ω be a bounded, Lipschitz domain and $u \in H^1(\Omega)$. Since $C^1(\overline{\Omega})$ is dense in $H^1(\Omega)$, $u \geq 0$ on $\partial\Omega$ if there exists a sequence $\{v_k\}_{k \geq 1} \subset C^1(\overline{\Omega})$ such that $v_k \to u$ in $H^1(\Omega)$ and $v_k \geq 0$. It is as if the trace of u on $\partial\Omega$ "inherits" the nonnegativity from the sequence $\{v_k\}_{k \geq 1}$.

Since $v_k \geq 0$ on $\partial\Omega$ is equivalent to saying that[5] the negative part $v_k^- = \max\{-v_k, 0\}$ has zero trace on $\partial\Omega$, it then turns out that $u \geq 0$ on $\partial\Omega$ if and only if $u^- \in H_0^1(\Omega)$. Similarly, $u \leq 0$ on $\partial\Omega$ if and only if $u^+ \in H_0^1(\Omega)$.

Other inequalities follow in a natural way. For instance, we have $u \leq v$ on $\partial\Omega$ if $u - v \leq 0$ on $\partial\Omega$. Thus, we may define:

$$\sup_{\partial\Omega} u = \inf\{k \in \mathbb{R} : u \leq k \text{ on } \partial\Omega\}, \qquad \inf_{\partial\Omega} u = \sup\{k \in \mathbb{R} : u \geq k \text{ on } \partial\Omega\}$$

which coincide with the usual *greatest lower bound* and *lowest upper bound* when $u \in C(\partial\Omega)$.

Consider the equation

$$B(u, v) = \int_\Omega \{(\mathbf{A}\nabla u - \mathbf{b}u)\nabla v + \mathbf{c}v \cdot \nabla u + a_0 uv\} \, d\mathbf{x} = 0, \qquad (8.66)$$

for every $v \in H_0^1(\Omega)$. We have:

Theorem 8.9. (*Weak maximum principle*). Assume that $u \in H^1(\Omega)$ satisfies (8.66) and that (8.46) and (8.47) hold. Moreover, let $\mathbf{b}$ Lipschitz and

$$\text{div}\mathbf{b} + a_0 \geq 0 \quad \text{a.e. in } \Omega. \qquad (8.67)$$

Then

$$\sup_\Omega u \leq \sup_{\partial\Omega} u^+ \qquad \text{and} \qquad \inf_\Omega \geq \inf_{\partial\Omega} u^-. \qquad (8.68)$$

Proof. For simplicity, we give the proof only for $\mathbf{b} = \mathbf{c} = \mathbf{0}$, and therefore $a_0 \geq 0$ a.e. in Ω. We have:

$$\int_\Omega \mathbf{A}\nabla u \cdot \nabla v \, d\mathbf{x} = -\int_\Omega a_0 uv \, d\mathbf{x}, \qquad \forall v \in H_0^1(\Omega).$$

Let

$$l = \sup_{\partial\Omega} u^+$$

We may assume that $l < \infty$, otherwise there is nothing to be proved. Select as a test function $v = \max\{u - l, 0\} \geq 0$, which belongs to $H_0^1(\Omega)$.

---

[5] Recall from subsection 7.5.2 that, if $u \in H^1(\Omega)$ then its positive and negative part, $u^+ = \max\{u, 0\}$ and $u^- = \max\{-u, 0\}$, belong to $H^1(\Omega)$ as well.

Now, observe that in the set $\{u > l\}$, where $v > 0$ a.e., we have $\nabla v = \nabla u$ so that, using the uniform ellipticity condition and (8.67), we obtain

$$\alpha \int_{\{u>l\}} |\nabla v|^2 \, d\mathbf{x} \leq \int_{\Omega} \mathbf{A}\nabla u \cdot \nabla v \, d\mathbf{x} = -\int_{\{u>l\}} a_0 u \, (u - l) \, d\mathbf{x} \leq 0.$$

Thus, either $|\{u > l\}| = 0$ or $\nabla v = 0$. In any case, since $v \in H_0^1(\Omega)$, we infer $v = 0$, whence $u \leq l$.

The second inequality in (8.68) may be proved in similar way. $\square$

Remark 8.14. Note that Theorem 8.9 implies that if $u \leq 0$ or $u \geq 0$ on $\partial\Omega$, then $u \leq 0$ or $u \geq 0$ in Ω. In particular, if $u = 0$ on $\partial\Omega$ then $u = 0$ in Ω.

Also, it is not possible to substitute $\sup_{\partial\Omega} u^+$ by $\sup_{\partial\Omega} u$ or $\inf_{\partial\Omega} u^-$ with $\inf_{\partial\Omega} u$ in (8.68). A counterexample in dimension *one* is shown in figure 8.2. The solution of $-u'' + u = 0$ in $(0, 1)$, $u(0) = u(1) = -1$, has a *negative maximum* which is greater than -1.
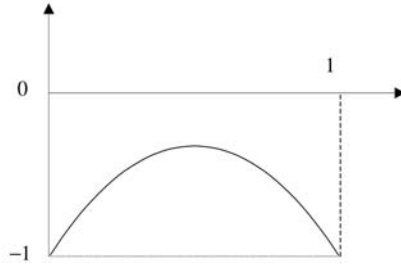
Fig. 8.2. The solution of $-u'' + u = 0$ in $(0, 1)$, $u(0) = u(1) = -1$

Using Theorem 6.12, we have:

Corollary 8.1. *Under the hypotheses of Theorem 8.9, the Dirichlet problem*

$$\begin{cases} \mathcal{E}u = f + \mathrm{div}\mathbf{f} & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

has a unique solution $u \in H_0^1(\Omega)$ and

$$\|\nabla u\|_0 \leq C(n, \alpha, K, \beta, \gamma) \{\|f\|_0 + \|\mathbf{f}\|_0\}.$$

A similar maximum principle holds for Robin or mixed conditions, yielding uniqueness and therefore well-posedness, for the corresponding problems.

Suppose for instance that $u \in H^1(\Omega)$ satisfies the equation

$$B(u, v) = 0, \quad \forall v \in H_{0,\Gamma_D}^1(\Omega). \tag{8.69}$$

Then u is a solution of a mixed problem with $f = g = 0$. We may prove the following result (compare with Example 8.18).

Theorem 8.10. *Let $\Gamma_D \subset \partial\Omega$, open, $\Gamma_D \neq \emptyset$. Assume that $u \in H^1(\Omega)$ satisfies (8.69) and that (8.46) and (8.47) hold. Moreover, let $\mathbf{b}$ be Lipschitz and*

$$\mathbf{b} \cdot \boldsymbol{\nu} \leq 0 \quad \text{a.e. on } \Gamma_N, \quad \text{div}\mathbf{b} + a_0 \geq 0 \quad \text{a.e. in } \Omega.$$

Then

$$\sup_{\Omega} u \leq \sup_{\Gamma_D} u^+ \qquad and \qquad \inf_{\Omega} \geq \inf_{\Gamma_D} u^-.$$

8.6 Regularity

An important task, in general technically rather complicated, is to establish the optimal regularity of a weak solution in relation to the degree of smoothness of the data: *the domain Ω, the boundary data, the coefficients of the operator and the forcing term*. To get a clue of what happens, consider for example the following Poisson problem:

$$\begin{cases} -\Delta u + u = F & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

where $F \in H^{-1}(\Omega)$. Under this hypothesis, the Lax-Milgram Theorem yields a solution $u \in H_0^1(\Omega)$ and we cannot get much more, in terms of smoothness. Indeed, from Sobolev inequalities (see subsection. 7.10.4) it follows that $u \in L^p(\Omega)$ with $p = \frac{2n}{n-2}$, if $n \geq 3$, or $u \in L^p(\Omega)$, with $2 \leq p < \infty$, if $n = 2$. However, this gain in integrability does not seriously increase the smoothness of u.

Reversing our point of view, we may say that, starting from a function in $H_0^1(\Omega)$ and applying to it a second order operator "two orders of differentiability are lost": the loss of one order drives from $H_0^1(\Omega)$ into $L^2(\Omega)$ while a further loss leads to $H^{-1}(\Omega)$. It is as if the upper index -1 indicates a "lack" of one order of differentiability.

Nevertheless, consider the case in which $u \in H^1(\mathbb{R}^n)$ is a solution of the equation

$$-\Delta u + u = f \quad \text{in } \mathbb{R}^n. \tag{8.70}$$

We ask: *if $f \in L^2(\mathbb{R}^n)$ what is the optimal regularity of u?*

Following the above argument, our conclusions would be: it is true that we start from $u \in H^1(\mathbb{R}^n)$, but applying the second order operator $-\Delta + I$, where I denotes the *identity operator*, we find $f \in L^2(\mathbb{R})$. Thus we conclude that the starting function should actually be in $H^2(\mathbb{R}^n)$ rather than $H^1(\mathbb{R}^n)$. Indeed this is true and can be easily proved using the Fourier transform. Since

$$\widehat{\partial_{x_i} u}(\boldsymbol{\xi}) = i\xi_i \widehat{u}(\boldsymbol{\xi}), \qquad \widehat{\partial_{x_i x_j} u}(\boldsymbol{\xi}) = -\xi_i \xi_j \widehat{u}(\boldsymbol{\xi})$$

we have

$$-\widehat{\Delta u}(\boldsymbol{\xi}) = |\boldsymbol{\xi}|^2 \widehat{u}(\boldsymbol{\xi})$$

and equation (8.70) becomes

$$(1 + |\boldsymbol{\xi}|^2)\widehat{u}(\boldsymbol{\xi}) = \widehat{f}(\boldsymbol{\xi})$$

whence

$$\widehat{u}(\boldsymbol{\xi}) = \frac{\widehat{f}(\boldsymbol{\xi})}{1 + |\boldsymbol{\xi}|^2}. \tag{8.71}$$

From (8.71) we easily draw the information we were looking for: *every second derivative of u belongs to* $L^2(\mathbb{R}^n)$. This comes from the following facts:

- formula (7.27):
$$\|\widehat{v}\|_{L^2(\mathbb{R}^n)}^2 = (2\pi)^n \|v\|_{L^2(\mathbb{R}^n)}^2,$$

- the elementary inequality
$$2\left|\xi_i\xi_j\right| < 1 + |\boldsymbol{\xi}|^2, \quad \forall i, j = 1, ..., n,$$

- the simple computation
$$\int_{\mathbb{R}^n} |\partial_{x_i x_j} u(\mathbf{x})|^2 \, d\mathbf{x} = \int_{\mathbb{R}^n} \xi_i^2 \xi_j^2 |\widehat{u}(\boldsymbol{\xi})|^2 \, d\boldsymbol{\xi} = \int_{\mathbb{R}^n} \frac{\xi_i^2 \xi_j^2}{(1 + |\boldsymbol{\xi}|^2)^2} \left|\widehat{f}(\boldsymbol{\xi})\right|^2 \, d\boldsymbol{\xi}$$
$$< \frac{1}{4}\int_{\mathbb{R}^n} \left|\widehat{f}(\boldsymbol{\xi})\right|^2 \, d\boldsymbol{\xi} = \frac{(2\pi)^n}{4}\int_{\mathbb{R}^n} |f(\mathbf{x})|^2 \, d\mathbf{x}.$$

Thus, $u \in H^2(\mathbb{R}^n)$ and moreover, we have obtained the estimate

$$\|u\|_{H^2(\mathbb{R}^n)} \leq C \|f\|_{L^2(\mathbb{R}^n)}.$$

We may go further. If $f \in H^1(\mathbb{R}^n)$, i.e. if f has first partials in $L^2(\mathbb{R}^n)$, a similar computation yields $u \in H^3(\mathbb{R}^n)$. Iterating this argument, we conclude that for every $m \geq 0$,

$$\text{if } f \in H^m(\mathbb{R}^n) \qquad \text{then} \qquad u \in H^{m+2}(\mathbb{R}^n).$$

Using the Sobolev embedding theorems of subsection 7.10.4, we infer that, if m is sufficiently large, *u is a classical solution*. In fact, if $u \in H^{m+2}(\mathbb{R}^n)$ then

$$u \in C^k(\mathbb{R}^n) \text{ for } k < m + 2 - \frac{n}{2},$$

and therefore it is enough that $m > \frac{n}{2}$ to have u at least in $C^2(\mathbb{R}^n)$. An immediate consequence is:

$$\text{if } f \in C^\infty(\mathbb{R}^n) \qquad \text{then} \qquad u \in C^\infty(\mathbb{R}^n).$$

This kind of results can be extended to *uniformly elliptic* operators $\mathcal{E}$ in divergence form and to the solutions of the *Dirichlet, Neumann* and *Robin* problems.

The regularity for mixed problems is more delicate and requires compatibility conditions along the border between Γ_D and Γ_N. We will not insist on this subject.

There are two kinds of regularity results, concerning *interior regularity* and *global regularity,* respectively. Since the proofs are quite technical (see *Evans,* 1998), we only state the main results.

In all the theorems below u is a weak solution of

$$\mathcal{E}u = f \quad \text{in } \Omega.$$

We keep the hypotheses (8.46) and (8.47).

• *Interior regularity.* The next theorem is a H^2-interior regularity result. Note that the boundary of the domain does not play any role. We have:

Theorem 8.11. (H^2 interior regularity). *Let the coefficients a_{ij} be Lipschitz in Ω. Then $u \in H_{loc}^2(\Omega)$ and if $\Omega' \subset\subset \Omega$,*

$$\|u\|_{H^2(\Omega')} \le C \left\{ \|f\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)} \right\}. \tag{8.72}$$

Thus, u is a *strong solution* (see Section 8.2) in Ω. The constant C depends on all the relevant parameters $\alpha, \beta, \gamma, \gamma_0, M$ and also on the distance of Ω' from $\partial\Omega$ and the Lipschitz constant of a_{ij} and b_j, $i, j = 1, ..., n$.

Remark 8.15. The presence of the norm $\|u\|_{L^2(\Omega)}$ in the right hand side of (8.72) is necessary[6] and due to the fact that the bilinear form B associated to $\mathcal{E}$ is only weakly coercive.

If we increase the regularity of the coefficients, the smoothness of u increases according to the following theorem:

Theorem 8.12. (Higher interior regularity). *Let $a_{ij}, b_j \in C^{m+1}(\Omega)$ and $c_j, a_0 \in C^m(\Omega)$, $m \ge 1$, $i, j = 1, ..., n$. Then $u \in H_{loc}^{m+2}(\Omega)$ and if $\Omega' \subset\subset \Omega$,*

$$\|u\|_{H^{m+2}(\Omega')} \le C \left\{ \|f\|_{H^m(\Omega)} + \|u\|_{L^2(\Omega)} \right\}.$$

As a consequence, if $a_{ij}, b_j, c_j, a_0, f \in C^\infty(\Omega)$, then $u \in C^\infty(\Omega)$ as well.

• *Global regularity.* We focus on the optimal regularity of a solution (non necessarily unique!) of the boundary value problems we have considered in the previous sections.

Consider first H^2-regularity. If $u \in H^2(\Omega)$, its trace on $\partial\Omega$ belongs to $H^{3/2}(\partial\Omega)$ so that a Dirichlet data g_D has to be taken in this space. On the other hand, the trace of the normal derivative belongs to $H^{1/2}(\partial\Omega)$ and hence we have to assign a Neumann or a Robin data g_N in this space. Also, the domain has to be smooth enough, say C^2, in order to define the traces of u and $\partial_\nu u$.

---

[6] For instance, $u(x) = \sin x$ is a solution of the equation $u'' + u = 0$. Clearly we cannot control any norm of u with the norm of the right hand side alone!

Thus, assume that u is a solution of $\mathcal{E}u = f$ in Ω, with one of the following boundary conditions:

$$u = g_D \in H^{3/2}(\partial\Omega)$$

or

$$\partial_{\boldsymbol{\nu}}^{\mathcal{E}} + hu = g_N \in H^{1/2}(\partial\Omega),$$

with

$$0 \leq h(\boldsymbol{\sigma}) \leq h_0 \quad \text{a.e. on } \partial\Omega.$$

We have:

Theorem 8.13. *Let Ω be a bounded, C^2-domain. Assume that a_{ij}, b_j, $i, j = 1, ..., n$, are Lipschitz in Ω and $f \in L^2(\Omega)$. Then $u \in H^2(\Omega)$ and*

$$\|u\|_{H^2(\Omega)} \leq C\left\{\|u\|_0 + \|f\|_0 + \|g_D\|_{H^{3/2}(\partial\Omega)}\right\} \qquad \text{(Dirichlet)},$$

$$\|u\|_{H^2(\Omega)} \leq C\left\{\|u\|_0 + \|f\|_0 + \|g_R\|_{H^{1/2}(\partial\Omega)}\right\} \qquad \text{(Neumann/Robin)}.$$

If we increase the regularity of the domain, the coefficients and the data, the smoothness of u increases accordingly to the following theorem.

Theorem 8.14. *Let Ω be a bounded $C^{m+2}-$domain. Assume that $a_{ij}, b_j \in C^{m+1}(\overline{\Omega}), c_j, a_0 \in C^m(\overline{\Omega}), i, j = 1, ..., n$, $f \in H^m(\Omega)$. If $g_D \in H^{m+3/2}(\partial\Omega)$ or $g_N \in H^{m+1/2}(\partial\Omega)$ and $h \in C^{m+1}(\partial\Omega)$, then $u \in H^{m+2}(\Omega)$ and moreover,*

$$\|u\|_{H^{m+2}(\Omega)} \leq C\left\{\|u\|_0 + \|f\|_{H^m(\Omega)} + \|g_D\|_{H^{m+3/2}(\partial\Omega)}\right\} \qquad \text{(Dirichlet)},$$

$$\|u\|_{H^{m+2}(\Omega)} \leq C\left\{\|u\|_0 + \|f\|_{H^m(\Omega)} + \|g_R\|_{H^{m+1/2}(\partial\Omega)}\right\} \qquad \text{(Neumann, Robin)}.$$

In particular, if Ω is a $C^\infty-$domain, all the coefficients are in $C^\infty(\overline{\Omega})$ and the boundary data are in $C^\infty(\partial\Omega)$, then $u \in C^\infty(\overline{\Omega})$.

• *A particular case.* Let Ω be a C^2-domain and $f \in L^2(\Omega)$. The Lax-Milgram Theorem and Theorem 8.11 imply that the solution of the Dirichlet problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

belongs to $H^2(\Omega) \cap H_0^1(\Omega)$ and that

$$\|u\|_{H^2(\Omega)} \leq C\|f\|_0 = C\|\Delta u\|_0. \tag{8.73}$$

Since clearly we have

$$\|\Delta u\|_0 \leq \|u\|_{H^2(\Omega)}$$

we draw the following important conclusion:

Corollary 8.2. *If $u \in H^2(\Omega) \cap H_0^1(\Omega)$, then*

$$\|\Delta u\|_0 \leq \|u\|_{H^2(\Omega)} \leq C_b \|\Delta u\|_0 \,.$$

In other words, $\|\Delta u\|_0$ and $\|u\|_{H^2(\Omega)}$ are equivalent norms in $H^2(\Omega) \cap H_0^1(\Omega)$.

In the next section, we will see an application of Corollary 8.2 to an equilibrium problem for a bent plate.

• *Domains with corners.* The above regularity results hold for smooth domains. However, in several applied situations, Lipschitz domains are the relevant ones. For these domains the regularity theory is not elementary and goes beyond the purposes of the present book. Thus, we only give an idea of what happens by means of two examples.
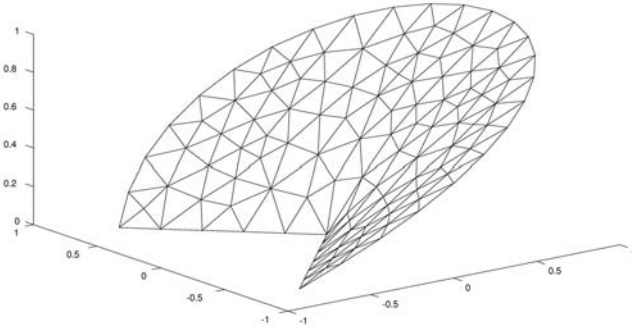
Fig. 8.3. The case $\alpha = \dfrac{3}{2}\pi$ in Example 8.17

Example 8.2. Consider the plane sector:

$$S_\alpha = \{(r, \theta) : 0 < r < 1, \, -\alpha/2 < \theta < \alpha/2\} \qquad (0 < \alpha < 2\pi).$$

The function

$$u(r, \theta) = r^{\frac{\pi}{\alpha}} \cos \frac{\pi}{\alpha}\theta$$

is harmonic in S_α, since it is the real part of $f(z) = z^{\frac{\pi}{\alpha}}$, which is holomorphic in S_α. Furthermore,

$$u(r, -\alpha/2) = u(r, \alpha/2) = 0, \qquad 0 \leq r \leq 1 \tag{8.74}$$

and

$$u(1, \theta) = \cos \frac{\pi}{\alpha}\theta, \qquad 0 \leq \theta \leq \alpha. \tag{8.75}$$

We focus on a neighborhood of the origin. If $\alpha = \pi$, S_α is a semicircle and

$$u(r, \theta) = \operatorname{Re} z = x \in C^\infty\left(\overline{S}_\alpha\right).$$

Suppose $\alpha \neq \pi$. Since

$$|\nabla u|^2 = u_r^2 + \frac{1}{r^2}u_\theta^2 = \frac{\pi^2}{\alpha^2}r^{2\left(\frac{\pi}{\alpha}-1\right)},$$

we have

$$\int_{S_\alpha}|\nabla u|^2\,dx_1dx_2 = \frac{\pi^2}{\alpha}\int_0^1 r^{2\frac{\pi}{\alpha}-1}dr = \frac{\pi}{2}$$

so that $u \in H^1(S_\alpha)$ and is the unique weak solution of $\Delta u = 0$ in S_α with the boundary conditions (8.74), (8.75). It is easy to check that for every $i,j = 1,2$,

$$\left|\partial_{x_ix_j}u\right| \sim r^{\frac{\pi}{\alpha}-2} \qquad \text{as } r \to 0$$

whence

$$\int_{S_\alpha}\left|\partial_{x_ix_j}u\right|^2\,dx_1dx_2 \simeq \int_0^1 r^{2\frac{\pi}{\alpha}-3}dr.$$

This integral is convergent only for

$$2\frac{\pi}{\alpha} - 3 > -1.$$

The conclusion is that $u \in H^2(S_\alpha)$ *if and only if* $\alpha \leq \pi$, i.e. if *the sector is convex.* If $\alpha > \pi$, $u \notin H^2(S_\alpha)$.

Conclusion: *in a neighborhood of a non convex angle, we expect a low degree of regularity of the solution* (less than H^2).
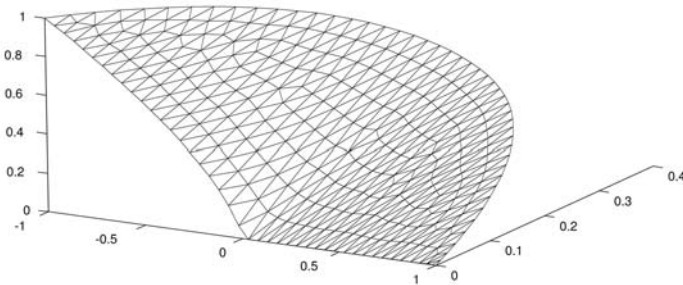
Fig. 8.4. The solution of the mixed problem in Example 8.18

Example 8.3. As a second example, the function $u(r,\theta) = r^{\frac{1}{2}}\sin\frac{\theta}{2}$ is a weak solution in the half circle

$$S_\pi = \{(r,\theta) : 0 < r < 1, \quad 0 < \theta < \pi\}$$

of the mixed problem

$$\begin{cases} \Delta u = 0 & \text{in } S_\pi \\ u(1, \theta) = \sin \frac{\theta}{2} & 0 < \theta < \pi \\ u(r, 0) = 0 \text{ and } \partial_{x_2} u(r, \pi) = 0 & 0 \le r < 1. \end{cases}$$

Namely,

$$|\nabla u|^2 = \frac{1}{4r},$$

so that

$$\int_{S_\pi} |\nabla u|^2 \, dx_1 dx_2 = \frac{\pi}{4}$$

whence $u \in H^1(S_\pi)$. Moreover,

$$\partial_{x_2} u = u_r \sin \theta + \frac{1}{r} u_\theta \cos \theta = \frac{1}{2\sqrt{r}} \cos \frac{\theta}{2}$$

hence

$$\partial_{x_2} u(r, \pi) = 0.$$

However, along the half-line $\theta = \pi/2$, for example, we have

$$\left| \partial_{x_i x_j} u \right| \sim r^{-\frac{3}{2}} \qquad r \sim 0$$

so that

$$\int_{S_\alpha} \left| \partial_{x_i x_j} u \right|^2 \, dx_1 dx_2 \sim \int_0^1 r^{-2} dr = \infty.$$

and therefore $u \notin H^2(S_\pi)$.

Thus, the solution has a low order of regularity near the origin, even though the boundary of S_π is flat there. Note that the origin separates the Dirichlet and Neumann regions (see Fig. 8.4).

Conclusion: *in general, the optimal regularity of the solution of a mixed problem is less than H^2 near the boundary between the Dirichlet and Neumann regions.*

8.7 Equilibrium of a plate

The range of application of the variational theory is not confined to second order equations. In this section we consider the vertical deflection $u = u(x, y)$ of a bent plate of small thickness (compared with the other dimensions) under the action of a normal load. If $\Omega \subset \mathbb{R}^2$ represents the transversal section of the plate, it can be shown that u is governed by the fourth order equation

$$\Delta \Delta u = \Delta^2 u = \frac{q}{D} \equiv f \qquad \text{in } \Omega,$$

where q is the density of loading and D encodes the elastic properties of the material. The operator Δ^2 is called **biharmonic** or **bi-laplacian** and the solutions of

$\Delta^2 u = 0$ are called *biharmonic functions*. In two dimensions, the explicit expression of Δ^2 is given by[7]

$$\Delta^2 = \frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4}.$$

If the plate is rigidly fixed along its boundary (*clamped plate*), then u and its normal derivative must vanish on $\partial\Omega$. Thus, we are led to the following boundary value problem:

$$\begin{cases} \Delta^2 u = f & \text{in } \Omega \\ u = \partial_{\boldsymbol{\nu}} u = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.76a}$$

We want to derive a variational formulation. To obtain it, choose $C_0^2(\Omega)$ as space of test functions, i.e. the set of functions in $C^2(\Omega)$, compactly supported in Ω. This choice takes into account the boundary conditions. Now, multiply the biharmonic equation by a function $v \in C_0^2(\Omega)$ and integrate over Ω:

$$\int_\Omega \Delta^2 u \; v \; d\mathbf{x} = \int_\Omega f \; v \; d\mathbf{x}. \tag{8.77}$$

Integrating by parts twice and using the conditions $v = \partial_{\boldsymbol{\nu}} v = 0$ on $\partial\Omega$, we get:

$$\int_\Omega \Delta^2 u \; v \; d\mathbf{x} = \int_\Omega (\operatorname{div}\nabla\Delta u)v \; d\mathbf{x} = \int_{\partial\Omega} \partial_{\boldsymbol{\nu}}(\Delta u)v \; d\sigma - \int_\Omega \nabla\Delta u \cdot \nabla v \; d\mathbf{x}$$

$$= -\int_{\partial\Omega} \Delta u \partial_{\boldsymbol{\nu}} v \; d\sigma + \int_\Omega \Delta u \Delta v \; d\mathbf{x} = \int_\Omega \Delta u \Delta v \; d\mathbf{x}.$$

Thus, (8.77) becomes

$$\int_\Omega \Delta u \Delta v \; d\mathbf{x} = \int_\Omega fv \; d\mathbf{x}. \tag{8.78}$$

Now we enlarge the space of test functions by taking the closure of $C_0^2(\Omega)$ in $H^2(\Omega)$, which is $H_0^2(\Omega)$. Note that (see subsection 7.9.2) this is precisely the space of functions u such that u and $\partial_{\boldsymbol{\nu}} u$ have zero trace on $\partial\Omega$.

Since $H_0^2(\Omega) \subset H_0^1(\Omega) \cap H^2(\Omega)$, from Corollary 8.2 we know that in this space we may choose $\|u\|_2 = \|\Delta u\|_0$ as a norm. We are led to the following **variational formulation**:

---

[7] It is possible to give the definition of *ellipticity* for an operator of order higher than two (see *Renardy-Rogers*, 2004). For instance, consider the linear operator with constant coefficients $\mathcal{L} = \sum_{|\alpha|=m} a_\alpha D^\alpha$, $m \geq 2$, where $\alpha = (\alpha_1, ..., \alpha_n)$ is a multi-index. Associate with $\mathcal{L}$ its **symbol**, given by

$$S_{\mathcal{L}}(\boldsymbol{\xi}) = \sum_{|\alpha|=m} a_\alpha (i\boldsymbol{\xi})^\alpha.$$

Then $\mathcal{L}$ is said to be **elliptic** if $S_{\mathcal{L}}(\boldsymbol{\xi}) \neq 0$ for every $\boldsymbol{\xi} \in \mathbb{R}^n$, $\boldsymbol{\xi} \neq \mathbf{0}$. The symbol of $\mathcal{L} = \Delta^2$ in 2 dimensions is $-\xi_1^4 - 2\xi_1^2\xi_2^2 - \xi_2^4$, which is negative if $(\xi_1, \xi_2) \neq (0,0)$. Thus Δ^2 is elliptic. Note that, for $m = 2$, we recover the usual definition of ellipticity.

Determine $u \in H_0^2(\Omega)$ such that

$$\int_\Omega \Delta u \Delta v \, d\mathbf{x} = \int_\Omega f v \, d\mathbf{x}, \qquad \forall v \in H_0^2(\Omega). \tag{8.79}$$

The following result holds:

Proposition 8.9. *If $f \in L^2(\Omega)$, there exists a unique solution $u \in H_0^2(\Omega)$ of (8.79). Moreover,*

$$\|\Delta u\|_0 \le C_b \|f\|_0.$$

Proof. Note that the bilinear form

$$B(u, v) = \int_\Omega \Delta u \cdot \Delta v \, d\mathbf{x}$$

coincides with the inner product in $H_0^2(\Omega)$. On the other hand, setting,

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

from Corollary 8.2, we have:

$$|L(v)| = \int_\Omega |f v| \, d\mathbf{x} \le \|f\|_0 \|v\|_0 \le C_b \|f\|_0 \|\Delta v\|_0$$

so that $L \in H_0^2(\Omega)^*$. We conclude the proof directly from the Riesz Representation Theorem. $\square$

Remark 8.16. Let u be the solution of problem (8.79). Setting $w = \Delta u$, we have $\Delta w = f$ with $f \in L^2(\Omega)$. Thus, Corollary 8.2 implies $w \in H^2(\Omega)$ which, in turn, yields $u \in H^4(\Omega)$.

8.8 A Monotone Iteration Scheme for Semilinear Equations

The weak maximum principle can be used to construct iteration schemes for solving nonlinear boundary value problems. We consider here the following problem:

$$\begin{cases} -\Delta u = f(u) & \text{in } \Omega \\ u = g & \text{on } \partial\Omega. \end{cases} \tag{8.80}$$

We assume that Ω is a smooth domain and that $f \in C^1(\mathbb{R})$, $g \in H^{1/2}(\partial\Omega)$. A weak solution of problem (8.80) is a function $u \in H^1(\Omega)$ such that $u = g$ on $\partial\Omega$ and

$$\int_\Omega \nabla u \cdot \nabla v \, d\mathbf{x} = \int_\Omega f(u) v \, d\mathbf{x} \qquad \forall v \in H_0^1(\Omega). \tag{8.81}$$

We need to introduce *weak sub and super solutions*. We say that $u_* \in H^1(\Omega)$ is a *weak subsolution* of problem (8.80) if $u_* \leq g$ on $\partial\Omega$ and

$$\int_\Omega \nabla u_* \cdot \nabla v \, d\mathbf{x} \leq \int_\Omega f(u_*) v \, d\mathbf{x} \qquad \forall v \in H_0^1(\Omega), \; v \geq 0 \text{ a.e. in } \Omega.$$

Similarly, we say that $u^* \in H^1(\Omega)$ is a *weak supersolution* of problem (8.80) if $u^* \geq g$ on $\partial\Omega$ and

$$\int_\Omega \nabla u^* \cdot \nabla v \, d\mathbf{x} \geq \int_\Omega f(u^*) v \, d\mathbf{x} \qquad \forall v \in H_0^1(\Omega), \; v \geq 0 \text{ a.e. in } \Omega.$$

We want to prove the following theorem.

Theorem 8.15. *Assume that g is bounded on $\partial\Omega$ and that there exist a weak subsolution u_* and a weak supersolution u^* of problem (8.80) such that:*

$$a \leq u_* \leq g \leq u^* \leq b \qquad a, b \in \mathbb{R}.$$

Then, there exists a solution u of problem (8.80) such that

$$u_* \leq u \leq u^*.$$

Proof. Let $M = \max_{[a,b]} |f'|$. Then the function $F(s) = f(s) + Ms$ is nondecreasing. Write Poisson's equation in the form

$$-\Delta u + Mu = F(u).$$

The idea is to exploit the linear theory to define recursively the following sequence $\{u_k\}_{k \geq 1}$ of functions: let u_1 be the solution of

$$\begin{cases} -\Delta u_1 + M u_1 = F(u_*) & \text{in } \Omega \\ \qquad\qquad u_1 = g & \text{on } \partial\Omega. \end{cases}$$

Given u_k, let u_{k+1} be the solution of

$$\begin{cases} -\Delta u_{k+1} + M u_{k+1} = F(u_k) & \text{in } \Omega \\ \qquad\qquad u_k = g & \text{on } \partial\Omega. \end{cases} \tag{8.82}$$

We claim that u_k is non decreasing and trapped between u_* and u^*:

$$u_* \leq u_k \leq u_{k+1} \leq u^* \qquad \text{a.e. in } \Omega.$$

Assuming the claim, we deduce that u_k converges a.e in Ω to some bounded function u, as $k \to +\infty$. Since $F(a) \leq F(u_k) \leq F(b)$, by the Dominated Convergence Theorem we infer that

$$\int_\Omega F(u_k) v d\mathbf{x} \to \int_\Omega F(u) v d\mathbf{x} \qquad \text{as } k \to \infty,$$

for every $v \in H_0^1(\Omega)$. Now it is enough to show that there is a subsequence $\{u_{k_j}\}$ which converges weakly in $H^1(\Omega)$ to u, in order to pass to the limit in the equation

$$\int_\Omega (\nabla u_{k_j+1} \cdot \nabla v + M u_{k_j+1} v) \, d\mathbf{x} = \int_\Omega F(u_{k_j}) v \, d\mathbf{x} \qquad \forall v \in H_0^1(\Omega)$$

and obtain (8.81).

We now prove the claim. Let us check that $u_* \leq u_1$ a.e in Ω. Set $h_0 = u_* - u_1$. Then $\sup_{\partial\Omega} h_0^+ = 0$ and

$$\int_\Omega (\nabla h_0 \cdot \nabla v + M h_0 v) d\mathbf{x} \leq 0, \qquad \forall v \in H_0^1(\Omega), \, v \geq 0 \text{ a.e. in } \Omega.$$

From the proof of Theorem 8.9 we deduce $h_0 \leq 0$. Similarly, we infer that $u_1 \leq u^*$. Now assume inductively that

$$u_* \leq u_{k-1} \leq u_k \leq u^* \qquad \text{a.e. in } \Omega.$$

We prove that $u_* \leq u_k \leq u_{k+1} \leq u^*$ a.e. in Ω. Let $w_k = u_k - u_{k+1}$. We have $w_k = 0$ on $\partial\Omega$ and

$$\int_\Omega (\nabla w_k \cdot \nabla v + M w_k v) \, d\mathbf{x} = \int_\Omega [F(u_{k-1}) - F(u_k)] v \, d\mathbf{x} \qquad \forall v \in H_0^1(\Omega).$$

Since F is nondecreasing, we deduce that $F(u_{k-1}) - F(u_k) \leq 0$ a.e. in Ω so that

$$\int_\Omega (\nabla w_k \cdot \nabla v + M w_k v) \, d\mathbf{x} \leq 0 \qquad \forall v \in H_0^1(\Omega), \, v \geq 0 \text{ a.e. in } \Omega.$$

Again, the proof of Theorem 8.9 yields $w_k \leq 0$ a.e. in Ω. Similarly, we infer that $u_* \leq u_k$ and $u_{k+1} \leq u^*$.

To complete the proof we have to show that $u_k \rightharpoonup u$, weakly in $H^1(\Omega)$. This follows from the estimate for the nonhomogeneous Dirichlet problem (8.82):

$$\|u_k\|_{1,2} \leq C(n, M, \Omega) \left\{ \|F(u_{k-1})\|_0 + \|g\|_{H^{1/2}(\partial\Omega)} \right\}$$
$$\leq C_1(n, M, \Omega) \left\{ F(b) + \|g\|_{H^{1/2}(\partial\Omega)} \right\}.$$

Since $\{u_k\}$ is bounded in $H^1(\Omega)$, there exists a subsequence weakly convergent to u. $\square$

The functions u_* and u^* in the above theorem are called *lower* and *upper* barrier, respectively. Thus, Theorem 8.17 reduces the solvability of problem (8.80) to finding a *lower* and an *upper* barrier. In general we cannot assert that the solution is unique. Here is an example of non uniqueness.

Example 8.4. Consider the following problem for the stationary Fisher equation:

$$\begin{cases} -\Delta u = u(1 - u) & \text{in } \Omega \\ \quad\quad u = 0 & \text{on } \partial\Omega. \end{cases}$$

Clearly, $u_* \equiv 0$ is a solution. If we assume that the domain Ω is smooth and that the first Dirichlet eigenvalue for the Laplace operator is $\lambda_1 < 1$, we can show that there exists a solution which is positive in Ω. In fact, $u^* \equiv 1$ is an *upper* barrier. We now exhibit a positive lower barrier. Let w_1 be the nonnegative normalized eigenfunction corresponding to λ_1. From Remark 8.8 we know that $w_1 > 0$ inside Ω and from elliptic regularity, w_1 is smooth up to $\partial\Omega$. Let $u_* = \sigma w_1$. We claim that, if σ is positive and small enough, u_* is a lower barrier. Indeed, since $-\Delta w_1 = \lambda_1 w_1$, we have,

$$-\Delta u_* - u_* (1 - u_*) = \sigma w_1 (\lambda_1 - 1 + \sigma w_1). \tag{8.83}$$

If $m = \max_{\bar{\Omega}} w_1$ and $\sigma < (1 - \lambda_1)/m$, then the right hand side of (8.83) is negative and u_* is a lower barrier.

From Theorem 8.17 we infer the existence of a solution u such that $w_1 \le u \le 1$.
□

The uniqueness of the solution of problem (8.80) is guaranteed if, for instance, f is nonincreasing:
$$f'(s) \le 0, \ s \in \mathbb{R}.$$

Then, if u_1 and u_2 are two solutions of (8.80), we have $w = u_1 - u_2 \in H_0^1(\Omega)$ and we can write
$$-\Delta w = f(u_1) - f(u_2) = c(\mathbf{x}) w$$
where $c(\mathbf{x}) = f'(\bar{u}(\mathbf{x}))$, for a suitable $\bar{u}$ between u_1 and u_2. Since $c \le 0$ we conclude from the maximum principle that $w \equiv 0$ or $u_1 = u_2$.

8.9 A Control Problem

Control problems are more and more important in modern technology. We give here an application of the variational theory we have developed so far, to a fairly simple temperature control problem.

8.9.1 Structure of the problem

Suppose that the temperature u of a homogeneous body, occupying a smooth bounded domain $\Omega \subset \mathbb{R}^3$, satisfies the following stationary conditions:

$$\begin{cases} \mathcal{E}u \equiv -\Delta u + \operatorname{div}(\mathbf{b}u) = z & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{8.84}$$

where $\mathbf{b} \in C^1(\overline{\Omega}; \mathbb{R}^3)$ is given, with $\operatorname{div}\mathbf{b} \ge 0$ in Ω.

In (8.84) we distinguish two types of dependent variables: the **control** variable z, that we take in $H = L^2(\Omega)$, and the **state** variable u.

Coherently, (8.84) is called the **state system**. Given a control z, from Corollary 8.1, (8.84) has a unique weak solution

$$u[z] \in V = H_0^1(\Omega).$$

Thus, setting

$$a\left(u,v\right) = \int_{\Omega} \left(\nabla u \cdot \nabla v - u\mathbf{b}\cdot\nabla v\right) \, d\mathbf{x},$$

$u\left[z\right]$ satisfies the **state** equation

$$a\left(u\left[z\right], v\right) = \left(z, v\right)_0 \qquad \forall v \in V \tag{8.85}$$

and

$$\|u\left[z\right]\|_1 \leq \|z\|_0. \tag{8.86}$$

From elliptic regularity (Theorem 8.13) it follows that $u \in H^2\left(\Omega\right) \cap H_0^1\left(\Omega\right)$, so that u is a *strong solution* of the state equation and satisfies it in the a.e. pointwise sense as well.

Our problem is to choose the source term z in order to minimize the "distance" of u from a given target state u_d.

Of course there are many ways to measure the distance of u from u_d. If we are interested in a distance which involves u and u_d over an open subset $\Omega_0 \subseteq \Omega$, a reasonable choice may be

$$J\left(u, z\right) = \frac{1}{2} \int_{\Omega_0} \left(u - u_d\right)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} z^2 d\mathbf{x} \tag{8.87}$$

where $\beta > 0$.

$J\left(u, z\right)$ is called **cost functional** or **performance index.** The second term in (8.87) is called *penalization term*; its role is, on one hand, to avoid using "too large" controls in the minimization of J, on the other hand, to assure coercivity for J, as we shall see later on.

Summarizing, we may write our control problem in the following way:

Find $(u^*, z^*) \in H \times V$, such that

$$\begin{cases} J\left(u^*, z^*\right) = \min_{(u,z)\in V \times H} J\left(u, z\right) \\[2mm] \quad under \ the \ conditions \\[2mm] \mathcal{E}u = z \ \ in \ \Omega, \ \ u = 0 \ \ on \ \partial\Omega \end{cases} \tag{8.88}$$

If (u^*, z^*) is a minimizing pair, u^* and z^* are called **optimal state and optimal control**, respectively.

Remark 8.17. When the control z is defined in an open subset Ω_0 of Ω, we say that it is a *distributed control.* In some cases, z may be defined only on $\partial\Omega$ and then is called *boundary control.*

Similarly, when the cost functional (8.87) involves the observation of u in $\Omega_0 \subseteq \Omega$, we say that the observation is *distributed.* On the other hand, one may observe u or $\partial_{\boldsymbol{\nu}}u$ on $\Gamma \subseteq \partial\Omega$. These cases correspond to *boundary observations* and the cost functional has to take an appropriate form. Some examples are given in Problems 8.20–8.22.

The main questions to face in a control problem are:

- Establish existence and/or uniqueness of an optimal pair (u^*, z^*).
- Derive necessary and/or sufficient optimality conditions.
- Construct algorithms for the numerical approximation of (u^*, z^*).

8.9.2 Existence and uniqueness of an optimal pair

Given $z \in H$, we may substitute into J the unique solution $u = u\,[z]$ of (8.85) to get the functional

$$\tilde{J}\,(z) = J\,(u\,[z]\,, z) = \frac{1}{2} \int_{\Omega_0} (u\,[z] - u_d)^2\,d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} z^2 d\mathbf{x},$$

depending only on z. Thus, our minimization problem (8.88) is reduced to find an optimal control $z^* \in H$ such that

$$\tilde{J}\,(z^*) = \min_{z \in H} \tilde{J}\,(z)\,. \tag{8.89}$$

Once z^* is known, the optimal state is given by $u^* = u\,[z^*]$.

The strategy to prove existence and uniqueness of an optimal control is to use the relationship between minimization of quadratic functionals and abstract variational problems corresponding to symmetric bilinear forms, expressed in Proposition 6.4. The key point is to write $\tilde{J}\,(z)$ in the following way:

$$\tilde{J}\,(z) = \frac{1}{2} b\,(z, z) + Lz + q \tag{8.90}$$

where $q \in \mathbb{R}$ (irrelevant in the optimization) and:

- $b\,(z, w)$ is a bilinear form in H, *symmetric, continuous and $H-$coercive*;
- L is a *linear, continuous functional* in H.

Then, by Proposition 6.4, there exists a unique minimizer $z^* \in H$. Moreover z^* is the minimizer if and only if z^* satisfies the Euler equation (see (6.40))

$$\tilde{J}'\,(z^*)\,w = b\,(z^*, w) - Lw = 0 \qquad \forall w \in H. \tag{8.91}$$

This procedure yields the following result.

Theorem 8.16. *There exists a unique optimal control $z^* \in H$. Moreover, z^* is optimal if and only if the following Euler equation holds ($u^* = u\,[z^*]$):*

$$\tilde{J}'\,(z^*)\,w = \int_{\Omega_0} (u^* - u_d)\,u\,[w]\ d\mathbf{x} + \beta \int_{\Omega} z^* w = 0 \qquad \forall w \in H. \tag{8.92}$$

Proof. According to the above strategy, we write $\tilde{J}(z)$ in the form (8.90).

First note that the map $z \mapsto u[z]$ is *linear*. In fact, if $\alpha_1, \alpha_2 \in \mathbb{R}$, then $u[\alpha_1 z_1 + \alpha_2 z_2]$ is the solution of $\mathcal{E}u[\alpha_1 z_1 + \alpha_2 z_2] = \alpha_1 z_1 + \alpha_2 z_2 u_1$. Since $\mathcal{E}$ is linear,

$$\mathcal{E}(\alpha_1 u[z_1] + \alpha_2 u[z_2]) = \alpha_1 \mathcal{E}u[z_1] + \alpha_2 \mathcal{E}u[z_2] = \alpha_1 z_1 + \alpha_2 z_2$$

and therefore, by uniqueness, $u[\alpha_1 z_1 + \alpha_2 z_2] = \alpha_1 u[z_1] + \alpha_2 u[z_2]$.

As a consequence,

$$b(z,w) = \int_{\Omega_0} u[z]\, u[w]\, d\mathbf{x} + \beta \int_\Omega zw \tag{8.93}$$

is a bilinear form and

$$Lw = \int_{\Omega_0} u[w]\, u_d\, d\mathbf{x} \tag{8.94}$$

is a linear functional in H.

Moreover, b is symmetric (obvious), continuous and $H-$coercive. In fact, from (8.86) and the Schwarz and Poincaré inequalities, we have, since $\Omega_0 \subseteq \Omega$,

$$|b(z,w)| \leq \|u[z]\|_{L^2(\Omega_0)} \|u[w]\|_{L^2(\Omega_0)} + \beta \|z\|_0 \|w\|_0$$
$$\leq (C_P^2 + \beta) \|z\|_0 \|w\|_0$$

which gives the continuity of b. The $H-$coercivity of b follows from

$$b(z,z) = \int_{\Omega_0} u^2[z]\, d\mathbf{x} + \beta \int_\Omega z^2 \geq \beta \|z\|_0^2.$$

Finally, from (8.86) and Poincarè's inequality,

$$|Lw| \leq \|u_d\|_{L^2(\Omega_0)} \|u[w]\|_{L^2(\Omega_0)} \leq C_P \|u_d\|_0 \|w\|_0,$$

and we deduce that L is continuous in H.

Now, if we set: $q = \int_{\Omega_0} u_d^2\, d\mathbf{x}$, it is easy to check that

$$\tilde{J}(z) = \frac{1}{2} b(z,z) - Lz + q.$$

Then, Proposition 6.4 yields existence and uniqueness of the optimal control and Euler equation (8.91) translates into (8.92) after simple computations. $\square$

8.9.3 Lagrange multipliers and optimality conditions

The Euler equation (8.92) gives a characterization of the optimal control z^* but it is not suitable for its computation.

To obtain more manageable optimality conditions, let us change point of view by regarding the state equation $\mathcal{E}u[z] = -\Delta u + \mathrm{div}(\mathbf{b}u) = z$, with $u = 0$ on $\partial\Omega$,

as a *constraint* for our minimization problem. Then, the key idea is to introduce a *multiplier* $p \in V$, to be chosen suitably later on, and write $\tilde{J}(z)$ in the augmented form

$$\frac{1}{2} \int_{\Omega_0} (u[z] - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_{\Omega} z^2 d\mathbf{x} + \int_{\Omega} p(z - \mathcal{E}u[z]) \, d\mathbf{x}. \qquad (8.95)$$

In fact, we have just added zero. Since $z \longmapsto u[z]$ is a linear map,

$$\tilde{L}z = \int_{\Omega} p(z - \mathcal{E}u[z]) \, d\mathbf{x}$$

is a linear functional in H and therefore Theorem 8.15 yields the Euler equation:

$$\tilde{J}'(z^*) w = \int_{\Omega_0} (u^* - u_d) u[w] \, d\mathbf{x} + \int_{\Omega} (p + \beta z^*) w \, d\mathbf{x} - \int_{\Omega} p \, \mathcal{E}u[w] \, d\mathbf{x} = 0 \quad (8.96)$$

for every $w \in H$. Now we integrate twice by parts the last term, recalling that $u[w] = 0$ on $\partial\Omega$. We find:

$$\int_{\Omega} p\mathcal{E}u[w] \, d\mathbf{x} = \int_{\partial\Omega} p(-\partial_\nu u[w] + (\mathbf{b}\cdot\nu)u[w]) \, d\sigma + \int_{\Omega} (-\Delta p - \mathbf{b}\cdot\nabla p) u[w] \, d\mathbf{x}$$

$$= -\int_{\partial\Omega} p \, \partial_\nu u[w] \, d\sigma + \int_{\Omega} \mathcal{E}^* p \, u[w] \, d\mathbf{x},$$

where the operator $\mathcal{E}^* = -\Delta - \mathbf{b}\cdot\nabla$ is the formal adjoint of $\mathcal{E}$.

Now we choose the multiplier: let p^* be the solution of the following **adjoint** problem:

$$\begin{cases} \mathcal{E}^* p = (u^* - u_d) \chi_{\Omega_0} & \text{in } \Omega \\ p = 0 & \text{on } \partial\Omega. \end{cases} \qquad (8.97)$$

Using (8.97), the Euler equation (8.96) becomes

$$\tilde{J}'(z^*) w = \int_{\Omega} (p^* + \beta z^*) w \, d\mathbf{x} = 0 \qquad \forall w \in H, \qquad (8.98)$$

equivalent to $p^* + \beta z^* = 0$.

Summarizing, we have proved the following result:

Theorem 8.17. *The control z^* and the state $u^* = u(z^*)$ are optimal if and only if there exists a multiplier $p^* \in V$ such that z^*, u^* and p^* satisfy the following optimality conditions:*

$$\begin{cases} \mathcal{E}u^* = -\Delta u^* + \mathrm{div}\,(\mathbf{b}u^*) = z^* & \text{in } \Omega, & u^* = 0 \text{ on } \partial\Omega \\ \mathcal{E}^* p^* = -\Delta p^* - \mathbf{b}\cdot\nabla p^* = (u^* - u_d)\chi_{\Omega_0} & \text{in } \Omega, & p^* = 0 \text{ on } \partial\Omega \\ p^* + \beta z^* = 0. & \text{(Euler equation)}. \end{cases}$$

Remark 8.18. The optimal multiplier p^* is also called **adjoint state**.

Remark 8.19. We may generate the state and the adjoint equations in weak form, introducing the *Lagrangian* $\mathcal{L} = \mathcal{L}(u, z, p)$, given by

$$\mathcal{L}(u, z, p) = J(u, z) - a(u, p) + (z, p)_0.$$

Notice that $\mathcal{L}$ is linear in p, therefore[8]

$$\mathcal{L}'_p(u^*, z^*, p^*) v = -a(u^*, v) + (z^*, v)_0 = 0$$

corresponds to the state equation. Moreover

$$\mathcal{L}'_u(u^*, z^*, p^*) \varphi = J'_u(u^*, z^*) \varphi - a(\varphi, p^*)$$
$$= (u^* - u_d, \varphi)_{L^2(\Omega_0)} - a^*(p^*, \varphi) = 0$$

generates the adjoint equation, while

$$\mathcal{L}'_z(u^*, z^*, p^*) w = \beta(w, z^*)_0 + (w, p^*)_0 = 0$$

constitutes Euler equation.

Remark 8.20. It is interesting to examine the behavior of $\tilde{J}(z^*)$ as $\beta \to 0$. In our case it is possible to show that $\tilde{J}(z^*) \to 0$ as $\beta \to 0$.

8.9.4 An iterative algorithm

From Euler equation (8.98) and the Riesz Representation Theorem, we infer that

$$p^* + \beta z^* \text{ is the Riesz element associated with } \tilde{J}'(z^*),$$

called the **gradient of J at** z^* and denoted by the usual symbol $\nabla J(z^*)$ or by $\delta z(z^*, p^*)$. Thus, we have

$$\nabla J(z^*) = p^* + \beta z^*.$$

It turns out that $-\nabla J(z^*)$ plays the role of the *steepest descent* direction *for J*, as in the finite-dimensional case. This suggests an iterative procedure to compute a sequence of controls $\{z_k\}_{k \geq 0}$, convergent to the optimal one.

Select an initial control z_0. If z_k is known $(k \geq 0)$, then z_{k+1} is computed according to the following scheme.

1. Solve the state equation $a(u_k, v) = (z_k, v)_0, \forall v \in V$.

2. Knowing u_k, solve the adjoint equation

$$a^*(p_k, \varphi) = (u_k - u_d, \varphi)_{L^2(\Omega_0)} \qquad \forall \varphi \in V.$$

3. Set

$$z_{k+1} = z_k - \tau_k \nabla J(z_k) \tag{8.99}$$

---

[8] $\mathcal{L}'_p$, $\mathcal{L}'_z$ and $\mathcal{L}'_u$ denote the derivatives of the quadratic functional $\mathcal{L}$ with respect to p, z, u, respectively.

and select the *relaxation parameter* τ_k in order to assure that

$$J\left(z_{k+1}\right) < J\left(z_k\right). \tag{8.100}$$

Clearly, (8.100) implies the convergence of the sequence $\{J\left(z_k\right)\}$, though in general not to zero. Concerning the choice of the relaxation parameter, there are several possibilities. For instance, if $\beta \ll 1$, we know that the optimal value $J\left(z^*\right)$ is close to zero (Remark 8.23) and then we may chose

$$\tau_k = J\left(z_k\right)\left|\nabla J\left(z_k\right)\right|^{-2}.$$

With this choice, (8.99) is a Newton type method:

$$z_{k+1} = z_k - \frac{\nabla J\left(z_k\right)}{\left|\nabla J\left(z_k\right)\right|^2}J\left(z_k\right).$$

Also $\tau_k = \tau$, constant, may work, as in the following example, where $\tau = 10$.

Example 8.20. Let $\Omega = (0,4) \times (0,4) \subset \mathbb{R}^2$ and $\Omega_0 = (2.5, 3.5) \times (2.5, 3.5)$. Consider problem (8.88), with $u_d = \chi_{\Omega_0}$, $\beta = 10^{-4}$ and state system

$$-\Delta u + 3.5u_x + 1.5u_y = z, \text{ in } \Omega \text{ and } u = 0 \text{ on } \partial\Omega.$$

According to Theorem 8.16, there exists a unique optimal control z^*. The adjoint system is

$$-\Delta p - 3.5p_x - 1.5p_y = (u-1)\chi_{\Omega_0}, \text{ in } \Omega \text{ and } p = 0 \text{ on } \partial\Omega.$$

Figures 8.5 and 8.6 show the optimal state and the optimal control, respectively, with their isolines. Note the hole at the center of Ω_0 in the graph of z^*, in which z^* attains a negative minimum. This is due to the fact that, without control, the
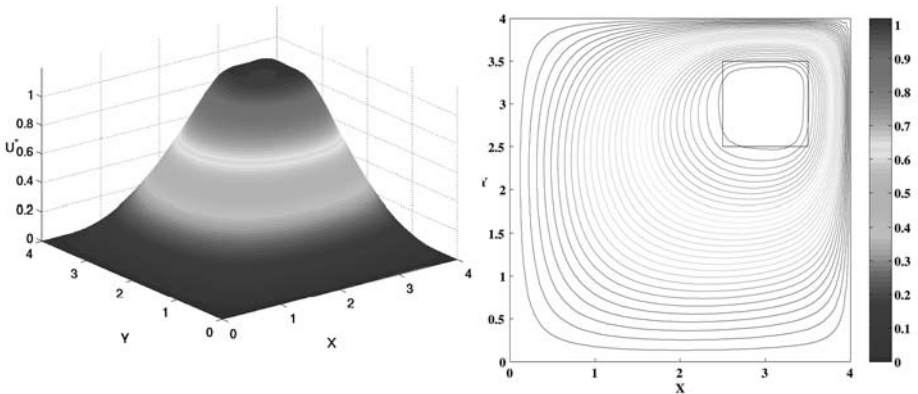
Fig. 8.5. Optimal state u^* in Example 8.20
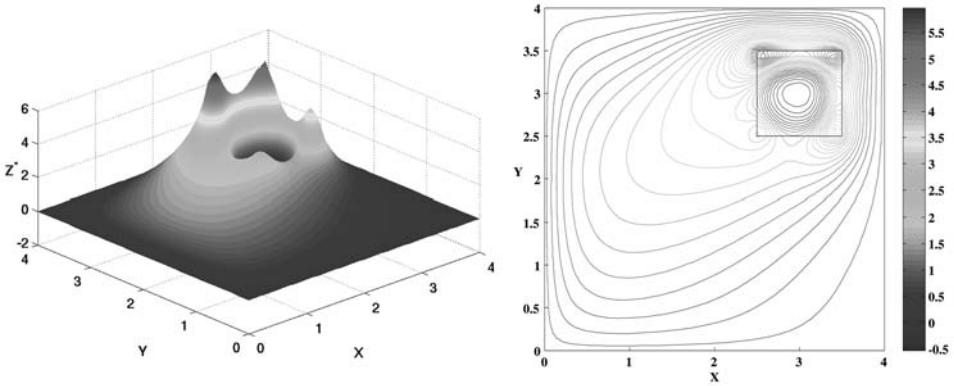
Fig. 8.6. Optimal control z^* in Example 8.20

solution of the state equation tends to be smooth and greater than one on Ω_0 so that the control has to counterbalance this effect.[9]

Problems

8.1. Consider the Dirichlet problem

$$\begin{cases} -(a(x)u')' + b(x)u' + a_0(x)u = f(x), & a < x < b) \\ u(a) = A, \ u(b) = B. \end{cases}$$

State and prove an existence, uniqueness and stability theorem.
[*Hint:* use Remark 8.3].

8.2. Write the weak formulation of the following problem:

$$\begin{cases} (x^2 + 1)u'' - xu' = \sin 2\pi x & 0 < x < 1 \\ u(0) = u(1) = 0. \end{cases}$$

Show that there exists a unique solution $u \in H_0^1(0, 1)$ and that $\|u'\|_{L^2(0,1)} \leq 1/\sqrt{2}$.

8.3. Fill in the details of the weak formulations of the Robin and mixed problem, in subsection 8.3.3.

8.4. Write the weak formulation of the following problem:

$$\begin{cases} \cos x \ u'' - \sin x \ u' - xu = 1 & 0 < x < 1 \\ u'(0) = -u(0), \ u(\pi/4) = 0 \end{cases}$$

Discuss existence and uniqueness and derive a stability estimates.

---

[9] For more on control theory see e.g. A.K. Aziz, J.W. Wingate and M.J. Balas eds, *Control Theory of Systems Governed by Partial Differential Equations*, Academic Press, 1977.

8.5. *Legendre equation.* Let

$$X = \left\{ v \in L^2\left(-1, 1\right) \colon \left(1 - x^2\right)^{1/2} \in L^2\left(-1, 1\right) \right\}$$

with inner product

$$(u, v)_X = \int_{-1}^{1} \left[uv + \left(1 - x^2\right) u'v' \right] dx.$$

a) Check that $(u, v)_X$ is indeed an inner product and that X is a Hilbert space.

b) Study the variational problem

$$(u, v)_V = \int_{-1}^{1} fv \ dx \qquad \text{for every } v \in X \qquad (8.101)$$

where $f \in L^2\left(-1, 1\right)$.

c) Determine the boundary value problem whose variational formulation is (8.101).

[a] *Hint:* Use Theorem 7.4, with $V = L^2\left(-1, 1\right)$ and $Z = L^2_w\left(-1, 1\right)$, $w\left(x\right) = \left(1 - x^2\right)^{1/2}$. Check that $Z \hookrightarrow \mathcal{D}\left(-1, 1\right)$.

b) *Hint:* Use the Lax-Milgram Theorem.

c) *Answer:* The boundary value problem is

$$\begin{cases} -\left[(1 - x^2)u'\right]' + u = f & -1 < x < 1 \\ (1 - x^2)u'\left(x\right) \to 0 & \text{as } x \to \pm 1. \end{cases}$$

This is a Legendre equation with the natural Neumann conditions at both end points].

8.6. Let $V = H^1_{per}\left(0, 2\pi\right) = \left\{ u \in H^1\left(0, 2\pi\right) : u\left(0\right) = u\left(2\pi\right) \right\}$ and F be the linear functional

$$F : v \longmapsto \int_0^{2\pi} tv\left(t\right) \ dt.$$

(a) Check that $F \in V^*$.

(b) According to Riesz's Theorem, there is a unique element $u \in V$ such that $(u, v)_{1,2} = \langle F, v \rangle_*$, for every $v \in V$. Determine explicitly u.

8.7. *Transmission conditions* (I). Consider the problem

$$\begin{cases} \left(p\left(x\right) u'\right)' = f & \text{in } \left(a, b\right) \\ u\left(a\right) = u\left(b\right) = 0 \end{cases} \qquad (8.102)$$

where $f \in L^2\left(a, b\right)$, $p\left(x\right) = p_1 > 0$ in $\left(a, c\right)$ and $p\left(x\right) = p_2 > 0$ in $\left(c, b\right)$.

Show that problem (8.102) has a unique weak solution in $H^1(a,b)$, satisfying the conditions:

$$\begin{cases} p_1 u'' = f & \text{in } (a,c) \\ p_2 u'' = f & \text{in } (c,b) \\ p_1 u'(c-) = p_2 u'(c+). \end{cases}$$

Observe the jump of the derivative of u at $x = c$ (Fig. 8.7).
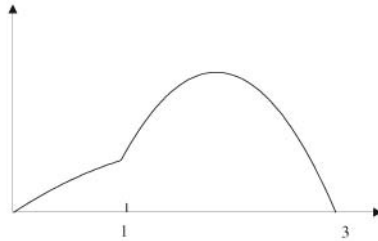
Fig. 8.7. The solution of the transmission problem $(p(x)u')' = -1$, $u(0) = u(3) = 0$, with $p(x) = 3$ in $(0,1)$ and $p(x) = 1/2$ in $(1,3)$

8.8. Let $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$. Prove that the functional

$$E(v) = \frac{1}{2} \int_\Omega \left\{ |\nabla v|^2 - xv \right\} dx dy$$

has a unique minimizer $u \in H_0^1(\Omega)$. Write the Euler equation and find an explicit formula for u.

8.9. Consider the following subspace of $H^1(\Omega)$:

$$V = \left\{ u \in H^1(\Omega) : \frac{1}{|\Omega|} \int_\Omega u \, dx = 0 \right\}.$$

a) Show that V is a Hilbert space with inner product $(\cdot, \cdot)_1$ and find which boundary value problem has the following weak formulation:

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega fv \, dx, \qquad \forall v \in V.$$

b) Show that if $f \in L^2(\Omega)$ there exists a unique solution.

8.10. Consider the following subspace of $H^1(\Omega)$:

$$V = \left\{ u \in H^1(\Omega) : \frac{1}{|\partial\Omega|} \int_{\partial\Omega} u \, d\sigma = 0 \right\}.$$

Show that V is a Hilbert space with inner product $(\cdot, \cdot)_{1,2}$ and find which boundary value problem has the following weak formulation:

$$\int_\Omega \left\{ \nabla u \cdot \nabla v + uv \right\} dx = \int_\Omega fv \, dx, \qquad \forall v \in V.$$

May we apply the Lax-Milgram Theorem?

[*Answer*: $-\Delta u + u = f$, $\partial_\nu u = $ constant; yes, we may apply it].

8.11. Let $\Omega \subset \mathbb{R}^n$ and $g \in H^{1/2}(\partial\Omega)$. Define

$$H_g^1(\Omega) = \left\{ v \in H^1(\Omega) : v = g \quad \text{on } \partial\Omega \right\}.$$

Prove the following theorem, known as **Dirichlet principle**: *Among all the functions* $v \in H_g$, *the harmonic one minimizes the Dirichlet integral*

$$D(v) = \int_\Omega |\nabla v|^2 \, d\mathbf{x}.$$

[*Hint*: In $H^1(\Omega)$ use the inner product

$$(u, v)_{1,\partial} = \int_{\partial\Omega} uv \, d\sigma + \int_\Omega \nabla u \cdot \nabla v \, d\mathbf{x}$$

and the norm (see Problem 7.24):

$$\|u\|_{1,\partial} = \left(\int_{\partial\Omega} u^2 \, d\sigma + \int_\Omega |\nabla u|^2 \, d\mathbf{x} \right)^{1/2}. \tag{8.103}$$

Then, minimizing $D(v)$ over $H_g^1(\Omega)$ amounts to minimizing $\|v\|_{1,\partial}^2$. Let $u \in H_g^1(\Omega)$, be harmonic in Ω. If $v \in H_g^1(\Omega)$, write $v = u + w$, with $w \in H_0^1(\Omega)$. Show that $(u, w)_{1,\partial} = 0$ and conclude that $\|u\|_{1,\partial}^2 \leq \|v\|_{1,\partial}^2$].

8.12. Let $\mathcal{E} = -\text{div}(\mathbf{A}(\mathbf{x})\nabla)$, with $\mathbf{A}$ symmetric. State and prove the analogues of Theorems 8.5 and 8.6.

8.13. *A simple system.* Consider the Neumann problem for the following system:

$$\begin{cases} -\Delta u_1 + u_1 - u_2 = f_1 & \text{in } \Omega \\ -\Delta u_2 + u_1 + u_2 = f_1 & \text{in } \Omega \\ \partial_\nu u_1 = \partial_\nu u_2 = 0 & \text{on } \partial\Omega. \end{cases}$$

Derive a variational formulation and establish a well-posedness theorem.

[*Hint:* Variational formulation:

$$\int_\Omega \left\{ \nabla u_1 \cdot \nabla v_1 + \nabla u_2 \cdot \nabla v_2 + u_1 v_1 - u_2 v_1 + u_1 v_2 + u_2 v_2 \right\} = \int_\Omega (f_1 v_1 + f_2 v_2)$$

for every $(v_1, v_2) \in H^1(\Omega) \times H^1(\Omega)$].

8.14. *Transmission conditions* (II). Let Ω_1 and Ω be bounded, Lipschitz domains in $\mathbb{R}^n$ such that $\Omega_1 \subset\subset \Omega$. Let $\Omega_2 = \Omega \backslash \overline{\Omega}_1$. In Ω_1 and Ω_2 consider the following bilinear forms

$$a_k(u, v) = \int_{\Omega_k} \mathbf{A}^k(\mathbf{x}) \nabla u \cdot \nabla v \, d\mathbf{x} \qquad (k = 1, 2)$$

with $\mathbf{A}^k$ *uniformly elliptic*. Assume that the entries of A^k are **continuous in** $\overline{\Omega}_k$, but that the matrix

$$\mathbf{A}(\mathbf{x}) = \begin{cases} \mathbf{A}^1(\mathbf{x}) \text{ in } \overline{\Omega}_1 \\ \mathbf{A}^2(\mathbf{x}) \text{ in } \Omega_2 \end{cases}$$

may have *a jump across* $\Gamma = \partial\Omega_1$. Let $u \in H_0^1(\Omega)$ be the weak solution of the equation

$$a(u, v) = a_1(u, v) + a_2(u, v) = (f, v)_0 \qquad \forall v \in H_0^1(\Omega),$$

where $f \in L^2(\Omega)$.

a) Which boundary value problem does u satisfy?

b) Which conditions on Γ do express the coupling between u_1 and u_2?

[*Hint: b*) $u_{1|\Gamma} = u_{2|\Gamma}$ and $\mathbf{A}^1 \nabla u_1 \cdot \boldsymbol{\nu} = \mathbf{A}^2 \nabla u_2 \cdot \boldsymbol{\nu}$, where $\boldsymbol{\nu}$ points outward with respect to Ω_1].

8.15. Find the mistake in the following argument. Consider the Neumann problem

$$\begin{cases} -\Delta u + \mathbf{c}\cdot\nabla u = f & \text{in } \Omega \\ \partial_\nu u = 0 & \text{on } \partial\Omega \end{cases} \qquad (8.104)$$

with Ω smooth, $\mathbf{c} \in C^1(\overline{\Omega})$ and $f \in L^2(\Omega)$. Let $V = H^1(\Omega)$ and

$$B(u, v) = \int_\Omega \{\nabla u \cdot \nabla v + (\mathbf{c}\cdot\nabla u)v\}.$$

If $\text{div}\mathbf{c} = 0$, we may write

$$\int_\Omega (\mathbf{c} \cdot \nabla u) u \, d\mathbf{x} = \frac{1}{2} \int_\Omega \mathbf{c} \cdot \nabla(u^2) \, d\mathbf{x} = \int_{\partial\Omega} u^2 \mathbf{c} \cdot \boldsymbol{\nu} \, d\sigma.$$

Thus, if $\mathbf{c} \cdot \boldsymbol{\nu} \geq c_0 > 0$ then, recalling Problem 8.11,

$$B(u, u) \geq \|\nabla u\|_0^2 + c_0\|u\|_{L^2(\partial\Omega)}^2 \geq C \|u\|_{1,2}^2$$

so that B is $V-$coercive and problem (8.104) has a unique solution!!

8.16. Let $\Omega = (0, \pi) \times (0, \pi)$. Study the solvability of the Dirichlet problem

$$\begin{cases} \Delta u + 2u = f & \text{in } Q \\ u = 0 & \text{on } \partial Q. \end{cases}$$

In particular, examine the cases $f(x, y) = 1$ and $f(x, y) = x - \pi/2$.

8.17. Let $B_1^+ = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, y > 0\}$. Examine the solvability of the Robin problem

$$\begin{cases} -\Delta u = f & \text{in } B_1^+ \\ \partial_\nu u + yu = 0 & \text{on } \partial B_1^+. \end{cases}$$

8.18. Let $\Omega = (0,1) \times (0,1)$, $a_0 \in \mathbb{R}$. Examine the solvability of the mixed problem

$$\begin{cases} \Delta u + a_0 u = 1 & \text{in } \Omega \\ \quad u = 0 & \text{on } \partial\Omega \setminus \{y=0\} \\ \quad \partial_\nu u = x & \text{on } \{y=0\}. \end{cases}$$

8.19. Derive a variational formulation of the following problem

$$\begin{cases} \Delta^2 u = f & \text{in } \Omega \\ \Delta u + \rho \partial_\nu u = 0 & \text{on } \partial\Omega \end{cases}$$

where ρ is a positive constant.

Show that the right functional setting is $H^2(\Omega) \cap H_0^1(\Omega)$, i.e. the space of functions in $H^2(\Omega)$ with zero trace. Prove the well-posedness of the resulting problem.

[*Hint:* The variational formulation is

$$\int_\Omega \Delta u \cdot \Delta v \, d\mathbf{x} + \int_{\partial\Omega} \rho \partial_\nu u \cdot \partial_\nu v \, d\sigma = \int_\Omega fv \, d\mathbf{x}, \qquad \forall v \in H^2(\Omega) \cap H_0^1(\Omega).$$

To show the well-posedness, use $\|\partial_\nu v\|_{L^2(\partial\Omega)} \leq \|\Delta v\|_0$, $\forall v \in H^2(\Omega) \cap H_0^1(\Omega)$].

8.20. *Distributed observation and control, Neumann conditions.* Let $\Omega \subset \mathbb{R}^n$ be a bounded, smooth domain and Ω_0 an **open** (non empty) subset of Ω. Set $V = H^1(\Omega)$, $H = L^2(\Omega)$ and consider the following control problem:

Minimize the cost functional

$$J(u,z) = \frac{1}{2} \int_{\Omega_0} (u - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_\Omega z^2 d\mathbf{x}$$

over $(u,z) \in H^1(\Omega) \times L^2(\Omega)$, with state system

$$\begin{cases} \mathcal{E}u = -\Delta u + a_0 u = z & \text{in } \Omega \\ \quad \partial_\nu u = g & \text{on } \partial\Omega \end{cases} \qquad (8.106)$$

where a_0 is a positive constant, $g \in L^2(\partial\Omega)$ and $z \in L^2(\Omega)$.

a) Show that there exists a unique minimizer.

b) Write the optimality conditions: adjoint problem and Euler equations.

[*a*] *Hint:* Follow the proof of Theorem 8.15, observing that, if $u[z]$ is the solution of (8.106) the map $z \longmapsto u[z] - u[0]$ is linear. Then write

$$\tilde{J}(z) = \frac{1}{2} \int_{\Omega_0} (u[z] - u[0] + u[0] - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_\Omega z^2 d\mathbf{x}$$

and adjust the bilinear form (8.93) accordingly.

b) *Answer:* The adjoint problem is $(\mathcal{E} = \mathcal{E}^*)$

$$\begin{cases} -\Delta p + a_0 p = (u - z_d)\chi_{\Omega_0} & \text{in } \Omega \\ \partial_\nu p = 0 & \text{on } \partial\Omega. \end{cases}$$

Where χ_{Ω_0} is the characteristic function of Ω_0. The Euler equation is: $p + \beta z = 0$ in $L^2(\Omega)$].

8.21. *Distributed observation and boundary control, Neumann conditions.* Let $\Omega \subset \mathbb{R}^n$ be a bounded, smooth domain. Consider the following control problem:

Minimize the cost functional

$$J(u, z) = \frac{1}{2}\int_\Omega (u - u_d)^2\, dx + \frac{\beta}{2}\int_{\partial\Omega} z^2 d\mathbf{x}$$

over $(u, z) \in H^1(\Omega) \times L^2(\partial\Omega)$, with state system

$$\begin{cases} -\Delta u + a_0 u = f & \text{in } \Omega \\ \partial_\nu u = z & \text{on } \partial\Omega \end{cases}$$

where a_0 is a positive constant, $f \in L^2(\Omega)$ and $z \in L^2(\partial\Omega)$.

a) Show that there exists a unique minimizer.

b) Write the optimality conditions: adjoint problem and Euler equations.

[a) *Hint:* See problem 8.20, a).

b) *Answer:* The adjoint problem is is

$$\begin{cases} -\Delta p + a_0 p = u - z_d & \text{in } \Omega \\ \partial_\nu p = 0 & \text{on } \partial\Omega \end{cases}$$

The Euler equation is: $p + \beta z = 0$ in $L^2(\partial\Omega)$].

8.22. *Boundary observation and distributed control, Dirichlet conditions.* Let $\Omega \subset \mathbb{R}^n$ be a bounded, smooth domain. Consider the following control problem:

Minimize the cost functional

$$J(u, z) = \frac{1}{2}\int_{\partial\Omega} (\partial_\nu u - u_d)^2\, d\sigma + \frac{\beta}{2}\int_\Omega z^2 d\mathbf{x}$$

over $(u, z) \in H_0^1(\Omega) \times L^2(\Omega)$, with state system

$$\begin{cases} -\Delta u + \mathbf{c} \cdot \nabla u = f + z, & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

where $\mathbf{c}$ is a constant vector and $f \in L^2(\Omega)$.

a) Show that, by elliptic regularity, $J(u, z)$ is well defined and that there exists a unique minimizer.

b) Write the optimality conditions: adjoint problem and Euler equations.

9

Weak Formulation of Evolution Problems

Parabolic Equations – Diffusion Equation – General Equations – The Wave Equation

9.1 Parabolic Equations

In Chapter 2 we have considered the diffusion equation and some of its generalizations, as in the reaction-diffusion model (Section 2.5) or in the Black-Scholes model (Section 2.9). This kind of equations belongs to the class of *parabolic equations,* that we have already classified in spatial dimension 1, in subsection 2.4.1 and that we are going to define in a more general setting.
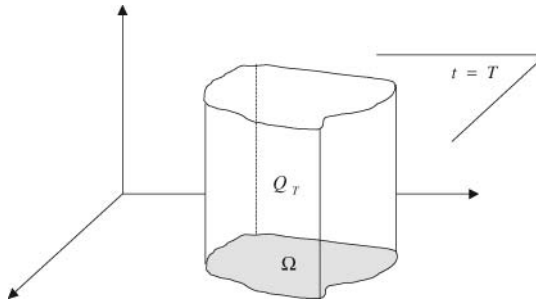
Fig. 9.1. Space-time cylinder

Let $\Omega \subset \mathbb{R}^n$ be a *bounded* domain, $T > 0$ and consider the space-time cylinder $Q_T = \Omega \times (0, T)$. Let $\mathbf{A} = \mathbf{A}(\mathbf{x}, t)$ be a square matrix of order n, $\mathbf{b} = \mathbf{b}(\mathbf{x}, t)$, $\mathbf{c} = \mathbf{c}(\mathbf{x}, t)$ vectors in $\mathbb{R}^n$, $a_0 = a_0(\mathbf{x}, t)$ and $f = f(\mathbf{x}, t)$ real functions. Equations *in divergence form* of the type

$$u_t - \mathrm{div}(\mathbf{A}\nabla u - \mathbf{b}u) + \mathbf{c} \cdot \nabla u + a_0 u = f \tag{9.1}$$

or in *non divergence form* of the type

$$u_t - \text{Tr} \left(\mathbf{A} D^2 u \right) + \mathbf{b} \cdot \nabla u + a_0 u = f \tag{9.2}$$

are called **parabolic** in Q_T if

$$A \left(\mathbf{x}, t \right) \boldsymbol{\xi} \cdot \boldsymbol{\xi} > 0 \qquad \forall (\mathbf{x}, t) \in Q_T, \forall \boldsymbol{\xi} \in \mathbb{R}^n, \ \boldsymbol{\xi} \neq \mathbf{0}.$$

For parabolic equations we may repeat the arguments concerning elliptic equations in Sections 9.1 and 9.2. Also in this case, different notions of solutions may be given, with the obvious corrections due to the evolutionary nature of (9.1) and (9.2). For identical reasons, we develop the theory for divergence form equations. Thus, let

$$\mathcal{E} u = -\text{div}(\mathbf{A} \nabla u - \mathbf{b} u) + \mathbf{c} \cdot \nabla u + a_0 u.$$

Given f in Q_T, we want to determine a solution u, of the *parabolic* equation

$$u_t + \mathcal{E} u = f \qquad \text{in } Q_T$$

satisfying an *initial (or Cauchy)* condition

$$u \left(\mathbf{x}, 0 \right) = u_0 \left(\mathbf{x} \right) \text{ in } \Omega$$

and one of the usual boundary conditions (*Dirichlet, Neumann, mixed or Robin*) on the lateral boundary $S_T = \partial \Omega \times [0, T]$.

The *star* among parabolic equations is clearly the heat equation. We use the Cauchy-Dirichlet problem for this equation to introduce a possible *weak formulation*. This approach requires the use of integrals for function with values in a Hilbert space and of Sobolev spaces involving time. A brief account of these notions is presented in Section 7.11.

9.2 Diffusion Equation

9.2.1 The Cauchy-Dirichlet problem

Suppose we are given the problem

$$\begin{cases} u_t - \alpha \Delta u = f & \text{in } Q_T \\ u \left(\mathbf{x}, 0 \right) = g \left(\mathbf{x} \right) & \text{in } \Omega \\ u \left(\boldsymbol{\sigma}, t \right) = 0 & \text{on } S_T \end{cases} \tag{9.3}$$

where $\alpha > 0$.

We want to find a weak formulation. Let us proceed formally. As we did several times in Chapter 1, we multiply the diffusion equation by a smooth function $v = v \left(\mathbf{x} \right)$, vanishing at the boundary of Ω, and integrate over Ω. We find

$$\int_\Omega u_t \left(\mathbf{x}, t \right) v \left(\mathbf{x} \right) d\mathbf{x} - \alpha \int_\Omega \Delta u \left(\mathbf{x}, t \right) v \left(\mathbf{x} \right) d\mathbf{x} = \int_\Omega f \left(\mathbf{x}, t \right) v \left(\mathbf{x} \right) d\mathbf{x}.$$

Integrating by parts the second term, we get

$$\int_{\Omega} u_t(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x} + \alpha \int_{\Omega} \nabla u(\mathbf{x},t) \cdot \nabla v(\mathbf{x})\, d\mathbf{x} = \int_{\Omega} f(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x}. \qquad (9.4)$$

This looks like what we did for elliptic equations, except for the presence of u_t. Moreover, here we will have somehow to take into account the initial condition. Which could be a correct functional setting?

First of all, since we are dealing with evolution equations, it is convenient to adopt the point of view of section 7.11, and consider $u = u(\mathbf{x}, t)$ as a function of t with values into a suitable Hilbert space V:

$$u \colon [0, T] \to V.$$

When we adopt this convention, we write $u(t)$ instead of $u(\mathbf{x}, t)$ and $\dot{u}$ instead of u_t. Accordingly, we write $f(t)$ instead of $f(\mathbf{x}, t)$. With these notations, (9.4) becomes

$$\int_{\Omega} \dot{u}(t)\, v\, d\mathbf{x} + \alpha \int_{\Omega} \nabla u(t) \cdot \nabla v\, d\mathbf{x} = \int_{\Omega} f(t)\, v\, d\mathbf{x}. \qquad (9.5)$$

The homogeneous Dirichlet condition, i.e. $u(t) = 0$ on $\partial\Omega$ for $t \in [0, T]$, suggests that the natural space for $u(t)$ is $V = H_0^1(\Omega)$, at least for a.e. $t \in [0, T]$. As usual, in $H_0^1(\Omega)$ we use the inner product

$$(w, v)_1 = (\nabla w, \nabla v)_0$$

with corresponding norm $\|\cdot\|_1$. Thus, the second integral in (9.5) may be written in the form

$$(\nabla u(t), \nabla v)_0\,.$$

Also, it would seem to be appropriate that $\dot{u}(t) \in L^2(\Omega)$, looking at the first integral. This however is not coherent with the choice $u(t) \in H_0^1(\Omega)$, since we have $\Delta u(t) \in H^{-1}(\Omega)$ and

$$\dot{u}(t) = \alpha \Delta u(t) + f(t) \qquad (9.6)$$

from the diffusion equation. Thus, we deduce that $H^{-1}(\Omega)$ is the natural space for $\dot{u}$ as well. Consequently, the first integral in (9.5) has to be interpreted as

$$\langle \dot{u}(t), v \rangle_*$$

where $\langle \cdot, \cdot \rangle_*$ denotes the pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

A reasonable hypothesis on f is $f \in L^2(Q_T)$, which in the new notations becomes[1]

$$f \in L^2\left(0, T; L^2(\Omega)\right).$$

Coherently, from (9.6) we require $u \in L^2\left(0, T; H_0^1(\Omega)\right)$ and $\dot{u} \in L^2\left(0, T; H^{-1}(\Omega)\right)$. Now, from Theorem 7.22 we know that

$$u \in C\left([0, T]; L^2(\Omega)\right)$$

so that the initial condition $u(0) = g$ makes perfect sense if we choose $g \in L^2(\Omega)$.

---

[1] Also $f \in L^2(0, T; V^*)$ is fine.

The above arguments motivate the following definition. Consider the Hilbert triplet (V, H, V^*), where $V = H_0^1(\Omega)$, $H = L^2(\Omega)$ and $V^* = H^{-1}(\Omega)$. Recall that Poincaré's inequality holds in V:

$$\|v\|_0 \leq C_P \|v\|_1.$$

Finally, let

$$a(w, v) = \alpha(\nabla w, \nabla v)_0$$

Definition 9.1. *A function $u \in L^2(0, T; V)$ is called **weak solution** of problem (9.3) if $\dot{u} \in L^2(0, T; V^*)$ and:*

1. for every $v \in V$,

$$\langle \dot{u}(t), v \rangle_* + a(u(t), v) = (f(t), v)_0 \qquad a.e.\ t \in [0, T]. \qquad (9.7)$$

2. $u(0) = g$.

Remark 9.1. Equation (9.7) may be interpreted in the sense of distributions. To see this, observe that, for every $v \in V$, the real function

$$w(t) = \langle \dot{u}(t), v \rangle_*$$

is a distribution in $\mathcal{D}'(0, T)$ and

$$\langle \dot{u}(t), v \rangle_* = \frac{d}{dt}(u(t), v)_0 \qquad \text{in } \mathcal{D}'(0, T). \qquad (9.8)$$

This means that, for every $\varphi \in \mathcal{D}(0, T)$, we have

$$\int_0^T \langle \dot{u}(t), v \rangle_* \varphi(t)\, dt = -\int_0^T (u(t), v) \dot{\varphi}(t)\, dt.$$

In fact, since $u(t) \in V$, by Bochner's Theorem 7.20 and the definition of $\dot{u}$, we may write

$$\int_0^T \langle \dot{u}(t), v \rangle_* \varphi(t)\, dt = \langle \int_0^T \dot{u}(t)\varphi(t)\, dt, v \rangle_* = \langle -\int_0^T u(t)\dot{\varphi}(t)\, dt, v \rangle_*.$$

On the other hand, $\int_0^T u(t)\dot{\varphi}(t)\, dt \in V$ so that [2]

$$\langle -\int_0^T u(t)\dot{\varphi}(t)\, dt, v \rangle_* = (-\int_0^T u(t)\dot{\varphi}(t)\, dt\ , v)_0 = -\int_0^T (u(t), v)_0 \dot{\varphi}(t)\, dt.$$

Thus $w \in L_{loc}^1(0, T) \subset \mathcal{D}'(0, T)$ and (9.8) is true. As a consequence, equation (9.7) may be written in the form

$$\frac{d}{dt}(u(t), v)_0 + a(u(t), v) = (f, v)_0 \qquad (9.9)$$

in the sense of distributions in $\mathcal{D}'(0, T)$, for all $v \in V$.

Remark 9.2. We leave it to the reader to check that if a *weak* solution u is *smooth*, i.e. $u \in C^{2,1}(\overline{Q}_T)$, then u is a classical solution.

---

[2] Recall from Section 6.8 that if $u \in H$ and $v \in V$, $\langle u, v \rangle_* = (u, v)_0$.

9.2.2 Faedo-Galerkin method (I)

We want to show that problem (9.3) has exactly one weak solution, which depends continuously on the data in a suitable norm.

Although there are variants of the Lax-Milgram Theorem perfectly adapted to solve evolution problems, we shall use the so-called Faedo-Galerkin method, also more convenient for numerical approximations. Let us describe the main strategy.

1. We select a sequence of smooth functions $\{w_k\}_{k=1}^\infty$ constituting[3]

$$\text{an } orthogonal \text{ basis in } V = H_0^1(\Omega)$$

and

$$\text{an } orthonormal \text{ basis in } H = L^2(\Omega).$$

In particular, we can write

$$g = \sum_{k=1}^\infty g_k w_k$$

where $g_k = (g, w_k)_0$ and the series converges in H.

2. We construct the sequence of finite-dimensional subspaces

$$V_m = \text{span}\{w_1, w_2, ..., w_m\}.$$

Clearly

$$V_m \subset V_{m+1} \quad \text{and} \quad \overline{\cup V_m} = V.$$

For m fixed, let

$$u_m(t) = \sum_{k=1}^m c_k(t) w_k, \qquad G_m = \sum_{k=1}^m g_k w_k. \tag{9.10}$$

We solve the following approximate problem: *Determine* $u_m \in H^1(0, T; V)$, *satisfying, for every* $s = 1, ..., m$,

$$\begin{cases} (\dot{u}_m(t), w_s)_0 + a(u_m(t), w_s) = (f(t), w_s)_0, & \text{a.e } t \in [0, T] \\ u_m(0) = G_m. \end{cases} \tag{9.11}$$

Note that the differential equation in (9.11) is true for each element of the basis w_s, $s = 1, ..., m$, if and only if it is true for every $v \in V_m$. Moreover, since $\dot{u}_m \in L^2(0, T; V)$, we have

$$(\dot{u}_m(t), v)_0 = \langle \dot{u}_m(t), v \rangle_*.$$

We call u_m a *Galerkin approximation* of the solution u.

3. We show that $\{u_m\}$ and $\{\dot{u}_m\}$ are bounded in $L^2(0, T; V)$ and $L^2(0, T; V^*)$, respectively (*energy estimates*). Then, the weak compactness Theorem 6.11 implies that a subsequence $\{u_{m_k}\}$ converges weakly in $L^2(0, T; V)$ to some element u, while $\{\dot{u}_{m_k}\}$ converges weakly in $L^2(0, T; V^*)$ to $\dot{u}$.

4. We prove that u in step 3 is the unique weak solution of problem (9.3).

---

[3] This is possible since V is a separable Hilbert space. In particular, here we can choose as w_k the Dirichlet eigenfunctions of the Laplace operator, normalized with respect to the norm in H (see Theorem 8.5).

9.2.3 Solution of the approximate problem

The following lemma holds:

Lemma 9.1. *For all m, there exists a unique solution u_m of problem (9.11). In particular, since $u_m \in H^1(0, T; V_m)$, we have $u_m \in C([0, T]; V_m)$.*

Proof. Since $w_1, ..., w_m$ are mutually orthonormal in $L^2(\Omega)$, we have

$$(\dot{u}_m(t), w_s)_0 = \left(\sum_{k=1}^{m} \dot{c}_k(t) w_k, w_s \right)_0 = \dot{c}_s(t).$$

Also, $w_1, ..., w_m$ is an orthogonal system in V_m, hence

$$a \left(\sum_{k=1}^{m} c_k(t) w_k, w_s \right) = \alpha (\nabla w_s, \nabla w_s)_0 \, c_s(t) = \alpha \, \|\nabla w_s\|_0^2 \, c_s(t).$$

Let

$$F_s(t) = (f(t), w_s), \qquad \mathbf{F}_m(t) = (F_1(t), ..., F_m(t))$$

and

$$\mathbf{C}_m(t) = (c_1(t), ..., c_m(t)), \qquad \mathbf{g}_m = (g_1, ..., g_m).$$

If we introduce the diagonal matrix

$$\mathbf{W} = \text{diag} \left\{ \|\nabla w_1\|_0^2, \|\nabla w_2\|_0^2, ..., \|\nabla w_m\|_0^2 \right\}$$

of order m, problem (9.11) is equivalent to the following system of m *uncoupled* linear ordinary differential equations, with constant coefficients:

$$\dot{\mathbf{C}}_m(t) = -\alpha \mathbf{W} \mathbf{C}_m(t) + \mathbf{F}_m(t), \qquad \text{a.e. } t \in [0, T] \tag{9.12}$$

with initial condition

$$\mathbf{C}_m(0) = \mathbf{g}_m.$$

Since $\mathbf{F} \in L^2(0, T; \mathbb{R}^m)$, there exists a unique solution $\mathbf{C}_m(t) \in H^1(0, T; \mathbb{R}^m)$. From

$$u_m(t) = \sum_{k=1}^{m} c_k(t) w_k,$$

we deduce that $u_m \in H^1(0, T; V_m)$. $\square$

Remark 9.3. We have chosen a basis $\{w_k\}$ orthonormal in L^2 and orthogonal in H_0^1 because with respect to this base, the Laplace operator becomes a *diagonal operator*, as it is reflected by the approximate problem (9.12). However, the method works using any countable basis for both spaces. Problem (9.11) becomes

$$\dot{\mathbf{C}}_m(t) = -\mathbf{M}^{-1} \mathbf{W} \mathbf{C}_m(t) + \mathbf{M}^{-1} \mathbf{F}_m(t) \qquad \text{a.e. } t \in [0, T]$$

where[4]

$$\mathbf{M} = (M_{sk}), \qquad M_{sk} = (w_s, w_k)_0,$$

$$\mathbf{W} = (W_{sk}), \qquad W_{sk} = \alpha (\nabla w_s, \nabla w_k)_0.$$

This is particularly important in the numerical implementation of the method, where, in general, the elements of the basis in V_m are not mutually orthogonal.

9.2.4 Energy estimates

Our purpose is to show that we can extract from the sequence of Galerkin approximations $\{u_m\}$ a subsequence converging in some sense to a solution of problem (9.3). This is a typical compactness problem in Hilbert spaces. The key tool is Theorem 6.11: *let H be a Hilbert space and $\{x_m\} \subset H$ a bounded sequence. Then, $\{x_m\}$ has a subsequence $\{x_{m_k}\}$ weakly convergent to $x \in H$. Moreover*

$$\|x\| \leq \lim_{k \to \infty} \inf \|x_{m_k}\|. \tag{9.13}$$

Thus, what we need is to show that suitable Sobolev norms of u_m can be estimated by suitable norms of the data, **and the estimates are independent of m**. Moreover, these estimates must be powerful enough in order to pass to the limit as $m \to +\infty$ in the approximating equation

$$(\dot{u}_m, v)_0 + \alpha (\nabla u_m, \nabla v)_0 = (f, v)_0.$$

In our case we will be able to control the norms of u_m in $L^\infty(0, T; H)$ and $L^2(0, T; V)$, and the norm of $\dot{u}_m$ in $L^2(0, T; V^*)$, that is the norms

$$\max_{t \in [0,T]} \|u_m(t)\|_0, \quad \int_\Omega \|u_m(t)\|_1^2 \, dt \quad \text{and} \quad \int_\Omega \|\dot{u}_m(t)\|_*^2 \, dt$$

Thus, let $u_m = \sum_{k=1}^m c_k(t) w_k$ be the solution of problem (9.11).

Theorem 9.1. (Estimate of u_m). *For every $t \in [0, T]$, the following estimate holds:*

$$\|u_m(t)\|_0^2 + \alpha \int_0^t \|u_m(s)\|_1^2 \, ds \leq \|g\|_0^2 + \frac{C_P^2}{\alpha} \int_0^t \|f(s)\|_0^2 \, ds. \tag{9.14}$$

Note in particular how estimate (9.14) deteriorates as α approaches to zero. An alternative estimate is given in Problem 9.3.

Proof. Multiplying equation (9.11) by $c_k(t)$ and summing for $k = 1, ..., m$, we get

$$(\dot{u}_m(t), u_m(t))_0 + a(u_m(t), u_m(t)) = (f(t), u_m(t))_0 \tag{9.15}$$

---

[4] Since $w_1, ..., w_m$ is a basis in V_m, the matrix $\mathbf{M}$ is positive, hence non singular.

for a.e. $t \in [0, T]$. Now, note that

$$\left(\dot{u}_m\left(t\right), u_m\left(t\right)\right)_0 = \frac{1}{2}\frac{d}{dt}\left\|u_m\left(t\right)\right\|_0^2, \qquad \text{a.e. } t \in (0, T)$$

and

$$a\left(u_m\left(t\right), u_m\left(t\right)\right) = \alpha\left\|\nabla u_m\left(t\right)\right\|_0^2 = \alpha\left\|u_m\left(t\right)\right\|_1^2.$$

From the inequalities of Schwarz and Poincaré and the elementary inequality

$$\left|ab\right| \leq \frac{a^2}{2\varepsilon} + \frac{\varepsilon}{2}b^2 \qquad \forall a, b \in \mathbb{R}, \forall \varepsilon > 0 \tag{9.16}$$

with $\varepsilon = \alpha$, we deduce

$$\left(f\left(t\right), u_m\left(t\right)\right)_0 \leq \left\|f\left(t\right)\right\|_0 \left\|u_m\left(t\right)\right\|_0 \leq C_P \left\|f\left(t\right)\right\|_0 \left\|u_m\left(t\right)\right\|_1$$
$$\leq \frac{C_P^2}{2\alpha}\left\|f\left(t\right)\right\|_0^2 + \frac{\alpha}{2}\left\|u_m\left(t\right)\right\|_1^2.$$

Thus, from (9.15) we obtain

$$\frac{d}{dt}\left\|u_m\left(t\right)\right\|_0^2 + \alpha\left\|u_m\left(t\right)\right\|_1^2 \leq \frac{C_P^2}{\alpha}\left\|f\left(t\right)\right\|_0^2.$$

We now integrate over $(0, t)$, using formula (7.70) in Remark 7.34. Since $u_m\left(0\right) = G_m$ and observing that

$$\left\|G_m\right\|_0^2 \leq \left\|g\right\|_0^2$$

by the orthogonality of $w_1, ..., w_m$ in $L^2\left(\Omega\right)$, we may write:

$$\left\|u_m\left(t\right)\right\|_0^2 + \alpha\int_0^t\left\|u_m\left(s\right)\right\|_1^2 ds \leq \left\|G_m\right\|_0^2 + \frac{C_P^2}{\alpha}\int_0^t\left\|f\left(s\right)\right\|_0^2 ds \tag{9.17}$$

$$\leq \left\|g\right\|_0^2 + \frac{C_P^2}{\alpha}\int_0^t\left\|f\left(s\right)\right\|_0^2 ds$$

which is (9.14). $\square$

We now give an estimate of the norm of $\dot{u}_m$ in $L^2\left(0, T; V^*\right)$.

Theorem 9.2. (Estimate of $\dot{u}_m$). *The following estimate holds:*

$$\int_0^T\left\|\dot{u}_m\left(t\right)\right\|_*^2 \, dt \leq 2\alpha\left\|g\right\|_0^2 + 4C_P^2\int_0^T\left\|f\left(t\right)\right\|_0^2 dt \tag{9.18}$$

Proof. Let $v \in V$ and write

$$v = w + z$$

where $w \in V_m$ and $z \in V_m^\perp$. We have

$$\left\|w\right\|_1 \leq \left\|v\right\|_1.$$

Let $v = w$ in problem (9.11); this yields

$$\left(\dot{u}_m\left(t\right),v\right)_0 = \left(\dot{u}_m\left(t\right),w\right)_0 = -a\left(u_m\left(t\right),w\right) + \left(f\left(t\right),w\right)_0.$$

Since

$$\left|a\left(u_m\left(t\right),w\right)\right| \leq \alpha \left\|u_m\left(t\right)\right\|_1 \left\|w\right\|_1$$

we infer, using the Schwarz and Poincaré inequalities,

$$\begin{aligned}\left|\left(\dot{u}_m\left(t\right),v\right)_0\right| &\leq \alpha \left\|u_m\left(t\right)\right\|_1 \left\|w\right\|_1 + \left\|f\left(t\right)\right\|_0 \left\|w\right\|_0 \\ &\leq \left\{\alpha \left\|u_m\left(t\right)\right\|_1 + C_P \left\|f\left(t\right)\right\|_0\right\} \left\|w\right\|_1 \\ &\leq \left\{\alpha \left\|u_m\left(t\right)\right\|_1 + C_P \left\|f\left(t\right)\right\|_0\right\} \left\|v\right\|_1 .\end{aligned}$$

Then, by the definition of norm in V^*, we may write

$$\left\|\dot{u}_m\left(t\right)\right\|_* \leq \alpha \left\|u_m\left(t\right)\right\|_1 + C_P \left\|f\left(t\right)\right\|_0 .$$

Squaring both sides and integrating over $(0,t)$ we get[5]

$$\int_0^t \left\|\dot{u}_m\left(s\right)\right\|_*^2 \, ds \leq 2\alpha^2 \int_0^t \left\|u_m\left(s\right)\right\|_1^2 \, ds + 2C_P^2 \int_0^t \left\|f\left(s\right)\right\|_0^2 \, ds.$$

Using (9.14) to estimate $2\alpha^2 \int_0^t \left\|u_m\left(s\right)\right\|_1^2 \, ds$, we easily obtain (9.18). $\square$

9.2.5 Existence, uniqueness and stability

Theorems 9.1 and 9.2. show that the sequence of Galerkin's approximations $\{u_m\}$ is bounded in $L^\infty\left(0,T;V\right)$, hence in $L^2\left(0,T;V\right)$, while $\{\dot{u}_m\}$ is bounded in $L^2\left(0,T;V^*\right)$.

We now use the compactness Theorem 6.11 and deduce that there exists a subsequence, which for simplicity we still denote by $\{u_m\}$, such that, as $m \to \infty$,

$$u_m \rightharpoonup u \quad \text{weakly in } L^2\left(0,T;V\right)$$

and[6]

$$\dot{u}_m \rightharpoonup \dot{u} \quad \text{weakly in } L^2\left(0,T;V^*\right).$$

This u is the unique solution of problem (9.3). Precisely:

Theorem 9.3. Let $f \in L^2\left(0,T;L^2\left(\Omega\right)\right)$ and $g \in L^2\left(\Omega\right)$. Then, u is the unique solution of problem (9.3). Moreover

$$\left\|u\left(t\right)\right\|_0^2 + \alpha \int_0^T \left\|u\left(t\right)\right\|_1^2 dt \leq \left\|g\right\|_0^2 + \frac{2C_P^2}{\alpha} \int_0^T \left\|f\left(t\right)\right\|_0^2 dt \qquad (9.19)$$

for every $t \in [0,T]$, and

$$\int_0^T \left\|\dot{u}\left(t\right)\right\|_*^2 dt \leq 2\alpha \left\|g\right\|_0^2 + 4C_P^2 \int_0^T \left\|f\left(t\right)\right\|_0^2 dt. \qquad (9.20)$$

---

[5] $(a+b)^2 \leq 2a^2 + 2b^2$
[6] Rigorously: $\dot{u}_m \to v$ in $L^2\left(0,T;V^*\right)$ and one checks that $v = \dot{u}$.

Proof. **Existence.** To say that $u_m \rightharpoonup u$, *weakly in* $L^2(0, T; V)$ as $m \to \infty$, means that

$$\int_0^T (\nabla u_m(t), \nabla v(t))_0 \, dt \to \int_0^T (\nabla u(t), \nabla v(t))_0 \, dt$$

for all $v \in L^2(0, T; V)$. Similarly, $\dot{u}_m \rightharpoonup \dot{u}$, *weakly in* $L^2(0, T; V^*)$, means that

$$\int_0^T (\dot{u}_m(t), v(t))_0 \, dt = \int_0^T \langle \dot{u}_m(t), v(t) \rangle_* \, dt \to \int_0^T \langle \dot{u}(t), v(t) \rangle_* \, dt$$

for all $v \in L^2(0, T; V)$.

We want to use these properties to pass to the limit as $m \to +\infty$ in problem (9.11), keeping in mind that the test functions have to be chosen in V_m. Fix $v \in L^2(0, T; V)$; we may write

$$v(t) = \sum_{k=1}^{\infty} b_k(t) w_k$$

with the series convergent in V, for a.e. $t \in [0, T]$. Let

$$v_N(t) = \sum_{k=1}^{N} b_k(t) w_k \tag{9.21}$$

and keep N fixed, for the time being. If $m \ge N$, then $v_N \in L^2(0, T; V_m)$. Multiplying equation (9.11) by $b_k(t)$ and summing for $k = 1, ..., N$, we get

$$(\dot{u}_m(t), v_N(t))_0 + \alpha (\nabla u_m(t), \nabla v_N(t))_0 = (f(t), v_N(t))_0.$$

An integration over $(0, T)$ yields

$$\int_0^T \{(\dot{u}_m, v_N)_0 + \alpha (\nabla u_m, \nabla v_N)_0\} \, dt = \int_0^T (f, v_N)_0 \, dt. \tag{9.22}$$

Thanks to the weak convergence of u_m and $\dot{u}_m$ in their respective spaces, we can let $m \to +\infty$. Since

$$\int_0^T (\dot{u}_m, v_N)_0 \, dt = \int_0^T \langle \dot{u}_m, v_N \rangle_* \, dt \to \int_0^T \langle \dot{u}, v_N \rangle_* \, dt,$$

we obtain

$$\int_0^T \{\langle \dot{u}, v_N \rangle_* + \alpha (\nabla u, \nabla v_N)_0\} \, dt = \int_0^T (f, v_N)_0 \, dt.$$

Now, let $N \to \infty$ observing that $v_N \to v$ in $L^2(0, T; V)$ and in particular weakly in this space as well. We obtain

$$\int_0^T \{\langle \dot{u}, v \rangle_* + \alpha (\nabla u, \nabla v)_0\} \, dt = \int_0^T (f, v)_0 \, dt. \tag{9.23}$$

Then, (9.23) is valid for all $v \in L^2 (0, T; V)$. This entails[7]

$$\langle \dot{u} (t) , v \rangle_* + \alpha (\nabla u (t) , \nabla v)_0 \, dt = (f (t) , v)_0$$

for all $v \in V$ and a.e. $t \in [0, T]$. Therefore u satisfies (9.7). From Theorem 7.22, we know that $u \in C ([0, T]; H)$.

It remains to check that $u (t)$ satisfies the initial condition $u (0) = g$. Let $v \in C^1 ([0, T]; V)$ with $v (T) = 0$. Integrating by parts (see Theorem 7.22, b)), we obtain

$$\int_0^T (\dot{u}_m, v_N)_0 \, dt = (G_m, v_N (0))_0 - \int_0^T (u_m, \dot{v}_N)_0 \, dt$$

so that, from (9.22) we find

$$- \int_0^T \{ (u_m, \dot{v}_N)_0 + \alpha (\nabla u_m, \nabla v_N)_0 \} \, dt = - (G_m, v_N (0))_0 + \int_0^T (f, v_N)_0 \, dt.$$

Let first $m \to \infty$ and then $N \to \infty$; we get

$$- \int_0^T \{ (u, \dot{v})_0 + \alpha (\nabla u, \nabla v)_0 \} \, dt = - (g, v (0))_0 + \int_0^T (f, v)_0 \, dt. \qquad (9.25)$$

On the other hand, integrating by parts in formula (9.23) (see again Theorem 7.22, b)) we find

$$- \int_0^T \{ (u, \dot{v})_0 + \alpha (\nabla u, \nabla v)_0 \} \, dt = (u (0) , v (0))_0 + \int_0^T (f (t) , v (t))_0 \, dt. \qquad (9.26)$$

Subtracting (9.25) from (9.26), we deduce

$$(u (0) , v (0))_0 = (g, v (0))_0$$

and the arbitrariness of $v (0)$ forces

$$u (0) = g.$$

Uniqueness. Let u_1 and u_2 be weak solutions of the same problem. Then, $w = u_1 - u_2$ is a weak solution of

$$\langle \dot{w} (t) , v \rangle_* + \alpha (\nabla w (t) , \nabla v)_0 = 0$$

for all $v \in V$ and a.e. $t \in [0, T]$, with initial data $w (0) = 0$. Choosing $v = w (t)$ we have

---

[7] Precisely: equation (9.23) is valid, in particular, for $v (t)$ of the form $w_k \varphi (t)$, with $\varphi \in L^2 (0, T)$. Therefore, for each k there is a set E_k of measure zero, such that

$$\langle \dot{u} (t) , w_k \rangle_* + \alpha (\nabla u (t) , \nabla w_k)_0 \, dt = (f (t) , w_k)_0 \qquad (9.24)$$

for all $t \notin E_k$. Then, (9.24) holds for every k, as long as $t \notin \cup_{k \geq 1} E_k$. Since $|\cup_{k \geq 1} E_k| = 0$ and $\{w_k\}$ is a basis in V, we conclude that (9.24) holds for all $v \in V$, a.e. $t \in [0, T]$.

$$\langle \dot{w}(t), w(t) \rangle_* + \alpha \left(\nabla w(t), \nabla w(t) \right)_0 = 0$$

or, using Remark 7.34,

$$\frac{1}{2} \frac{d}{dt} \|w(t)\|_0^2 = -\alpha \|w(t)\|_1^2$$

whence, since $\|w(0)\|_0^2 = 0$,

$$\|w(t)\|_0^2 = -\int_0^T -\alpha \|w(t)\|_1^2 \, dt < 0$$

which entails $w(t) = 0$ for all $t \in [0,T]$. This gives uniqueness of the weak solution.

Stability estimates. Letting $m \to +\infty$ in (9.14) and (9.18) we get, using (9.13) and Proposition 7.16,

$$\|u\|_{L^\infty(0,T;H)}^2, \ \|u\|_{L^2(0,T;V)}^2 \leq \|g\|_0^2 + \frac{C_P^2}{\alpha} \int_0^T \|f\|_0^2 \, dt$$

and

$$\|\dot{u}\|_{L(0,T;V^*)}^2 \leq 2\alpha \|g\|_0^2 + 4C_P^2 \int_0^T \|f\|_0^2 \, dt$$

which give (9.19) and (9.20). □

Remark 9.4. As a by-product of the above proof, we deduce that, if $f=0$, u satisfies the equation

$$\frac{d}{dt} \|u(t)\|_0^2 = -2\alpha \|u(t)\|_1^2 \leq 0$$

which shows the *dissipative nature of the diffusion equation.*

9.2.6 Regularity

As in the elliptic case, the regularity of the solution improves with the regularity of the data. Precisely, we have:

Theorem 9.4. *Let Ω be a C^2-domain and u be the weak solution of problem (9.3). If $g \in V$, then $u \in L^2(0,T;H^2(\Omega)) \cap L^\infty(0,T;V)$ and $\dot{u} \in L^2(0,T;H)$. Moreover*

$$\|u\|_{L^2(0,T;H^2(\Omega))} + \|u\|_{L^\infty(0,T;V)} + \|\dot{u}\|_{L^2(0,T;H)} \leq C(\alpha) \left\{ \|g\|_V + \|f\|_{L^2(0,T;H)} \right\}. \tag{9.27}$$

Proof. Multiplying equation (9.11) by $\dot{c}_k(t)$ and summing for $k = 1, ..., m$, we get

$$\|\dot{u}_m(t)\|_0^2 + \alpha \left(\nabla u_m(t), \nabla \dot{u}_m(t) \right)_0 = \left(f(t), \dot{u}_m(t) \right)_0 \tag{9.28}$$

for a.e. $t \in [0, T]$. Now, note that

$$(\nabla u_m(t), \nabla \dot{u}_m(t))_0 = \frac{1}{2} \frac{d}{dt} \|\nabla u_m(t)\|_0^2, \qquad \text{a.e. } t \in (0, T)$$

and that, from Schwarz's inequality

$$(f(t), u_m(t))_0 \le \|f(t)\|_0 \|u_m(t)\|_0 \le \frac{1}{2} \|f(t)\|_0^2 + \frac{1}{2} \|\dot{u}_m(t)\|_0^2.$$

From this inequality and (9.28), we infer

$$\alpha \frac{d}{dt} \|\nabla u_m(t)\|_0^2 + \|\dot{u}_m(t)\|_0^2 \le \|f(t)\|_0^2 \qquad \text{a.e. } t \in (0, T).$$

An integration over $(0, t)$ yields

$$\alpha \|\nabla u_m(t)\|_0^2 + \int_0^t \|\dot{u}_m(s)\|_0^2 \, ds \le \int_0^t \|f(s)\|_0^2 \, ds + \|g_m\|_1^2. \tag{9.29}$$

Passing to the limit as $m \to \infty$ along an appropriate subsequence, we deduce that the same estimate holds for u and therefore, that $u \in L^\infty(0, T; V)$ and $\dot{u} \in L^2(0, T; H)$. In particular, we may write (9.7) in the form:

$$\alpha(\nabla u(t), \nabla v)_0 = (f(t) - \dot{u}(t), v)_0 \qquad \text{a.e. } t \in [0, T]$$

for all $v \in V$.

Now, the regularity theory for elliptic equations (Theorem 8.13) implies that $u(t) \in H^2(\Omega)$ for a.e. $t \in [0, T]$ and that

$$\|u(t)\|_{H^2(\Omega)}^2 \le C(\alpha, \Omega) \left\{ \|g\|_1^2 + \|f(t)\|_0^2 + \|\dot{u}(t)\|_0^2 \right\}.$$

Integrating and using (9.29), we obtain

$$u \in L^2(0, T; H^2(\Omega))$$

and the estimate (9.27). □

Further regularity requires compatibility conditions on f and g. We limit ourselves to consider the following situation in the case $f = 0$. Suppose we have $u \in C^\infty(\overline{Q}_T)$. Since $u = 0$ on the lateral side, we have

$$u = \partial_t u = \cdots = \partial_t^j u = \cdots = 0, \quad \forall j \ge 0, \text{ on } \partial\Omega \times (0, \infty)$$

which hold, by continuity, also for $t = 0$. On the other hand, the heat equation gives

$$\partial_t u = \alpha \Delta u, \qquad \partial_t^2 u = \alpha \Delta(\partial_t u) = \alpha \Delta^2 u$$

and, in general,

$$\partial_t^j u = \alpha \Delta^j u, \qquad \forall j \ge 0, \text{ in } Q_T.$$

Since $u \in C^{\infty}\left(\overline{Q}_T\right)$, these equations still hold for $t = 0$. As a consequence, we conclude that

$$g = \Delta g = \cdots = \Delta^j g = \cdots = 0 \qquad \forall j \geq 0, \text{on } \partial\Omega. \tag{9.30}$$

Thus, conditions (9.30) are necessary in order to have $u \in C^{\infty}\left(\overline{Q}_T\right)$. It turns out that they are sufficient as well, as stated by the following theorem[8].

Theorem 9.5. *Let u be the weak solution of problem (9.3). If $g \in H^m\left(\Omega\right)$, for every $m \geq 1$, and conditions (9.30) hold, then $u \in C^{\infty}\left(\overline{Q}_T\right)$.*

9.2.7 The Cauchy-Neuman problem

The Faedo-Galerkin method works with the other common boundary conditions, with small adjustments. Let us examine the weak formulation of the diffusion equation,

$$\langle \dot{u}(t), v \rangle_* + a\left(u(t), v\right) = \left(f(t), v\right)_0, \tag{9.31}$$

which must be true for all $v \in V$ and a.e. in $[0, T]$. For the Cauchy-Dirichlet problem, the bilinear form a is

$$a(w, v) = \alpha\left(\nabla w, \nabla v\right)_0$$

which is a multiple of the inner product in $V = H_0^1(\Omega)$. Thus, a is continuous but also $V-$coercive, which is crucial for the method, as in the elliptic case.

However, once the relevant Hilbert triplet (V, H, V^*) has been selected, for parabolic equation it is enough that a be **weakly coercive** i.e that there exists $\alpha > 0$, $\lambda \geq 0$ such that

$$a(v, v) + \lambda \|v\|_H^2 \geq \alpha \|v\|_V^2 \qquad \forall v \in V. \tag{9.32}$$

Indeed, if (9.32) holds, set

$$w(t) = e^{-\lambda t} u(t).$$

Then,

$$\dot{w}(t) = e^{-\lambda t} \dot{u}(t) - \lambda e^{-\lambda t} u(t) = e^{-\lambda t} \dot{u}(t) - \lambda w(t)$$

so that, if u solves (9.31), w solves

$$\langle \dot{w}(t), v \rangle_* + a\left(w(t), v\right) + \lambda\left(w(t), v\right)_H = \left(e^{-\lambda t} f(t), v\right)_H$$

which is an equation of the same type, with the *coercive* bilinear form

$$\widetilde{a}(w, v) = a(w, v) + \lambda\left(w, v\right)_H$$

and forcing term $e^{-\lambda t} f(t)$. In other words, if the bilinear form a is only weakly coercive, by a simple change of variable we may reduce ourselves to an equivalent

---

[8] For the proof, see *Evans,* 1998.

equation, associated with a modified coercive bilinear form (see however Problem 9.4).

For instance, consider the following Cauchy-Neumann problem[9]:

$$\begin{cases} u_t - \alpha \Delta u = f & \text{in } Q_T \\ u(\mathbf{x},0) = g(\mathbf{x}) & \text{in } \Omega \\ \partial_\nu u(\boldsymbol{\sigma},t) = 0 & \text{on } S_T. \end{cases} \tag{9.33}$$

where $f \in L^2(Q_T)$ and $g \in L^2(\Omega)$. For the weak formulation choose $H = L^2(\Omega)$ and $V = H^1(\Omega)$, where, we recall, inner product and norm are given by

$$(u,v)_{1,2} = (\nabla u, \nabla v)_0 + (u,v)_0, \quad \|u\|_{1,2}^2 = \|\nabla u\|_0^2 + \|u\|_0^2.$$

A weak formulation of the Cauchy-Neumann problem may be stated as follows:

Find $u \in L^2(0,T;V)$ such that $\dot{u} \in L^2(0,T;V^*)$ and

1. *for all $v \in V$ and a.e. $t \in [0,T]$,*

$$\langle \dot{u}(t), v \rangle_* + a(u(t), v) = (f(t), v)_0,$$

2. $u(0) = g.$

The bilinear form

$$a(u,v) = \alpha(\nabla u, \nabla v)_0$$

is weakly coercive: any $\lambda > 0$ works.

For simplicity, let $\lambda = \alpha$; then

$$\tilde{a}(w,v) = \alpha\{(\nabla u, \nabla v)_0 + (u,v)_0\}.$$

With the change of variable

$$w(t) = e^{-\alpha t} u(t)$$

we are reduced to the following equivalent formulation:

Find $w \in L^2(0,T;V)$ such that

$$\dot{w} \in L^2(0,T;V^*)$$

and

1. *for all $v \in V$ and a.e. $t \in [0,T]$,*

$$\langle \dot{w}(t), v \rangle_* + \tilde{a}(w(t), v) = \left(e^{-\alpha t} f(t), v\right)_0,$$

2. $w(0) = g.$

---

[9] For nonhomogeneous Neumann conditions, see Problem 9.2.

With small adjustments, the technique used for Dirichlet boundary conditions yields existence and uniqueness of a unique solution w of the above Cauchy-Neumann problem and therefore of the original problem. The stability estimates for w take the form

$$\|w\left(t\right)\|_0^2 + \int_0^t \|w\left(s\right)\|_{1,2}^2 \, ds \le c(\alpha) \left\{ \|g\|_0^2 + \int_0^t e^{-2\alpha s} \|f\left(s\right)\|_0^2 \, ds \right\}$$

and

$$\int_0^t \|\dot{w}\left(s\right)\|_*^2 \, ds \le c(\alpha) \left\{ \|g\|_0^2 + \int_0^t e^{-2\alpha s} \|f\left(s\right)\|_0^2 \, ds \right\}$$

for all $t \in [0,T]$. Going back to u, we obtain the following theorem.

Theorem 9.6. *There exists a unique weak solution u of (9.31) satisfying the initial condition $u\left(0\right) = g$. Moreover*

$$\int_0^T \left\{ \|u\left(s\right)\|_{1,2}^2 + \|\dot{u}\left(s\right)\|_*^2 \right\} ds \le C \left\{ \|g\|_0^2 + \int_0^T \|f\left(s\right)\|_0^2 \, ds \right\} \tag{9.34}$$

where $C = C\left(\alpha,T\right)$.

9.2.8 Cauchy-Robin and mixed problems

Consider the problem

$$\begin{cases} u_t - \alpha \Delta u = f & \text{in } Q_T \\ u\left(\mathbf{x},0\right) = g\left(\mathbf{x}\right) & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u\left(\boldsymbol{\sigma},t\right) + h\left(\boldsymbol{\sigma}\right) u\left(\boldsymbol{\sigma},t\right) = 0 & \text{on } S_T \end{cases}$$

where $h \in L^\infty\left(\partial\Omega\right)$ and $h \ge 0$. For the weak formulation we choose $H = L^2\left(\Omega\right)$ and $V = H^1\left(\Omega\right)$. As in the elliptic case, we consider the bilinear form

$$a\left(u,v\right) = \alpha\left(\nabla u, \nabla v\right)_0 + \int_{\partial\Omega} huv \, d\sigma. \tag{9.35}$$

A weak formulation is the following:

Determine $u \in L^2\left(0,T;V\right)$ such that $\dot{u} \in L^2\left(0,T;V^*\right)$ and

1. for all $v \in V$ and a.e. $t \in [0,T]$,

$$\langle \dot{u}\left(t\right), v \rangle_* + a\left(u\left(t\right), v\right) = \left(f\left(t\right), v\right),$$

2. $u\left(0\right) = u_0$.

The following theorem holds.

Theorem 9.7. *There exists a unique weak solution u of the Cauchy-Robin problem. Moreover, the inequality (9.34) holds for u, with $C = C(\alpha, \Omega, \|h\|_{L^\infty(\partial\Omega)})$.*

Proof. We may argue as in the case of Neumann conditions. The bilinear form (9.35) is continuous and weakly coercive for any $\lambda > 0$, since[10] $h \geq 0$ on $\partial\Omega$. Choosing $\lambda = \alpha$, we have

$$\widetilde{a}(u, v) = \alpha \left\{ (\nabla u, \nabla v)_0 + (u, v)_0 \right\} + \int_{\partial\Omega} huv \, d\sigma \geq \alpha (u, v)_{1,2}.$$

Moreover, thanks to the *trace inequality* (see Theorem 7.11)

$$\|u\|_{L^2(\partial\Omega)} \leq C_* \|u\|_{1,2} \tag{9.36}$$

we may write

$$|\widetilde{a}(u, v)| \leq \alpha \|u\|_{1,2} \|v\|_{1,2} + \|h\|_{L^\infty(\partial\Omega)} \|u\|_{L^2(\partial\Omega)} \|v\|_{L^2(\partial\Omega)}$$
$$\leq (\alpha + C_* \|h\|_{L^\infty(\partial\Omega)}) \|u\|_{1,2} \|v\|_{1,2}$$

whence $\widetilde{a}$ is continuous as well. Setting

$$w(t) = e^{-\alpha t} u(t),$$

we are led to *determine $w \in L^2(0, T; V)$ such that $\dot{w} \in L^2(0, T; V^*)$ and*

1. *for all $v \in V$ and a.e. $t \in [0, T]$,*

$$\langle \dot{w}(t), v \rangle_* + \widetilde{a}(w(t), v) = (e^{-\alpha t} f(t), v),$$

2. $w(0) = u_0.$

Then, the energy inequalities follow as in the case of the Dirichlet problem. For the existence and uniqueness of the weak solution, we need only to observe that the trace inequality (9.36) gives, for the Galerkin approximations $\{u_m\}$, the estimate

$$\int_0^T \|u_m(t)\|_{L^2(\partial\Omega)}^2 \, dt \leq C_*^2 \int_0^T \|u_m(t)\|_{1,2}^2 \, dt \leq \overline{C}(\alpha, T) \left\{ \|g\|_0^2 + \int_0^T \|f(s)\|_0^2 \, ds \right\}.$$

Thus, $\{u_m\}$ has a subsequence weakly convergent in $L^2(\partial\Omega)$. Therefore we can pass to the limit as $m \to +\infty$ in the term

$$\int_{\partial\Omega} hu_m(t) v d\sigma$$

as well. $\square$

---

[10] If $|h| \leq M$ on $\partial\Omega$, a is weakly coercive for λ large enough (check it).

Finally, we consider the **mixed problem**

$$\begin{cases} u_t - \alpha \Delta u = f & \text{in } Q_T \\ u(\mathbf{x},0) = g(\mathbf{x}) & \text{in } \Omega \\ \partial_{\boldsymbol{\nu}} u(\boldsymbol{\sigma},t) = 0 & \text{on } \Gamma_N \times [0,T] \\ u(\boldsymbol{\sigma},t) = 0 & \text{on } \Gamma_D \times [0,T]. \end{cases}$$

where Γ_D is a relatively open subset of $\partial\Omega$ and $\Gamma_N = \partial\Omega\backslash\Gamma_D$. For the weak formulation we choose $H = L^2(\Omega)$ and $V = H^1_{0,\Gamma_D}(\Omega)$, with inner product

$$(u,v)_1 = (\nabla u, \nabla v)_0$$

and norm $\|\cdot\|_1$. Recall that in $H^1_{0,\Gamma_D}(\Omega)$ Poincaré's inequality holds:

$$\|v\|_0 \le C_P \|v\|_1.$$

The bilinear form

$$a(w,v) = \alpha(\nabla w, \nabla v)_0$$

is continuous and $V-$coercive. Reasoning as in the case of the Dirichlet condition, we conclude that:

Theorem 9.8. *There exists a unique weak solution u of the initial-mixed problem. Moreover, the inequalities (9.19) and (9.20) hold.*

9.2.9 A control problem

Using the same techniques of Section 8.8, we may solve simple control problems for the diffusion equation. Consider, for instance, the problem of minimizing the cost functional

$$J(u,z) = \frac{1}{2}\int_\Omega |u(T) - u_d|^2\, d\mathbf{x} + \frac{\beta}{2}\int_{Q_T} z^2 d\mathbf{x}dt$$

under the condition (*state system*)

$$\begin{cases} u_t - \Delta u = z & \text{in } Q_T \\ u = 0 & \text{on } S_T \\ u(\mathbf{x},0) = g(\mathbf{x}) & \text{in } \Omega. \end{cases} \tag{9.37}$$

Thus, we want to control the distributed heat flux in Q_T, given by z, in order to **minimize the distance from u_d of the final observation of u, given by** $u(T)$.

If $g \in L^2(\Omega)$ and the class of admissible controls is $L^2(Q_T)$, we know from Theorem 9.3 that problem (9.37) has a unique weak solution $u = u[z]$ for every control z. Moreover, we assume that Ω is a C^2-domain and that $g \in H^1_0(\Omega)$, so that Theorem 9.4 implies that, actually, $u \in L^2(0,T;H^2(\Omega))$ and $\dot{u} \in L^2(0,T;L^2(\Omega))$.

Substituting $u[z]$ into J, we obtain the functional

$$\tilde{J}(z) = J(u[z], z) = \frac{1}{2} \int_{\Omega} (u[T; z] - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_{Q_T} z^2 d\mathbf{x} dt, \qquad (9.38)$$

where we have set $u[t; z] = u[z](t)$.

Since the mapping $z \mapsto u[z] - u[0]$ is linear (why?), we write

$$\tilde{J}(z) = \frac{1}{2} \int_{\Omega} (u[T; z] - u[T; 0] + u[T; 0] - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_{Q_T} z^2 d\mathbf{x} dt,$$

and then it is easy to check that $\tilde{J}$ has the form

$$\tilde{J}(z) = \frac{1}{2} b(z, z) + Lz + q$$

where

$$b(z, w) = \int_{\Omega} (u[T; z] - u[T; 0])(u[T; w] - u[T; 0]) \, d\mathbf{x} + \beta \int_{Q_T} zw \, d\mathbf{x} dt$$

and

$$Lz = \int_{\Omega} (u[T; z] - u[T; 0])(u[T; 0] - u_d) \, d\mathbf{x}$$

with $q = \frac{1}{2} \int_{\Omega} (u[T; 0] - u_d)^2 \, d\mathbf{x}$.

Following the proof of Theorem 8.15 we deduce that *there exists a unique optimal control* z^*, with corresponding optimal state $u^* = u[z^*]$. Moreover, the optimal control is characterized by the following *Euler equation*:

$$\tilde{J}'(z^*)[w] = b(z^*, w) + Lw = 0$$

which, after some adjustments becomes

$$\tilde{J}'(z^*)[w] = \int_{\Omega} (u^*(T) - u_d)(u[T; w] - u[T; 0]) \, d\mathbf{x} + \beta \int_{Q_T} z^* w \, d\mathbf{x} dt = 0$$

for every $w \in L^2(Q_T)$.

Using the method of Lagrange multiplier as in subsection 8.8.3 we may obtain a more manageable set of optimality conditions. In fact, let us write the cost functional (9.38) in the following augmented form, highlighting the role of the linear map $z \mapsto u[z] - u[0]$:

$$\tilde{J}(z) = \frac{1}{2} \int_{\Omega} (u[T; z] - u_d)^2 \, d\mathbf{x} + \frac{\beta}{2} \int_{Q_T} z^2 d\mathbf{x} dt$$

$$+ \int_{Q_T} p \{z - (\dot{u}[z] - \dot{u}[0]) + \Delta(u[z] - u[0])\} \, d\mathbf{x} dt$$

where p is a multiplier. Note that we have just added zero to $\tilde{J}$.

The Euler equation for the augmented functional becomes:

$$\tilde{J}'\left(z^{*}\right)[w] = \int_{\Omega}\left(u^{*}\left(T\right)-u_{d}\right)\left(u\left[T;w\right]-u\left[T;0\right]\right)\ d\mathbf{x} + \int_{Q_{T}}\left(\beta z^{*}+p\right)w\ d\mathbf{x}dt +$$

$$-\int_{Q_{T}}p\ \left\{\left(\dot{u}\left[w\right]-\dot{u}\left[0\right]\right)-\Delta(u\left[w\right]-u\left[0\right])\right\}d\mathbf{x}dt=0$$

for all $w \in L^{2}\left(Q_{T}\right)$. We now integrate by parts the last integral. We have, since $u\left[0;w\right]-u\left[0;0\right]=g-g=0$:

$$\int_{Q_{T}}p\ (\dot{u}\left[w\right]-\dot{u}\left[0\right])d\mathbf{x}dt = \int_{0}^{T}\int_{\Omega}p\ (\dot{u}\left[w\right]-\dot{u}\left[0\right])d\mathbf{x}dt$$

$$= \int_{\Omega}p\left(T\right)\left(u\left[T;w\right]-u\left[T;0\right]\right)d\mathbf{x} - \int_{Q_{T}}\dot{p}\ (u\left[w\right]-u\left[0\right])d\mathbf{x}dt.$$

Furthermore, since $u\left[w\right]-u\left[0\right]=0$ on S_{T},

$$\int_{Q_{T}}p\ \Delta(u\left[w\right]-u\left[0\right])d\mathbf{x}dt = \int_{S_{T}}p\left(u_{\boldsymbol{\nu}}\left[w\right]-u_{\boldsymbol{\nu}}\left[0\right]\right)d\sigma dt - \int_{Q_{T}}\nabla p\cdot\nabla(u\left[w\right]-u\left[0\right])d\mathbf{x}dt$$

$$= \int_{S_{T}}p\left(u_{\boldsymbol{\nu}}\left[w\right]-u_{\boldsymbol{\nu}}\left[0\right]\right)d\sigma dt + \int_{Q_{T}}\Delta p\ (u\left[w\right]-u\left[0\right])d\mathbf{x}dt.$$

Let the multiplier p be the unique solution of the following *adjoint* problem:

$$\begin{cases} p_{t}+\Delta p=0 & \text{in } Q_{T} \\ p=0 & \text{on } S_{T} \\ p\left(\mathbf{x},T\right)=-\left(u^{*}\left(T\right)-u_{d}\right) & \text{in } \Omega. \end{cases} \tag{9.39}$$

Then the Euler equation reduces to

$$\tilde{J}'\left(z^{*}\right)[w] = \int_{Q_{T}}\left(\beta z^{*}+p\right)w\ d\mathbf{x}dt=0 \qquad \forall w\in L^{2}\left(Q_{T}\right)$$

whence

$$\beta z^{*}+p=0. \tag{9.40}$$

Let us summarize the above results. *The control z^{*} and the state $u^{*}\left[z^{*}\right]$ are optimal if and only if there exist a multiplier $p^{*}\in L^{2}\left(0,T;H^{2}\left(\Omega\right)\right)$, with $\dot{p}^{*}\in L^{2}\left(0,T;L^{2}\left(\Omega\right)\right)$, such that z^{*}, u^{*} and p^{*} satisfy the state system (9.37), the adjoint system (9.39) and the Euler equation (9.40).*

Remark 9.5. Note that the adjoint system is a final value problem for the backward heat equation, which is a well posed problem.

9.3 General Equations

9.3.1 Weak formulation of initial value problems

We now consider divergence form operators[11]

$$\mathcal{E}u = -\text{div}\mathbf{A}\nabla u + \mathbf{c}\cdot\nabla u + a_0 u.$$

The matrix $\mathbf{A} = (a_{i,j}(\mathbf{x},t))$, in general different from a multiple of the identity matrix, encodes the anisotropy of the medium with respect to diffusion. For instance, (see subsection 2.6.2) a matrix of the type

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{pmatrix}$$

with $\alpha \gg \varepsilon > 0$, denotes higher propensity of the medium towards diffusion along the x_1-axis, than along the other directions. As in the stationary case, for the control of the stability of numerical algorithms, it is important to compare the effects of the drift, reaction and diffusion terms. We make the following hypotheses:

(a) the coefficients $\mathbf{c}$, a_0 are bounded (i.e. all belong to $L^\infty(Q_T)$), with

$$|\mathbf{c}| \leq \gamma, \ |a_0| \leq \gamma_0, \qquad \text{a.e. in } Q_T.$$

(b) $\mathcal{E}$ is *uniformly elliptic*:

$$\alpha|\boldsymbol{\xi}|^2 \leq \mathbf{A}(\mathbf{x},t)\boldsymbol{\xi}\cdot\boldsymbol{\xi} \leq K|\boldsymbol{\xi}|^2 \qquad \text{for all } \boldsymbol{\xi} \in \mathbb{R}^n, \ \boldsymbol{\xi} \neq \mathbf{0}, \text{ a.e. } (\mathbf{x},t) \in Q_T.$$

We consider initial value problems of the form:

$$\begin{cases} u_t + \mathcal{E}u = f & \text{in } Q_T \\ u(\mathbf{x},0) = g(\mathbf{x}) & \mathbf{x} \in \Omega \\ \mathcal{B}u(\boldsymbol{\sigma},t) = 0 & (\boldsymbol{\sigma},t) \in S_T \end{cases} \tag{9.41}$$

where $\mathcal{B}u$ stands for one of the usual *homogeneous* boundary conditions. For instance, $\mathcal{B}u = \partial_\nu u$ for the Neumann condition.

The weak formulation of problem (9.41) follows the pattern of the previous sections. Let us briefly review the main ingredients.

Functional setting. The functional setting is constituted by a Hilbert triplet (V, H, V^*), where $H = L^2(\Omega)$ and $H_0^1(\Omega) \subseteq V \subseteq H^1(\Omega)$. The choice of V depends on the type of boundary condition we are dealing with. The familiar choices are $V = H_0^1(\Omega)$ for the homogeneous Dirichlet condition, $V = H^1(\Omega)$ for the Neumann or Robin condition, $V = H_{0,\Gamma_D}^1(\Omega)$ in the case of mixed conditions.

---

[11] For simplicity we consider $\mathbf{b} = \mathbf{0}$, but the extension of the results to the case $\mathbf{b} \neq \mathbf{0}$ is straightforward,

The bilinear form. We set

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

and, in the case of Robin condition,

$$a(u, v; t) = \int_\Omega \{\mathbf{A}\nabla u \cdot \nabla v + (\mathbf{c} \cdot \nabla u) v + a_0 uv\} \, d\mathbf{x} + \int_{\partial\Omega} huv \, d\sigma$$

where we require $h \in L^\infty(\partial\Omega)$, $h \geq 0$ an $\partial\Omega$. Notice that a is time dependent, in general.

Under the stated hypotheses, it is not difficult to show that

$$|a(u, v; t)| \leq M \|u\|_V \|v\|_V$$

so that a is *continuous* in V. The constant M depends on K, γ, γ_0 that is, on the size of the coefficients a_{ij}, c_j, a_0 (and on $\|h\|_{L^\infty(\partial\Omega)}$ in the case of Robin condition).

Also, a is *weakly coercive*. In fact from (9.16), we have, for every $\varepsilon > 0$:

$$\int_\Omega (\mathbf{c} \cdot \nabla u) u \, d\mathbf{x} \geq -\gamma \|\nabla u\|_0 \|u\|_0$$

$$\geq -\frac{\gamma}{2} \left[\varepsilon \|\nabla u\|_0^2 + \frac{1}{\varepsilon} \|u\|_0^2 \right]$$

and

$$\int_\Omega a_0 u^2 d\mathbf{x} \geq -\gamma_0 \|u\|_0^2$$

whence, as $h \geq 0$ a.e. on $\partial\Omega$,

$$a(u, u; t) \geq \left[\alpha - \frac{\gamma_0 \varepsilon}{2}\right] \|\nabla u\|_0^2 - \left[\frac{\gamma}{2\varepsilon} + \gamma_0\right] \|u\|_0^2. \tag{9.42}$$

We distinguish three cases:

If $\gamma = 0$ and $\gamma_0 = 0$ the bilinear form is $V-$coercive when $V = H_0^1(\Omega)$. If $H_0^1(\Omega) \subseteq V \subseteq H^1(\Omega)$,

$$\tilde{a}(u, v; t) = a(u, v; t) + \lambda_0 (u, v)_0 \tag{9.43}$$

is $V-$coercive for any $\lambda_0 > 0$.

If $\gamma = 0$ and $\gamma_0 > 0$, (9.43) is $V-$coercive for any $\lambda_0 > \gamma$.

If $\gamma > 0$, choose in (9.42)

$$\varepsilon = \frac{\alpha}{\gamma} \quad \text{and} \quad \lambda_0 = 2\left[\frac{\gamma}{2\varepsilon} + \gamma_0\right] = 2\left[\frac{\gamma^2}{2\alpha} + \gamma_0\right].$$

Then

$$\tilde{a}(u, v; t) \geq \frac{\alpha}{2} \|\nabla u\|_0^2 + \frac{\lambda_0}{2} \|u\|_0^2 \geq \min\left\{\frac{\alpha}{2}, \frac{\lambda_0}{2}\right\} \|u\|_{1,2}^2$$

so that a is *weakly coercive*.

The data g and f. We assume $g \in H$ and $f \in L^2(0, T; V^*)$

The solution. We look for u such that $u(t) \in V$, at least a.e. $t \in [0, T]$. Since

$$v \longmapsto a(u, v; t)$$

and $f(t)$ are elements of V^*, a.e. $t \in [0, T]$, we ask $\dot{u}(t) \in V^*$ a.e. in $[0, T]$ as well. Moreover we require that $\|u(t)\|_V$ and $\|\dot{u}(t)\|_{V^*}$ belongs to $L^2(0, T)$.

The weak formulation. The above considerations lead to the following weak formulation of the initial-boundary value problem:

Given $f \in L^2(0, T; V^*)$ and $g \in L^2(\Omega)$, determine $u \in L^2(0, T; V)$ such that $\dot{u} \in L^2(0, T; V^*)$ and that:

1. for all $v \in V$ and a.e. $t \in [0, T]$,

$$\langle \dot{u}(t), v \rangle_* + a(u(t), v; t) = \langle f(t), v \rangle_*, \tag{9.44}$$

2. $u(0) = g$.

Again, since $u \in C([0, T]; H)$, condition 2 means $\|u(t) - g\|_0 \to 0$ when $t \to 0^+$. As for the heat equation, (9.44) may be written in the equivalent form

$$\frac{d}{dt}(u(t), v) + a(u(t), v; t) = \langle f(t), v \rangle_*$$

for all $v \in V$ and in the sense of *distributions in* $\mathcal{D}'[0, T]$.

If a is not coercive, we make the change of variable

$$w(t) = e^{-\lambda_0 t} u(t)$$

and (9.44) becomes

$$\langle \dot{u}(t), v \rangle_* + \tilde{a}(u(t), v; t) = \langle e^{-\lambda_0 t} \tilde{f}(t), v \rangle_*$$

with the coercive bilinear form

$$\tilde{a}(u, v; t) = a(u, u; t) + \lambda_0 (u, v)_0.$$

Every stability estimate for w translates into a corresponding estimate for u, times the factor $e^{\lambda_0 t}$.

9.3.2 Faedo-Galerkin method (II)

We want to show that our initial value problem has a unique weak solution, which continuously depends on the data, in the appropriate norms. The method of Faedo-Galerkin may be used also in this case, as in the previous sections, with small corrections only.

Choose an orthonormal basis $\{w_k\}$ in H, orthogonal in V, and let

$$V_m = \operatorname{span}\{w_1, w_2, ..., w_m\}.$$

Look at the projected equation

$$\langle \dot{u}_m, v \rangle_* + a\,(u_m, v; t) = \langle f, v \rangle_* \qquad \forall v \in V_m \tag{9.45}$$

where $u_m = u_m\,(t) = \sum_{k=1}^m c_k\,(t)\,w_k$.

Galerkin approximations. Inserting $v = w_s$, $s = 1, ..., m$, into (9.45) we are led to the following linear system of ordinary differential equations:

$$\begin{cases} \mathbf{C}'_m\,(t) = -\mathbf{W}\,(t)\,\mathbf{C}_m\,(t) + \mathbf{F}\,(t)\,, & \text{a.e. } t \in [0, T]\,, \\ \mathbf{C}_m\,(0) = \mathbf{g}_m. \end{cases} \tag{9.46}$$

where $\mathbf{C}_m\,(t) = (c_1\,(t)\,, ..., c_m\,(t))$, the entries of the matrix $\mathbf{W}$ are

$$W_{sk}\,(t) = a\,(w_k, w_s, t)$$

and

$$F_s\,(t) = \langle f\,(t)\,, w_s \rangle_*\,, \qquad \mathbf{F}_m\,(t) = (F_1\,(t)\,, ..., F_m\,(t))\,,$$

$$g_s = (g, w_s)\,, \qquad \mathbf{g}_m = (g_1, ..., g_m)\,.$$

Since $\mathbf{F} \in L^2\,(0, T; \mathbb{R}^m)$ and $W_{sk} \in L^\infty\,(0, T)$, for every $m \geq 1$ there exists a unique solution $u_m \in H^1\,(0, T; V_m)$ of Problem (9.46).

The *energy estimates* for u_m and their proofs, necessary to pass to the limit in (9.45), are perfectly analogous to those indicated in Theorems 9.1 and 9.2.

If a is not coercive, we make the change of variable

$$w\,(t) = e^{-\lambda_0 t} u\,(t)$$

and (9.44) becomes

$$\langle \dot{u}\,(t)\,, v \rangle_* + \tilde{a}\,[(u\,(t)\,, v; t)] = \langle e^{-\lambda_0 t} \tilde{f}\,(t)\,, v \rangle_*$$

with the coercive bilinear form

$$\tilde{a}\,(u, v; t) = a\,(u, u; t) + \lambda_0\,(u, v)_0\,.$$

Every stability estimate for w translates into a corresponding estimate for u, times the factor $e^{\lambda_0 t}$. Precisely, we have:

Estimates of u_m and $\dot{u}_m$. *Let u_m be the solution of problem (9.46). Then*

$$\max_{t \in [0, T]} \|u_m\,(t)\|_0^2 + \alpha \int_0^T \|u_m\|_V^2\, dt \leq C \left\{ \int_0^T \|f\|_*^2\, dt + \|g\|_0^2 \right\} \tag{9.47}$$

and

$$\int_0^T \|\dot{u}_m\|_*^2\, dt \leq C \left\{ \int_0^T \|f\|_*^2\, dt + \|g\|_0^2 \right\} \tag{9.48}$$

where C depends only on $\Omega, \alpha, K, \beta, \gamma, T$.

Existence and uniqueness. From (9.47), the sequence $\{u_m\}$ of Galerkin approximations is bounded in $L^2(0,T;V)$, while, from (9.48), $\{\dot{u}_m\}$ is bounded in $L^2(0,T;V^*)$. Thus, there exists a subsequence of $\{u_m\}$, which we still denote by $\{u_m\}$, such that, for $m \to \infty$,

$$u_m \rightharpoonup u \quad weakly \ in \ L^2(0,T;V)$$

and

$$\dot{u}_m \rightharpoonup \dot{u} \quad weakly \ in \ L^2(0,T;V^*).$$

Then:

Theorem 9.9. *If* $f \in L^2\left(0,T;L^2(\Omega)\right)$ *and* $g \in L^2(\Omega)$, u *is the unique weak solution of problem (9.41). Moreover*

$$\max_{t \in [0,T]} \|u(t)\|_0^2 + \alpha \int_0^T \|u\|_V^2 \, dt \le C \left\{ \int_0^T \|f\|_*^2 \, dt + \|g\|_0^2 \right\}$$

$$\int_0^T \|u(t)\|_*^2 \, dt \le C \left\{ \int_0^T \|f\|_*^2 \, dt + \|g\|_0^2 \right\}$$

where C *depends only on* Ω, $\alpha, K, \beta, \gamma, T$.
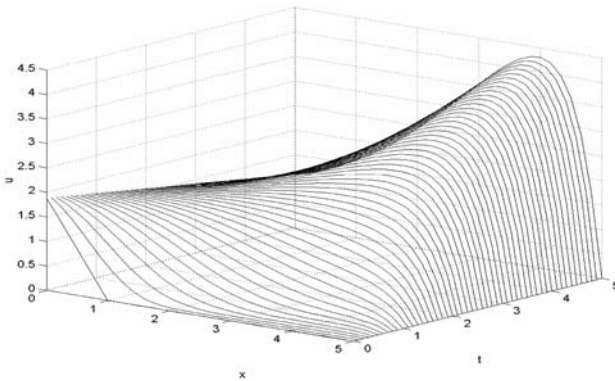
Fig. 9.2. The solution of problem (9.49) in Example 9.1

Remark 9.6. The method works with non homogeneous boundary conditions as well. For instance, for the initial-Dirichlet problem, if the data is the trace of a function $\varphi \in L^2\left(0,T;H^1(\Omega)\right)$ with $\dot{\varphi} \in L^2\left(0,T;L^2(\Omega)\right)$, the change of variable $w = u - \varphi$ reduces the problem to homogeneous boundary conditions.

Example 9.1. Figure 9.2 shows the graph of the solution of the Cauchy-Dirichlet problem

$$\begin{cases} u_t - u_{xx} + 2u_x = 0.2tx & 0 < x < 5, t > 0 \\ u(x,0) = \max(2 - 2x, 0) & 0 < x < 5 \\ u(0,t) = 2 - t/6, u(5,t) = 0 & t > 0 \end{cases} \tag{9.49}$$

Note the tendency of the drift term $2u_x$, to "transport to the right" initial data and the effect of the source term $0.2tx$ to increase the solution near $x = 5$, more and more with time.

9.4 The Wave Equation

9.4.1 Hyperbolic Equations

The wave propagation in a nonhomogeneous and anisotropic medium leads to second order *hyperbolic* equations. With the same notations of section 9.1, an equation in *divergence form* of the type

$$u_{tt} - \operatorname{div}(\mathbf{A}(\mathbf{x},t) \nabla u) + \mathbf{b}(\mathbf{x},t) \cdot \nabla u + c(\mathbf{x},t) u = f(\mathbf{x},t) \tag{9.50}$$

or in *non-divergence form* of the type

$$u_{tt} - \operatorname{tr}(\mathbf{A}(\mathbf{x},t) D^2 u) + \mathbf{b}(\mathbf{x},t) \cdot \nabla u + c(\mathbf{x},t) u = f(\mathbf{x},t) \tag{9.51}$$

is called **hyperbolic** in $Q_T = \Omega \times (0, T)$ if

$$\mathbf{A}(\mathbf{x},t) \boldsymbol{\xi} \cdot \boldsymbol{\xi} > 0 \qquad \text{a.e. } (\mathbf{x},t) \in Q_T, \forall \boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\xi} \neq \mathbf{0}.$$

The typical problems for hyperbolic equations are those already considered for the wave equation. Given f in Q_T, we want to determine a solution u of (9.50) or (9.51) satisfying the *initial* conditions

$$u(\mathbf{x},0) = g(\mathbf{x}), \quad u_t(\mathbf{x},0) = h(\mathbf{x}) \qquad \text{in } \Omega$$

and one of the usual boundary conditions (*Dirichlet, Neumann, mixed or Robin*) on the lateral boundary $S_T = \partial \Omega \times [0, T]$.

Even if from the phenomenological point of view, the hyperbolic equations display substantial differences from the parabolic ones, for *divergence form* equations it is possible to give a similar weak formulation,which can be analyzed by means of Faedo-Galerkin method. We will limit ourselves to the Cauchy-Dirichlet problem for the wave equation. For general equations, the theory is more complicated, unless we assume that the coefficients a_{jk}, entries of the matrix $\mathbf{A}$, are continuously differentiable with respect to both $\mathbf{x}$ and t.

9.4.2 The Cauchy-Dirichlet problem

Consider the problem

$$
\begin{cases}
u_{tt} - c^2 \Delta u = f & \text{in } Q_T \\
u(\mathbf{x},0) = g(\mathbf{x}), \ u_t(\mathbf{x},0) = g^1(\mathbf{x}) & \mathbf{x} \in \Omega \\
u(\boldsymbol{\sigma},t) = 0 & (\boldsymbol{\sigma},t) \in S_T.
\end{cases}
\tag{9.52}
$$

To find an appropriate weak formulation, multiply the wave equation by a function $v = v(\mathbf{x})$, vanishing at the boundary, and integrate over Ω. We find

$$
\int_\Omega u_{tt}(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x} - c^2 \int_\Omega \Delta u(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x} = \int_\Omega f(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x}.
$$

Integrating by parts the second term, we get

$$
\int_\Omega u_{tt}(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x} + c^2 \int_\Omega \nabla u(\mathbf{x},t) \cdot \nabla v(\mathbf{x})\, d\mathbf{x} = \int_\Omega f(\mathbf{x},t)\, v(\mathbf{x})\, d\mathbf{x}
\tag{9.53}
$$

which becomes, in the notations of the previous sections,

$$
\int_\Omega \ddot{u}(t)\, v\ d\mathbf{x} + \alpha \int_\Omega \nabla u(t) \cdot \nabla v\ d\mathbf{x} = \int_\Omega f(t)\, v\ d\mathbf{x}
$$

where $\ddot{u}$ stays for u_{tt}. Again the natural space for u is $L^2\left(0,T; H_0^1(\Omega)\right)$. Thus, a.e. $t > 0$, $u(t) \in V = H_0^1(\Omega)$, and $\Delta u(t) \in V^* = H^{-1}(\Omega)$. On the other hand, from the wave equation we have

$$
u_{tt} = c^2 \Delta u + f.
$$

If $f \in L^2(0,T; H)$, with $H = L^2(\Omega)$, it is natural to require $\ddot{u} \in L^2(0,T; V^*)$.

Accordingly, a reasonable assumption for $\dot{u}$ is $\dot{u} \in L^2(0,T; H)$, an intermediate space between $L^2(0,T; V)$ and $L^2(0,T; V^*)$. Thus, we look for solutions u such that

$$
u \in L^2(0,T; V), \quad \dot{u} \in L^2(0,T; H), \quad \ddot{u} \in L^2(0,T; V^*).
\tag{9.54}
$$

It can be shown[12] that, if u satisfies (9.54), then,

$$
u \in C([0,T]; V) \quad \text{and} \quad \dot{u} \in C([0,T]; H).
$$

Thus, it is reasonable to assume $u(0) = g \in V$, $\dot{u}(0) = g^1 \in H$.

The above considerations lead to the following weak formulation.

Given $f \in L^2(0,T; V^*)$ and $g \in V$, $g^1 \in H$, determine $u \in L^2(0,T; V)$ such that

$$
\dot{u} \in L^2(0,T; H), \quad \ddot{u} \in L^2(0,T; V^*)
$$

and that:

---

[12] *Lions-Magenes*, Chapter 3, 1972.

1. *for all $v \in V$ and a.e. $t \in [0, T]$,*

$$\langle \ddot{u}(t), v \rangle_* + c^2 (\nabla u(t), \nabla v)_0 = (f(t), v)_0, \tag{9.55}$$

2. $u(0) = g, \dot{u}(0) = g^1$

Remark 9.7. Equation (9.55) may be interpreted in the sense of distributions in $\mathcal{D}'(0, T)$. First observe that, for every $v \in V$, the real function

$$w(t) = \langle \ddot{u}(t), v \rangle_*$$

is a distribution $\mathcal{D}'(0, T)$ and

$$w(t) = \frac{d^2}{dt^2} (u(t), v)_0 \qquad \text{in } \mathcal{D}'(0, T). \tag{9.56}$$

This means that for every $\varphi \in \mathcal{D}(0, T)$ we have

$$\int_0^T w(t) \varphi(t) \, dt = \int_0^T (u(t), v) \ddot{\varphi}(t) \, dt.$$

In fact, since $u(t) \in V^*$, we may write, thanks to Bochner's Theorem,

$$\int_0^T w(t) \varphi(t) \, dt = \int_0^T \langle \ddot{u}(t), v \rangle_* \varphi(t) \, dt = \left\langle \int_0^T \ddot{u}(t) \varphi(t) \, dt, v \right\rangle_*$$

$$= \left(\int_0^T u(t) \ddot{\varphi}(t) \, dt, v \right)_0 = \int_0^T (u(t), v))_0 \ddot{\varphi}(t) \, dt.$$

for all $\varphi \in \mathcal{D}(0, T)$. Since the last integral is well defined, $w \in L^1_{loc}(0, T)$ and therefore $w \in \mathcal{D}'(0, T)$. Moreover, by definition,

$$\int_0^T (u(t), v) \ddot{\varphi}(t) \, dt = \int_0^T \frac{d^2}{dt^2} (u(t), v) \ \varphi(t) \, dt$$

in $\mathcal{D}'(0, T)$, which is (9.56). As a consequence, (9.55) may be written in the form

$$\frac{d^2}{dt^2} (u(t), v)_0 + c^2 (\nabla w(t), \nabla v)_0 = (f(t), v)_0 \tag{9.57}$$

for all $v \in V$ and in the sense of distributions in $[0, T]$.

Remark 9.8. We leave it to the reader to check that if a *weak* solution u is *smooth*, i.e. $u \in C^2(\overline{Q_T})$, then u is a classical solution.

9.4.3 Faedo-Galerkin method (III)

We want to show that problem (9.52) has a unique weak solution, which continuously depends on the data, in appropriate norms. Once more, we are going to use the method of Faedo-Galerkin, so that we briefly review the main steps, emphasizing the differences with the parabolic case.

1. We select a sequence of smooth functions $\{w_k\}_{k=1}^\infty$ constituting

$$\text{an } orthogonal \text{ basis in } V$$

and

$$\text{an } orthonormal \text{ basis in } H.$$

In particular, we can write

$$g = \sum_{k=1}^\infty g_k w_k, \qquad g^1 = \sum_{k=1}^\infty g_k^1 w_k$$

where $g_k = (g, w_k)_0$, $g_k^1 = (g^1, w_k)_0$, with the series converging in H.

2. Let

$$V_m = \text{span}\,\{w_1, w_2, ..., w_m\}$$

and

$$u_m(t) = \sum_{k=1}^m r_k(t)\, w_k, \qquad G_m = \sum_{k=1}^m g_k w_k, \qquad G_m^1 = \sum_{k=1}^m g_k^1 w_k.$$

We construct the sequence of Galerkin approximations u_m by solving the following *projected* problem:

Determine $u_m \in H^2(0, T; V)$ such that, for all $s = 1, ..., m$,

$$\begin{cases} (\ddot{u}_m(t), w_s)_0 + c^2 (\nabla u_m(t), \nabla w_s)_0 = (f(t), w_s)_0, & 0 \le t \le T \\ u_m(0) = G_m, \quad \dot{u}_m(0) = G_m^1. \end{cases} \tag{9.58}$$

Note that the differential equation in (9.58) is true for each element of the basis w_s, $s = 1, ..., m$, if and only if it is true for every $v \in V_m$. Moreover, since $u_m \in H^2(0, T; V)$ we have $\ddot{u}_m \in L^2(0, T; V)$, so that

$$(\ddot{u}_m(t), v)_0 = \langle \ddot{u}_m(t), v \rangle_*.$$

3. We show that $\{u_m\}$, $\{\dot{u}_m\}$ and $\{\ddot{u}_m\}$ are bounded in $L^2(0, T; V)$, $L^2(0, T; H)$ and $L^2(0, T; V^*)$, respectively (*energy estimates*). Then, the weak compactness Theorem 6.11 implies that a subsequence $\{u_{m_k}\}$ converges weakly in $L^2(0, T; V)$ to u, while $\{\dot{u}_{m_k}\}$ and $\{\ddot{u}_{m_k}\}$ converge weakly in $L^2(0, T; H)$ and $L^2(0, T; V^*)$ to $\dot{u}$ and $\ddot{u}$.

4. We prove that u in step **3** is the unique weak solution of problem (9.52).

9.4.4 Solution of the approximate problem

The following lemma holds.

Lemma 9.2. *For all $m \geq 1$, there exists a unique solution to problem (9.58). In particular, since $u_m \in H^2 (0, T; V)$, we have $u_m \in C^1 ([0, T]; V)$.*

   Proof. Observe that, since w_1, w_2, ...,w_m are orthonormal in H,

$$(\ddot{u}_m (t) , w_s)_0 = \left(\sum_{k=1}^{m} \ddot{r}_k (t) \, w_k, w_s \right)_0 = \ddot{r}_k (t)$$

and since they are orthogonal in V,

$$c^2 \left(\sum_{k=1}^{m} r_k (t) \, \nabla w_k, \nabla w_s \right)_0 = c^2 (\nabla w_s, \nabla w_s)_0 \, r_s (t) = c^2 \, \|\nabla w_s\|_0^2 \, r_s (t) \, .$$

Set

$$F_s (t) = (f (t) , w_s) , \qquad \mathbf{F} (t) = (F_1 (t) , ..., F_m (t))$$

and

$$\mathbf{R}_m (t) = (r_1 (t) , ..., r_m (t)) , \quad \mathbf{g}_m = (g_1, ..., g_m) , \quad \mathbf{g}_m^1 = \left(g_1^1, ..., g_m^1 \right) .$$

If we introduce the diagonal matrix

$$\mathbf{W} = \text{diag} \left\{ \|\nabla w_1\|_0^2 , \|\nabla w_2\|_0^2 , ..., \|\nabla w_m\|_0^2 \right\}$$

of order m, problem (9.58) is equivalent to the following system of m *uncoupled* linear ordinary differential equations, with constant coefficients:

$$\ddot{\mathbf{R}}_m (t) = -c^2 \mathbf{W} \mathbf{R}_m (t) + \mathbf{F}_m (t) , \qquad \text{a.e. } t \in [0, T] \tag{9.59}$$

with initial conditions

$$\mathbf{R}_m (0) = \mathbf{g}_m. \quad \dot{\mathbf{R}}_m (0) = \mathbf{g}_m^1.$$

Since $F_s \in L^2 (0, T)$, for all $s = 1, ..., m$, system (9.59) has a unique solution $\mathbf{R}_m (t) \in H^2(0, T; \mathbb{R}^m)$. From

$$u_m (t) = \sum_{k=1}^{m} r_k (t) \, w_k,$$

we deduce $u_m \in H^2 (0, T; V)$. $\square$

9.4.5 Energy estimates

We want to show that from the sequence of Galerkin approximations $\{u_m\}$ it is possible to extract a subsequence converging to the weak solution of the original problem. As in the parabolic case, we are going to prove that the relevant Sobolev norms of u_m can be controlled by the norms of the data, **in a way that does not depend on** m. Moreover, the estimates must be powerful enough in order to pass to the limit as $m \to +\infty$ in the approximating equation.

$$(\ddot{u}_m(t), v)_0 + c^2 (\nabla u_m(t), \nabla v)_0 = (f(t), v)_0.$$

In this case we can give a bound of the norms of u_m in $L^\infty(0, T; V)$, of $\dot{u}_m$ in $L^\infty(0, T; H)$ and of $\ddot{u}$ in $L^2(0, T; V^*)$, that is the norms

$$\max_{t \in [0,T]} \|u_m\|_1, \quad \max_{t \in [0,T]} \|\dot{u}_m\|_0 \quad \text{and} \quad \int_0^T \|\ddot{u}_m(t)\|_*^2 \, dt.$$

For the proof, we shall use the following elementary but very useful lemma.

Lemma 9.3. (Gronwall). *Let* Ψ, G *be continuous in* $[0, T]$, *with* G *nondecreasing and* $\gamma > 0$. *If*

$$\Psi(t) \le G(t) + \gamma \int_0^t \Psi(s) \, ds, \qquad \text{for all } t \in [0, T]$$

then

$$\Psi(t) \le G(t) e^{\gamma t}, \qquad \text{for all } t \in [0, T].$$

Proof. Let

$$R(s) = \gamma \int_0^s \Psi(r) \, dr.$$

Then, for all $t \in [0, T]$,

$$R'(s) = \gamma \Psi(s) \le \gamma \left[G(s) + \gamma \int_0^s \Psi(r) \, dr \right] = \gamma [G(s) + R(s)].$$

Multiplying both sides by $\exp(-\gamma t)$, we can write the above inequality in the form

$$\frac{d}{ds} [R(s) \exp(-\gamma t)] \le \gamma G(s) \exp(-\gamma t).$$

Integrating over $(0, t)$ gives ($R(0) = 0$):

$$R(t) \le \gamma \int_0^t G(s) e^{\gamma(t-s)} ds \le G(t) e^{\gamma t}, \qquad \text{for all } t \in [0, T].$$

$\square$

Theorem 9.10. (Estimate of $u_m, \dot{u}_m$). *Let u_m be the solution of problem (9.58). Then*

$$\max_{t \in [0,T]} \left\{ \|\dot{u}_m(t)\|_0^2 + 2c^2 \|u_m(t)\|_1^2 \right\} \leq e^T \left\{ \|g\|_1^2 + \|g^1\|_0^2 + \|f\|_{L^2(0,T;H)}^2 \right\}. \quad (9.60)$$

Proof. Since $u_m \in H^2(0,T;V)$, we may choose $v = \dot{u}_m(t)$ as a test function in (9.58). We find

$$(\ddot{u}_m(t), \dot{u}_m(t))_0 + c^2 (\nabla u_m(t), \nabla \dot{u}_m(t))_0 = (f(t), \dot{u}_m(t))_0 \quad (9.61)$$

for a.e. $t \in [0,T]$. Observe that

$$(\ddot{u}_m(t), \dot{u}_m(t))_0 = \frac{1}{2} \frac{d}{dt} \|\dot{u}_m(t)\|_0^2, \qquad \text{a.e. } t \in (0,T)$$

and

$$(\nabla u_m(t), \nabla \dot{u}_m(t))_0 = c^2 \frac{d}{dt} \|\nabla u_m(t)\|_0^2.$$

By Schwarz's inequality,

$$(f(t), \dot{u}_m(t))_0 \leq \|f(t)\|_0 \|\dot{u}_m(t)\|_0 \leq \frac{1}{2} \|f(t)\|_0^2 + \frac{1}{2} \|\dot{u}_m(t)\|_0^2$$

so that, from (9.61) we deduce

$$\frac{d}{dt} \left\{ \|\dot{u}_m(t)\|_0^2 + 2c^2 \|u_m(t)\|_1^2 \right\} \leq \|f(t)\|_0^2 + \|\dot{u}_m(t)\|_0^2.$$

Integrating over $(0,t)$ we get (Remark 7.34 applied to $\dot{u}_m$ and ∇u_m)

$$\|\dot{u}_m(t)\|_0^2 + 2c^2 \|u_m(t)\|_1^2$$
$$\leq \|G_m\|_1^2 + \|G_m^1\|_0^2 + \int_0^t \|f(s)\|_0^2 \, ds + \int_0^t \|\dot{u}_m(s)\|_0^2 \, ds$$
$$\leq \|g\|_1^2 + \|g^1\|_0^2 + \int_0^t \|f(s)\|_0^2 \, ds + \int_0^t \|\dot{u}_m(s)\|_0^2 \, ds,$$

since

$$\|G_m\|_1^2 \leq \|g\|_1^2, \quad \|G_m^1\|_0^2 \leq \|g^1\|_0^2.$$

Let

$$\Psi(t) = \|\dot{u}_m(t)\|_0^2 + 2c^2 \|u_m(t)\|_1^2, \quad G(t) = \|g\|_1^2 + \|g^1\|_0^2 + \int_0^t \|f(s)\|_0^2 \, ds.$$

Note that both Ψ and G are continuous in $[0,T]$. Then

$$\Psi(t) \leq G(t) + \int_0^t \Psi(s) \, ds$$

and Gronwall Lemma yields, for every $t \in [0, T]$,

$$\|\dot{u}_m(t)\|_0^2 + 2c^2 \|u_m(t)\|_1^2 \le e^t \left\{ \|g\|_1^2 + \|h\|_0^2 + \int_0^t \|f\|_0^2 \, ds \right\}$$

□

We now give a control of the norm of $\ddot{u}_m$ in $L^2(0, T; V^*)$.

Theorem 9.11. (Estimate of $\ddot{u}_m$). *Let u_m be the solution of problem (9.58). Then*

$$\int_0^T \|\ddot{u}_m(t)\|_*^2 \, dt \le C(c, T) \left\{ \|g\|_1^2 + \|g^1\|_0^2 + \int_0^T \|f(s)\|_0^2 \, ds. \right\} \tag{9.62}$$

Proof. Let $v \in V$ and write

$$v = w + z$$

with $w \in V_m = \text{span}\{w_1, w_2, ..., w_m\}$ and $z \in V_m^\perp$. Since $w_1, ..., w_k$ are orthogonal in V, we have

$$\|w\|_1 \le \|v\|_1 .$$

Choosing w as a test function in problem (9.58), we obtain

$$(\ddot{u}_m(t), v)_0 = (\ddot{u}_m(t), w)_0 = -c^2 (\nabla u_m(t), \nabla w)_0 + (f(t), w)_0 .$$

Since

$$|(\nabla u_m(t), \nabla w)_0| \le \|u_m(t)\|_1 \|w\|_1$$

we may write

$$\begin{aligned} |(\ddot{u}_m(t), v)_0| &\le \{c^2 \|u_m(t)\|_1 + \|f(t)\|_0\} \|w\|_1 \\ &\le \{c^2 \|u_m(t)\|_1 + \|f(t)\|_0\} \|v\|_1 . \end{aligned}$$

Thus, by the definition of norm in V^*, we infer

$$\|\ddot{u}_m(t)\|_* \le c^2 \|u_m(t)\|_1 + \|f(t)\|_0 .$$

Squaring and integrating over $(0, T)$ we obtain

$$\int_0^T \|\ddot{u}_m(t)\|_*^2 \, dt \le 2c^4 \int_0^T \|u_m(t)\|_1^2 \, dt + 2 \int_0^T \|f(t)\|_0^2 \, dt$$

and Theorem 9.10 gives (9.62). □

9.4.6 Existence, uniqueness and stability

Theorem 9.10 shows that the sequence $\{u_m\}$ of Galerkin approximations is bounded in $L^\infty(0, T; V)$, hence, in particular, in $L^2(0, T; V)$, while the sequence $\{\ddot{u}_m\}$ is bounded in $L^2(0, T; V^*)$.

Theorem 6.11 implies that there exists a subsequence, which for simplicity we still denote by $\{u_m\}$, such that, as $m \to \infty$,

$$u_m \rightharpoonup u \quad weakly\ in\ L^2(0, T; V)$$
$$\dot{u}_m \rightharpoonup \dot{u} \quad weakly\ in\ L^2(0, T; H)$$
$$\ddot{u}_m \rightharpoonup \ddot{u} \quad weakly\ in\ L^2(0, T; V^*).$$

The following theorem holds:

Theorem 9.12. *Let* $f \in L^2(0, T; H)$, $g \in V$, $g^1 \in H$. *Then* u *is the unique weak solution of problem (9.52). Moreover,*

$$\|u\|_{L^\infty(0,T;V)}^2 + \|\dot{u}\|_{L^\infty(0,T;H)}^2 + \|\ddot{u}\|_{L^2(0,T;V^*)}^2 \leq C \left\{ \|f\|_{L^2(0,T;H)}^2 + \|g\|_1^2 + \|g^1\|_0^2 \right\}$$

with $C = C(c, T)$.

Proof. **Existence**. We know that:

$$\int_0^T (\nabla u_m(t), \nabla v(t))_0 \ dt \to \int_0^T (\nabla u(t), \nabla v(t))_0 \ dt$$

for all $v \in L^2(0, T; V)$,

$$\int_0^T (\dot{u}_m(t), w(t))_0 \ dt \to \int_0^T (\dot{u}(t), w(t))_0 \ dt$$

for all $w \in L^2(0, T; H)$, and

$$\int_0^T (\ddot{u}_m(t), v(t))_0 = \int_0^T \langle \ddot{u}_m(t), v(t) \rangle_* \ dt \to \int_0^T \langle \ddot{u}(t), v(t) \rangle_* \ dt$$

for all $v \in L^2(0, T; V)$,

We want to use these properties to pass to the limit as $m \to +\infty$ in problem (9.58), keeping in mind that the test functions have to be chosen in V_m. Fix $v \in L^2(0, T; V)$; we may write

$$v(t) = \sum_{k=1}^\infty b_k(t) w_k$$

where the series converges in V for a.e. $t \in [0, T]$. Let

$$v_N(t) = \sum_{k=1}^N b_k(t) w_k \tag{9.63}$$

and keep N fixed, for the time being. If $m \geq N$, then $v_N \in L^2(0, T; V_m)$. Multiplying equation (9.58) by $b_k(t)$ and summing for $k = 1, ..., N$, we get

$$(\ddot{u}_m(t), v_N(t))_0 + c^2 (\nabla u_m, \nabla v_N)_0 = (f(t), v_N(t))_0.$$

An integration over $(0, T)$ yields

$$\int_0^T \{(\ddot{u}_m, v_N)_0 + c^2 (\nabla u_m, \nabla v_N)_0\} \, dt = \int_0^T (f, v_N)_0 \, dt. \qquad (9.64)$$

Thanks to the weak convergence of u_m and $\ddot{u}_m$ in their respective spaces, we can let $m \to +\infty$. Since

$$(\ddot{u}_m(t), v_N(t))_0 = \langle \ddot{u}_m(t), v_N(t) \rangle_* \to \langle \ddot{u}(t), v_N(t) \rangle_*,$$

we obtain

$$\int_0^T \{\langle \ddot{u}, v_N \rangle_* + c^2 (\nabla u, \nabla v_N)_0\} \, dt = \int_0^T (f, v_N)_0 \, dt.$$

Now, let $N \to \infty$, observing that $v_N \to v$ in $L^2(0, T; V)$ and, in particular, weakly in this space as well. We obtain

$$\int_0^T \{\langle \ddot{u}(t), v(t) \rangle_* + c^2 (\nabla u(t), \nabla v(t))_0\} \, dt = \int_0^T (f(t), v(t))_0 \, dt. \qquad (9.65)$$

Then, (9.65) is valid for all $v \in L^2(0, T; V)$. This entails, in particular (see footnote 7),

$$\langle \ddot{u}(t), v \rangle_* + c^2 (\nabla u(t), \nabla v)_0 \, dt = (f(t), v)_0$$

for all $v \in V$ and a.e. $t \in [0, T]$. Therefore u satisfies (9.55) and we know that $u \in C([0, T]; V)$, $\dot{u} \in C([0, T]; H)$.

To check the initial conditions, we proceed as in Theorem 9.3. We choose any function $v \in C^2([0, T]; V)$, with $v(T) = \dot{v}(T) = 0$. Integrating by parts twice in (9.65), we find

$$\int_0^T \{\langle u(t), \ddot{v}(t) \rangle_* + c^2 (\nabla u(t), \nabla v(t))_0\} \, dt \qquad (9.66)$$

$$= \int_0^T (f(t), v(t))_0 \, dt + (\dot{u}(0), \dot{v}(0)) - (u(0), v(0)).$$

On the other hand, integrating by parts twice in (9.64), and letting first $m \to +\infty$, then $N \to \infty$, we deduce

$$\int_0^T \{\langle u(t), \ddot{v}(t) \rangle_* + c^2 (\nabla u(t), \nabla v(t))_0\} \, dt \qquad (9.67)$$

$$= \int_0^T (f(t), v(t))_0 \, dt + (g^1, \dot{v}(0)) - (g, v(0)).$$

Comparing (9.66) and (9.67), we conclude

$$(\dot{u}(0), \dot{v}(0)) - (u(0), v(0)) = (g^1, \dot{v}(0)) - (g, v(0))$$

for every $v \in C^2([0, T]; V)$, with $v(T) = \dot{v}(T) = 0$. The arbitrariness of $\dot{v}(0)$ and $v(0)$ gives

$$\dot{u}(0) = g^1 \quad \text{and} \quad u(0) = g.$$

Uniqueness. Assume $g = g^1 \equiv 0$ and $f \equiv 0$. We want to show that $u \equiv 0$. The proof would be easy if we could choose $\dot{u}$ as a test function in (9.55), but $\dot{u}(t)$ does not belong to V. Thus, for fixed s, set[13]

$$v(t) = \begin{cases} \int_t^s u(r)\,dr & \text{if } 0 \le t \le s \\ 0 & \text{if } s \le t \le T. \end{cases}$$

We have $v(t) \in V$ for all $t \in [0, T]$, so that we may insert it into (9.55). After an integration over $(0, T)$, we deduce

$$\int_0^s \left\{ \langle \ddot{u}(t), v(t) \rangle_* + c^2 (\nabla u(t), \nabla v(t))_0 \right\} dt = 0. \tag{9.68}$$

An integration by parts yields

$$\int_0^s \langle \ddot{u}(t), v(t) \rangle_* dt = - \int_0^s (\dot{u}(t), \dot{v}(t))_0 \, dt = \int_0^s (\dot{u}(t), u(t))_0 \, dt$$

$$= \frac{1}{2} \int_0^s \frac{d}{dt} \|u(t)\|_0^2 \, dt$$

since $v(s) = \dot{u}(0) = 0$ and $\dot{v}(t) = -u(t)$ if $0 < t < s$. On the other hand,

$$\int_0^s (\nabla u(t), \nabla v(t))_0 \, dt = - \int_0^s (\nabla \dot{v}(t), \nabla v(t))_0 \, dt = -\frac{1}{2} \int_0^s \frac{d}{dt} \|\nabla v(t)\|_0^2 \, dt.$$

Hence, from (9.68),

$$\int_0^s \frac{d}{dt} \left\{ \|u(t)\|_0^2 - c^2 \|\nabla v(t)\|_0^2 \right\} dt = 0$$

or

$$\|u(s)\|_0^2 + c^2 \|\nabla v(0)\|_0^2 = 0$$

which entails $u(s) \equiv 0$.

Stability. To prove the estimate in Theorem 9.12, use Proposition 7.16. to pass to the limit as $m \to \infty$ in (9.60). This gives the estimates for u and $\dot{u}$. The estimate for $\ddot{u}$ follows from the weak lower semicontinuity of the norm in $L^2(0, T; V^*)$. $\square$

---

[13] We follow *Evans*, 1998.

Problems

9.1. Consider the problem

$$\begin{cases} u_t - (a\,(x)\,u_x)_x + b\,(x)\,u_x + c(x)u = f\,(x,t) & 0 < x < 1, 0 < t < T \\ u\,(x,0) = g\,(x)\,, & 0 \le x \le 1 \\ u\,(0,t) = 0,\ u\,(1,t) = k\,(t)\,. & 0 \le t \le T. \end{cases}$$

1) Modifying u suitably, reduce the problem to homogeneous Dirichlet conditions.

2) Write a weak formulation for the new problem.

3) Prove the well-posedness of the problem, under suitable hypotheses on the coefficients a, b, c and the data f, g. Write a stability estimate for the original u.

9.2. Consider the Neumann problem (9.33) with non-homogeneous boundary condition $\partial_\nu u = h$, with $h \in L^2\,(S_T)$.

a) Give a weak formulation of the problem and derive the main estimates for the Galerkin approximations.

b) Deduce existence and uniqueness of the solution.

9.3. Prove a variant of the energy estimate in Theorem 9.1, by showing first that

$$\frac{d}{dt}\,\|u_m\,(t)\|_0^2 + 2\alpha\,\|u_m\,(t)\|_1^2 \le \|f\,(t)\|_0^2 + \|u_m\,(t)\|_0^2$$

and then using Gronwall's Lemma.

9.4. Derive the energy estimate for the Galerkin approximations u_m of the solution of the Cauchy-Neumann problem, whithout using the change of variable $w\,(t) = e^{-\lambda t}u\,(t)$.

[*Hint.* Add and subtract $\lambda\,\|u_m\,(t)\|_H^2$; use the weak coercivity of a and Gronwall's Lemma.]

9.5. $H^2 - regularity.$ State and prove a H^2-regularity result for the heat equation with homogeneous Neumann boundary conditions.

9.6. Consider the problem

$$\begin{cases} u_{tt} - c^2\Delta u = f & \text{in } \Omega \times (0,T) \\ u\,(\mathbf{x},0) = g\,(\mathbf{x})\,, u_t\,(\mathbf{x},0) = h\,(\mathbf{x}) & \text{in } \Omega \\ u_\nu\,(0,t) = 0. & \text{on } \partial\Omega \times [0,T]\,. \end{cases}$$

Write a weak formulation of the problem and prove the analogues of Theorems 9.10, 9.11 and 9.12.

9.7. *Concentrated reaction.* Consider the problem

$$\begin{cases} u_{tt} - u_{xx} + u\,(x,t)\,\delta\,(x) = 0 & -1 < x < 1,\ 0 < t < T \\ u\,(x,0) = g\,(x)\,, u_t\,(x,0) = h\,(x) & -1 \le x \le 1 \\ u\,(-1,t) = u\,(1,t) = 0. & 0 \le t \le T. \end{cases}$$

where $\delta(x)$ denotes the Dirac δ at the origin.

a) Write a weak formulation for the problem.

b) Prove the well-posedness of the problem, under suitable hypotheses on g and h.

[*Hint. a*) Let $V = H_0^1(-1,1)$ and $H = L^2(-1,1)$. The weak formulation is: *find* $u \in C([-1,1],V)$, *with* $\dot{u} \in C([-1,1],H)$ *and* $\ddot{u} \in C([-1,1],V^*)$, *such that, for every* $v \in V$,

$$\langle \ddot{u}(t), v \rangle_* + (u_x(t), v_x) + u(0,t)v(0) = 0 \qquad \text{for a.e. } t \in (0,T)$$

and $\|u(t) - g\|_V \to 0$, $\|\dot{u}(t) - h\|_H \to 0$ *as* $t \to 0$].

9.8. Consider the minimization of the cost functional

$$J(u,z) = \frac{1}{2}\int_\Omega |u(T) - u_d|^2 + \frac{\beta}{2}\int_{Q_T} z^2 d\mathbf{x}dt$$

under the condition (*state system*)

$$\begin{cases} u_{tt} - \Delta u = f + z & \text{in } Q_T \\ u = 0 & \text{on } S_T \\ u(\mathbf{x},0) = u_t(\mathbf{x},0) = 0 & \text{in } \Omega \end{cases}$$

where Ω is a C^2-domain and $f \in L^2(Q_T)$. Show that there exists a unique optimal control $z^* \in L^2(Q_T)$ and determine the optimality conditions (adjoint equation and Euler equation).

Appendix A

Fourier Series

Fourier coefficients – Expansion in Fourier series

A.1 Fourier coefficients

Let u be a $2T$-periodic function in $\mathbb{R}$ and assume that u can be expanded in a trigonometric series as follows:

$$u(x) = U + \sum_{k=1}^{\infty} \{a_k \cos k\omega x + b_k \sin k\omega x\} \tag{A.1}$$

where $\omega = \pi/T$.

First question: how u and the coefficients U, a_k and b_k are related to each other? To answer, we use the following so called *orthogonality relations*, whose proof is elementary:

$$\int_{-T}^{T} \cos k\omega x \ \cos m\omega x \ dx = \int_{-T}^{T} \sin k\omega x \ \sin m\omega x \ dx = 0 \qquad \text{if } k \neq m$$

$$\int_{-T}^{T} \cos k\omega x \ \sin m\omega x \ dx = 0 \quad \text{for all } k, m \geq 0.$$

Moreover

$$\int_{-T}^{T} \cos^2 k\omega x \ dx = \int_{-T}^{T} \sin^2 k\omega x \ dx = T. \tag{A.2}$$

Now, suppose that the series (A.1) converges *uniformly* in $\mathbb{R}$. Multiplying (A.1) by $\cos n\omega x$ and integrating term by term over $(-T, T)$, the orhogonality relations and (A.2) yield, for $n \geq 1$,

$$\int_{-T}^{T} u(x) \cos n\omega x \ dx = Ta_n$$

or

$$a_n = \frac{1}{T} \int_{-T}^{T} u(x) \cos n\omega x \ dx. \tag{A.3}$$

For $n = 0$ we get

$$\int_{-T}^{T} u(x) \ dx = 2UT$$

or, setting $U = a_0/2$,

$$a_0 = \frac{1}{T} \int_{-T}^{T} u(x) \ dx \tag{A.4}$$

which is coherent with (A.3) as $n = 0$.

Similarly, we find

$$b_n = \frac{1}{T} \int_{-T}^{T} u(x) \sin n\omega x \ dx. \tag{A.5}$$

Thus, if u has the uniformly convergent expansion (A.1), the coefficients a_n, b_n (with $a_0 = 2U$) must be given by the formulas (A.3) and (A.5). In this case we say that the trigonometric series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \{a_k \cos k\omega x + b_k \sin k\omega x\} \tag{A.6}$$

is the *Fourier series of* u and the coefficients (A.3), (A.4) and (A.5) are called the *Fourier coefficients* of u.

• *Odd and even functions.* If u is an *odd* function, i.e. $u(-x) = -u(x)$, we have $a_k = 0$ for every $k \geq 0$, while

$$b_k = \frac{2}{T} \int_{0}^{T} u(x) \sin k\omega x \ dx.$$

Thus, if u is odd, its Fourier series is a *sine* Fouries series:

$$u(x) = \sum_{k=1}^{\infty} b_k \sin k\omega x.$$

Similarly, if u is *even*, i.e. $u(-x) = u(x)$, we have $b_k = 0$ for every $k \geq 1$, while

$$a_k = \frac{2}{T} \int_{0}^{T} u(x) \cos k\omega x \ dx.$$

Thus, if u is even, its Fourier series is a *cosine* Fouries series:

$$u(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos k\omega x.$$

• *Fourier coefficients of a derivative.* Let $u \in C^1(\mathbb{R})$ be $2T-$periodic. Then we may compute the Fourier coefficients a'_k and b'_k of u'. We have, integrating by parts, for $k \geq 1$:

$$
\begin{aligned}
a'_k &= \frac{1}{T} \int_{-T}^{T} u'(x) \cos k\omega x \; dx \\
&= \frac{1}{T} [u(x) \cos k\omega x]_{-T}^{T} + \frac{k\omega}{T} \int_{-T}^{T} u(x) \sin k\omega x \; dx \\
&= \frac{k\omega}{T} \int_{-T}^{T} u(x) \sin k\omega x \; dx \\
&= k\omega b_k
\end{aligned}
$$

and

$$
\begin{aligned}
b'_k &= \frac{1}{T} \int_{-T}^{T} u'(x) \sin k\omega x \; dx \\
&= \frac{1}{T} [u(x) \sin k\omega x]_{-T}^{T} - \frac{k\omega}{T} \int_{-T}^{T} u(x) \cos k\omega x \; dx \\
&= -\frac{k\omega}{T} \int_{-T}^{T} u(x) \cos k\omega x \; dx \\
&= -k\omega a_k.
\end{aligned}
$$

Thus, the Fourier coefficients a'_k and b'_k are related to a_k and b_k by the following formulas:

$$
a'_k = k\omega b_k, \quad b'_k = -k\omega a_k. \tag{A.7}
$$

• *Complex form of a Fourier series.* Using the Euler identities

$$
e^{\pm ik\omega x} = \cos k\omega x \pm i \sin k\omega x
$$

the Fourier series (A.6) can be expressed in the complex form

$$
\sum_{k=-\infty}^{\infty} c_k e^{ik\omega x},
$$

where the complex Fourier coefficients c_k are given by

$$
c_k = \frac{1}{2T} \int_{-T}^{T} u(z) e^{-ik\omega z} dz.
$$

The relations among the real and the complex Fourier coefficients are:

$$
c_0 = \frac{1}{2} a_0
$$

and

$$
c_k = \frac{1}{2} (a_k - b_k), \quad c_{-k} = \bar{c}_k \quad \text{for } k > 0.
$$

A.2 Expansion in Fourier series

In the above computations we started from a function u admitting a uniform convergent expansion in Fourier series. Adopting a different point of view, let u be a $2T-$periodic function and assume we can compute its Fourier coefficients, given by formulas (A.3) and (A.5). Thus, we can *associate* with u its Fourier series and write

$$u \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} \{a_k \cos k\omega x + b_k \sin k\omega x\}.$$

The main questions are now the following:

1. Which conditions on u do assure "the convergence" of its Fourier series? Of course there are several notions of convergence (e.g pointwise, uniform, least squares).

2. If the Fourier series is convergent in some sense, does it always have sum u?

A complete answer to the above questions is not elementary. The convergence of a Fourier series is a rather delicate matter. We indicate some basic results (for the proofs, see e.g. *Rudin,* 1964 and 1974, *Royden,* 1988, or *Zygmund and Wheeden,* 1977.

• *Least squares or L^2 convergence.* This is perhaps the most natural type of convergence for Fourier series (see subsection 6.4.2). Let

$$S_N (x) = \frac{a_0}{2} + \sum_{k=1}^{N} \{a_k \cos k\omega x + b_k \sin k\omega x\}$$

be the $N-$partial sum of the Fourier series of u. We have

Theorem A.1 *Let u be a square integrable function*[1] *on $(-T, T)$. Then*

$$\lim_{N \to +\infty} \int_{-T}^{T} [S_N (x) - u (x)]^2 \, dx = 0.$$

Moreover, the following Parseval relation holds:

$$\frac{1}{T} \int_{-T}^{T} u^2 = \frac{a_0^2}{2} + \sum_{k=1}^{\infty} \left(a_k^2 + b_k^2\right). \tag{A.8}$$

Since the numerical series in the right hand side of (A.8) is convergent, we deduce the following important consequence:

Corollary A.1 (Riemann-Lebesgue).

$$\lim_{k \to +\infty} a_k = \lim_{k \to +\infty} b_k = 0$$

• *Pointwise convergence.* We say that u satisfies the *Dirichlet conditions* in $[-T, T]$ if it is continuous in $[-T, T]$ except possibly at a finite number of points

---

[1] That is $\int_{-T}^{T} u^2 < \infty$.

of jump discontinuity and moreover if the interval $[-T, T]$ can be partitioned in a finite numbers of subintervals such that u is monotone in each one of them.

The following theorem holds.

Theorem A.2. *If u satisfies the Dirichlet conditions in $[-T, T]$ then the Fourier series of u converges at each point of $[-T, T]$. Moreover[2]:*

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \{a_k \cos k\omega x + b_k \sin k\omega x\} = \begin{cases} \dfrac{u(x+) + u(x-)}{2} & x \in (-T, T) \\[2mm] \dfrac{u(T-) + u(-T+)}{2} & x = \pm T \end{cases}$$

In particular, under the hypotheses of Theorem A.2, at every point x of continuity of u the Fourier series converges to $u(x)$.

• *Uniform convergence.* A simple criterion of uniform convergence is provided by the Weierstrass test (see Section 1.4). Since

$$|a_k \cos k\omega x + b_k \sin k\omega x| \leq |a_k| + |b_k|$$

we deduce: *If the numerical series*

$$\sum_{k=1}^{\infty} |a_k| \quad \text{and} \quad \sum_{k=1}^{\infty} |b_k|$$

are convergent, then the Fourier series of u is uniformly convergent in $\mathbb{R}$, with sum u.

This is the case, for instance, if $u \in C^1(\mathbb{R})$ and is $2T$ periodic. In fact, from (A.7) we have for every $k \geq 1$,

$$a_k = -\frac{1}{\omega k} b_k' \quad \text{and} \quad b_k = \frac{1}{\omega k} a_k'.$$

Therefore

$$|a_k| \leq \frac{1}{\omega k^2} + (b_k')^2$$

and

$$|b_k| \leq \frac{1}{\omega k^2} + (a_k')^2.$$

Now, the series $\sum \frac{1}{k^2}$ is convergent. On the other hand, also the series

$$\sum_{k=1}^{\infty} (a_k')^2 \quad \text{and} \quad \sum_{k=1}^{\infty} (b_k')^2$$

are convergent, by Parseval's relation (A.8) applied to u' in place of u. The conclusion is that *if $u \in C^1(\mathbb{R})$ and $2T$ periodic, its Fourier series is uniformly convergent in $\mathbb{R}$ with sum u.*

---

[2] We set $f(x\pm) = \lim_{y \to \pm x} f(y)$.

Another useful result is a refinement of Theorem A.2.

Theorem A.3 *Assume u satisfies the Dirichlet conditions in $[-T, T]$. Then:*

a) If u is continuous in $[a, b] \subset (-T, T)$, then its Fourier series converges uniformly in $[a, b]$.

b) If u is continuous in $[-T, T]$ and $u(-T) = u(T)$, then its Fourier series converges uniformly in $[-T, T]$ (and therefore in $\mathbb{R}$).

Appendix B

Measures and Integrals

Lebesgue Measure and Integral

B.1 Lebesgue Measure and Integral

B.1.1 A counting problem

Two persons, that we denote by $\mathcal{R}$ and $\mathcal{L}$, must compute the total value of M coins, ranging from 1 to 50 cents. $\mathcal{R}$ decides to group the coins arbitrarily in piles of, say, 10 coins each, then to compute the value of each pile and finally to sum all these values. $\mathcal{L}$, instead, decides to partition the coins according to their value, forming piles of 1-cent coins, of 5-cents coins and so on. Then he computes the value of each pile and finally sums all their values.

In more analytical terms, let

$$V : M \to \mathbb{N}$$

a *value function* that associates to each element of M (i.e. each coin) its value. $\mathcal{R}$ partitions the **domain** of V in disjoint subsets, sums the values of V in such subsets and then sums everything. $\mathcal{L}$ considers each point p in the **image** of V (the value of a single coin), considers the inverse image $V^{-1}(p)$ (the pile of coins with the same value p), computes the corresponding value and finally sums over every p.

These two ways of counting correspond to the strategy behind the definitions of the integrals of Riemann and Lebesgue, respectively. Since V is defined on a discrete set and is integer valued, in both cases there is no problem in summing its values and the choice is determined by an efficiency criterion. Usually, the method of $\mathcal{L}$ is more efficient.

In the case of a real (or complex) function f, the "sums of its values" corresponds to an integration of f. While the construction of $\mathcal{R}$ remains rather elementary, the one of $\mathcal{L}$ requires new tools.

Let us examine the particular case of a *bounded* and *positive* function, defined on an interval $[a, b] \subset \mathbb{R}$. Thus, let

$$f : [a, b] \to [\inf f, \sup f] \,.$$

To construct the Riemann integral, we partition $[a, b]$ in subintervals $I_1, ..., I_N$ (the piles of $\mathcal{R}$), then we choose in each interval I_k a point ξ_k and we compute $f(\xi_k) \, l(I_k)$, where $l(I_k)$ is the length of I_k, (i.e. the value of the $k - th$ pile). Now we sum the values $f(\xi_k) \, l(I_k)$ and set

$$(\mathcal{R}) \int_a^b f = \lim_{\delta \to 0} \sum_{k=1}^N f(\xi_k) \, l(I_k),$$

where $\delta = \max \{ l(I_1), ..., l(I_N) \}$. If the limit is finite and moreover is independent of the choice of the points ξ_k, then this limit defines the Riemann integral of f in $[a, b]$.

Now, let us examine the Lebesgue strategy. This time we partition the interval $[\inf f, \sup f]$ in subintervals $[y_{k-1}, y_k]$ (the values of each coin for $\mathcal{L}$) with

$$\inf f = y_0 < y_1 < ... < y_{N-1} < y_N = \sup f.$$

Then we consider the inverse images $E_k = f^{-1} ([y_{k-1}, y_k])$ (the piles of homogeneous coins) and we would like to compute their .... *length*. However, in general E_k is *not* an interval or a union of intervals and, in principle, it could be a very irregular set so that it is not clear what is the "length" of E_k.

Thus, the need arises to associate with every E_k a *measure*, which replaces the length when E_k is an irregular set. This leads to the introduction of the *Lebesgue measure* of (practically every) set $E \subseteq \mathbb{R}$, denoted by $|E|$.

Once we know how to measure E_k (the number of coins in the $k - th$ pile), we choose an arbitrary point $\overline{\alpha}_k \in [y_{k-1}, y_k]$ and we compute $\overline{\alpha}_k \, |E_k|$ (the value of the $k - th$ pile). Then, we sum all the values $\overline{\alpha}_k \, |E_k|$ and set

$$(L) \int_a^b f = \lim_{\rho \to 0} \sum_{k=1}^N \overline{\alpha}_k \, |E_k| \,.$$

where ρ is the maximum among the lengths of the intervals $[y_{k-1}, y_k]$. It can be seen that under our hypotheses, the limit exists, is finite and is independent of the choice of $\overline{\alpha}_k$. Thus, we may always choose $\overline{\alpha}_k = y_{k-1}$. This remark leads to the definition of the Lebesgue integral in subsection B.3: the number $\sum_{k=1}^N y_{k-1} |E_k|$ is nothing else that the integral of a *simple function*, which approximates f from below and whose range is the finite set $y_0 < ... < y_{N-1}$. The integral of f is the supremum of these numbers.

The resulting theory has several advantages with respect to that of Riemann. For instance, the class of integrable functions is much wider and there is no need to distinguish among bounded or unbounded functions or integration domains.

Especially important are the convergence theorems presented in subsection B.1.4, which allow the possibility of interchanging the operation of limit and integration, under rather mild conditions.

Finally, the construction of the Lebesgue measure and integral can be greatly generalized as we will mention in subsection B.1.5.

For the proofs of the theorems stated in this Appendix, the interested reader can consult *Rudin,* 1964 and 1974, *Royden,* 1988, or *Zygmund and Wheeden,* 1977.

B.1.2 Measures and measurable functions

A measure in a set Ω is a *set function*, defined on a particular class of subsets of Ω called *measurable set* which "behaves well" with respect to union, intersection and complementation. Precisely:

Definition B.1 *A collection $\mathcal{F}$ of subsets of Ω is called $\sigma-$algebra if:*

(i) $\varnothing, \Omega \in \mathcal{F}$;

(ii) $A \in \mathcal{F}$ *implies* $\Omega \backslash A \in \mathcal{F}$;

(iii) *if* $\{A_k\}_{k \in \mathbb{N}} \subset \mathcal{F}$ *then also* $\cup A_k$ *and* $\cap A_k$ *belong to* $\mathcal{F}$.

Example B.1. If $\Omega = \mathbb{R}^n$, we can define the smallest $\sigma-$algebra containing all the open subsets of $\mathbb{R}^n$, called the *Borel $\sigma-$algebra*. Its elements are called *Borel sets*, typically obtained by countable unions and/or intersections of open sets.

Definition B.2 *Given a $\sigma-$algebra $\mathcal{F}$ in a set Ω, a measure on $\mathcal{F}$ is a function*

$$\mu : \mathcal{F} \to \mathbb{R}$$

such that:

(i) $\mu(A) \geq 0$ *for every* $A \in \mathcal{F}$;

(ii)   *if* $A_1, A_2, ...$*are pairwise disjoint sets in $\mathcal{F}$, then*

$$\mu(\cup_{k \geq 1} A_k) = \sum_{k \geq 1} \mu(A_k) \qquad (\sigma - additivity).$$

The elements of $\mathcal{F}$ are called measurable sets.

The Lebesgue measure in $\mathbb{R}^n$ is defined on a $\sigma-$*algebra* $\mathcal{M}$ containing the Borel $\sigma-$algebra, through the following theorem.

Theorem B.1 *There exists in $\mathbb{R}^n$ a $\sigma-$algebra $\mathcal{M}$ and a measure*

$$|\cdot|_n : \mathcal{M} \to [0, +\infty]$$

with the following properties:

1. *Each open and closed set belongs to $\mathcal{M}$.*
2. *If $A \in \mathcal{M}$ and A has measure zero, every subset of A belongs to $\mathcal{M}$ and has measure zero.*

3. *If*
$$A = \{\mathbf{x} \in \mathbb{R}^n : a_j < x_j < b_j; \ j = 1, ..., n\}$$
then $|A| = \prod_{j=1}^{n} (b_j - a_j)$.

The elements of $\mathcal{M}$ are called *Lebesgue measurable sets* and $|\cdot|_n$ (or simply $|\cdot|$ if no confusion arises) is called the $n-dimensional$ *Lebesgue measure.* Unless explicitly said, from now on, *measurable* means *Lebesgue measurable* and the measure is the Lebesgue measure.

Not every subset of $\mathbb{R}^n$ is measurable. However, the nonmeasurable ones are quite ... pathological[1] !

The sets of measure zero are quite important. Here are some examples: all countable sets, e.g. the set $\mathbb{Q}$ of rational numbers; straight lines or smooth curves in $\mathbb{R}^2$; straight lines, hyperplanes, smooth curves and surfaces in $\mathbb{R}^3$.

Notice that a straight line segment has measure zero in $\mathbb{R}^2$ but, of course not in $\mathbb{R}$.

We say that a *property holds almost everywhere in* $A \in \mathcal{M}$ (in short, a.e. in A) *if it holds at every point of* A *except that in a subset of measure zero.*

For instance, the sequence $f_k(x) = \exp\left(-n \sin^2 x\right)$ converges to zero a.e. in $\mathbb{R}$, a Lipschitz function is differentiable a.e. in its domain (Rademacher's Theorem 1.1).

The Lebesgue integral is defined for *measurable* functions, characterized by the fact that the inverse image of every closed set is measurable.

Definition B.3 *Let* $A \subseteq \mathbb{R}^n$ *be measurable, and* $f : A \to \mathbb{R}$. *We say that* f *is measurable if*
$$f^{-1}(C) \in \mathcal{F}$$
for any closed set $C \subseteq \mathbb{R}$.

If f is continuous, is measurable. The sum and the product of a finite number of measurable functions is measurable. The pointwise limit of a sequence of measurable functions is measurable.

If $f : A \to \mathbb{R}$, is measurable, we define its *essential supremum* or *least upper bound* by the formula:
$$\text{ess sup } f = \inf\{K : f \leq K \text{ a.e. in } A\}.$$

Note that, if $f = \chi_{\mathbb{Q}}$, the characteristic functions of the rational numbers, we have $\sup f = 1$, but $\text{ess sup } f = 0$, since $|\mathbb{Q}| = 0$.

Every measurable function may be approximated by **simple functions**. A function $s : A \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be **simple** if its range is constituted by a *finite number* of values $s_1, ..., s_N$, attained respectively on measurable sets $A_1, ..., A_N$, contained in A. Introducing the characteristic functions χ_{A_j}, we may write
$$s = \sum_{j=1}^{N} s_j \chi_{A_j}.$$

---

[1] See e.g. *Rudin,* 1974.

We have:

Theorem B.2. *Let $f : A \to \mathbb{R}$, be measurable. There exists a sequence $\{s_k\}$ of simple functions converging pointwise to f in A. Moreover, if $f \geq 0$, we may choose $\{s_k\}$ increasing.*

B.1.3 The Lebesgue integral

We define the Lebesgue integral of a measurable function on a measurable set A. For a simple function $s = \sum_{j=1}^{N} s_j \chi_{A_j}$ we set:

$$\int_A s = \sum_{j=1}^{N} s_j \, |A_j|$$

with the convenction that, if $s_j = 0$ and $|A_j| = +\infty$, then $s_j \, |A_j| = 0$.

If $f \geq 0$ is measurable, we define

$$\int_A f = \sup \int_A s$$

where the supremum is computed over the set of all simple functions s such that $s \leq f$ in A.

In general, if f is measurable, we write $f = f^+ - f^-$, where $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$ are the positive and negative parts of f, respectively. Then we set:

$$\int_A f = \int_A f^+ - \int_A f^-$$

under the condition that at least one of the two integrals in the right hand side is finite.

If both these integrals are finite, the function f is said to be **integrable** or **summable** in A. From the definition, it follows immediately that a measurable functions f is *integrable if and only if $|f|$ is integrable*.

All the functions Riemann integrable in a set A are Lebesgue integrable as well. An interesting example of non integrable function in $(0, +\infty)$ is given by $h(x) = \sin x / x$. In fact[2]

$$\int_0^{+\infty} \frac{|\sin x|}{x} dx = +\infty.$$

On the contrary, it may be proved that

$$\lim_{N \to +\infty} \int_0^N \frac{\sin x}{x} dx = \frac{\pi}{2}.$$

and therefore the improper Riemann integral of h is finite.

---

[2] We may write

$$\int_0^{+\infty} \frac{|\sin x|}{x} dx = \sum_{k=1}^{\infty} \int_{(k-1)\pi}^{k\pi} \frac{|\sin x|}{x} dx \geq \sum_{k=1}^{\infty} \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin x| \, dx = \sum_{k=1}^{\infty} \frac{2}{k\pi} = +\infty.$$

The set of the integrable functions in A is denoted by $L^1(A)$. If we identify two functions when they agree a.e. in A, $L^1(A)$ becomes a Banach space with the norm[3]

$$\|f\|_{L^1(A)} = \int_A |f|.$$

We denote by $L^1_{loc}(A)$ the set of *locally summable functions*, i.e. of the functions which are summable in every compact subset of A.

B.1.4 Some fundamental theorems

The following theorems are among the most important and useful in the theory of integration.

Theorem B.3 (Dominated Convergence Theorem). *Let $\{f_k\}$ be a sequence of summable functions in A such that $f_k \to f$ a.e.in A. If there exists $g \geq 0$, summable in A and such that $|f_k| \leq g$ a.e. in A, then f is summable and*

$$\|f_k - f\|_{L^1(A)} \to 0 \quad as\ k \to +\infty.$$

In particular

$$\lim_{k \to \infty} \int_A f_k = \int_A f.$$

Theorem B.4 *Let $\{f_k\}$ be a sequence of summable functions in A such that $\|f_k - f\|_{L^1(A)} \to 0$ as $k \to +\infty$. Then there exists a subsequence $\{f_{k_j}\}$ such that $f_{k_j} \to f$ a.e. as $j \to +\infty$.*

Theorem B.5 (Monotone Convergence Theorem). *Let $\{f_k\}$ be a sequence of nonnegative, measurable functions in A such that*

$$f_1 \leq f_2 \leq ... \leq f_k \leq f_{k+1} \leq ... .$$

Then

$$\lim_{k \to \infty} \int_A f_k = \int_A \lim_{k \to \infty} f_k.$$

Example B.2. A typical situation we often encounter in this book is the following. Let $f \in L^1(A)$ and, for $\varepsilon > 0$, set $A_\varepsilon = \{|f| > \varepsilon\}$. Then, we have

$$\int_{A_\varepsilon} f \to \int_A f \quad as\ \varepsilon \to 0.$$

This follows from Theorem B.4 since, for every sequence $\varepsilon_j \to 0$, we have $|f|\chi_{A_{\varepsilon_j}} \leq |f|$ and therefore

$$\int_{A_{\varepsilon_j}} f = \int_A f\chi_{A_{\varepsilon_j}} \to \int_A f \quad as\ \varepsilon \to 0.$$

---

[3] See Chapter 6.

Let $C_0(A)$ be the set of continuous functions in A, compactly supported in A. An important fact is that any summable function may be approximated by a function in $C_0(A)$.

Theorem B.6. *Let $f \in L^1(A)$. Then, for every $\delta > 0$, there exists a continuous function $g \in C_0(A)$ such that*

$$\|f - g\|_{L^1(A)} < \delta.$$

The fundamental theorem of calculus extends to the Lebesgue integral in the following form:

Theorem B.7. (Differentiation). *Let $f \in L^1_{loc}(\mathbb{R})$. Then*

$$\frac{d}{dx} \int_a^x f(t)\, dt = f(x) \qquad \text{a.e. } x \in \mathbb{R}.$$

Finally, the integral of a summable function can be computed via iterated integrals in any order. Precisely, let

$$I_1 = \{\mathbf{x} \in \mathbb{R}^n : -\infty \le a_i < x_i < b_i \le \infty; \; i = 1, ..., n\}$$

and

$$I_2 = \{\mathbf{y} \in \mathbb{R}^m : -\infty \le a_j < y_j < b_j \le \infty; \; j = 1, ..., m\}.$$

Theorem B.8 (Fubini). *Let f be summable in $I = I_1 \times I_2 \subset \mathbb{R}^n \times \mathbb{R}^m$. Then*

1. $f(\mathbf{x}, \cdot) \in L^1(I_2)$ *for a.e.* $\mathbf{x} \in I_1$, *and* $f(\cdot, \mathbf{y}) \in L^1(I_1)$ *for a.e.* $\mathbf{y} \in I_2$,

2. $\int_{I_2} f(\cdot, \mathbf{y})\, d\mathbf{y} \in L^1(I_1)$ *and* $\int_{I_1} f(\mathbf{x}, \cdot)\, d\mathbf{x} \in L^1(I_2)$,

3. *the following formulas hold:*

$$\int_I f(\mathbf{x}, \mathbf{y})\, d\mathbf{x}d\mathbf{y} = \int_{I_1} d\mathbf{x} \int_{I_2} f(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = \int_{I_2} d\mathbf{x} \int_{I_1} f(\mathbf{x}, \mathbf{y})d\mathbf{y}.$$

B.1.5 Probability spaces, random variables and their integrals

Let $\mathcal{F}$ be a $\sigma-algebra$ in a set Ω. A *probability measure P on $\mathcal{F}$* is a measure in the sense of definition B.2, such that $P(\Omega) = 1$ and

$$P : \mathcal{F} \to [0, 1].$$

The triplet $(\Omega, \mathcal{F}, P)$ is called a *probability space*. In this setting, the elements ω of Ω are *sample points*, while a set $A \in \mathcal{F}$ has to be interpreted as an *event*. $P(A)$ is the probability of (occurrence of) A.

A typical example is given by the triplet

$$\Omega = [0, 1], \; \mathcal{F} = \mathcal{M} \cap [0, 1], \; P(A) = |A|$$

which models a *uniform random choice* of a point in $[0, 1]$.

A $1-dimensional$ *random variable in* $(\Omega, \mathcal{F}, P)$ is a function

$$X : \Omega \to \mathbb{R}$$

such that X is $\mathcal{F}-measurable,$ that is

$$X^{-1}(C) \in \mathcal{F}$$

for each closed set $C \subseteq \mathbb{R}$.

Example B.3. The *number k of steps to the right after N steps* in the random walk of Section 2.4 is a random variable. Here Ω is the set of walks of N steps.

By the same procedure used to define the Lebesgue integral we can define the integral of a random variable with respect to a probability measure. We sketch the main steps.

If X is *simple*, i.e. $X = \sum_{j=1}^{N} s_j \chi_{A_j}$, we define

$$\int_{\Omega} X \, dP = \sum_{j=1}^{N} s_j P(A_j).$$

If $X \geq 0$ we set

$$\int_{\Omega} X \, dP = \sup \left\{ \int_{\Omega} Y \, dP : Y \leq X, \ Y \text{ simple} \right\}.$$

Finally, if $X = X^+ - X^-$ we define

$$\int_{\Omega} X \, dP = \int_{\Omega} X^+ \, dP - \int_{\Omega} X^- \, dP$$

provided at least one of the integral on the right hand side is finite.

In particular, if

$$\int_{\Omega} |X| \, dP < \infty,$$

then

$$E(X) = \langle X \rangle = \int_{\Omega} X \, dP$$

is called *the expected value (or mean value or expectation)* of X, while

$$\text{Var}(X) = \int_{\Omega} (X - E(X))^2 \, dP$$

is called the *variance of X*.

Analogous definitions can be given componentwise for $n-dimensional$ *random variables*

$$\mathbf{X} : \Omega \to \mathbb{R}^n.$$

Appendix C

Identities and Formulas

Gradient, Divergence, Curl, Laplacian – Formulas

C.1 Gradient, Divergence, Curl, Laplacian

Let $\mathbf{F}$ be a smooth vector field and f a smooth real function, in $\mathbb{R}^3$.

Orthogonal cartesian coordinates

1. *gradient*:
$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

2. *divergence* ($\mathbf{F} = F_1\mathbf{i} + F_1\mathbf{j} + F_3\mathbf{k}$):
$$\text{div } \mathbf{F} = \frac{\partial}{\partial x}F_1 + \frac{\partial}{\partial y}F_2 + \frac{\partial}{\partial z}F_3$$

3. *laplacian*:
$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$$

4. *curl*:
$$\text{curl } \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ F_1 & F_2 & F_3 \end{vmatrix}$$

Cylindrical coordinates

$$x = r\cos\theta, \ y = r\sin\theta, \ z = z \qquad (r > 0, \ 0 \le \theta \le 2\pi)$$

$$\mathbf{e}_r = \cos\theta\mathbf{i} + \sin\theta\mathbf{j}, \quad \mathbf{e}_\theta = -\sin\theta\mathbf{i} + \cos\theta\mathbf{j}, \quad \mathbf{e}_z = \mathbf{k}.$$

1. *gradient*:
$$\nabla f = \frac{\partial f}{\partial r}\mathbf{e}_r + \frac{1}{r}\frac{\partial f}{\partial \theta}\mathbf{e}_\theta + \frac{\partial f}{\partial z}\mathbf{e}_z$$

2. *divergence* $(\mathbf{F} = F_r\mathbf{e}_r + F_\theta\mathbf{e}_\theta + F_z\mathbf{k})$:
$$\text{div }\mathbf{F} = \frac{1}{r}\frac{\partial}{\partial r}(rF_r) + \frac{1}{r}\frac{\partial}{\partial \theta}F_\theta + \frac{\partial}{\partial z}F_z$$

3. *laplacian*:
$$\Delta f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2} = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}$$

4. *curl*:
$$\text{curl }\mathbf{F} = \frac{1}{r}\begin{vmatrix} \mathbf{e}_r & r\mathbf{e}_\theta & \mathbf{e}_z \\ \partial_r & \partial_\theta & \partial_z \\ F_r & rF_\theta & F_z \end{vmatrix}$$

Spherical coordinates

$$x = r\cos\theta\sin\psi, \ y = r\sin\theta\sin\psi, \ z = r\cos\psi \qquad (r > 0,\ 0 \le \theta \le 2\pi,\ 0 \le \psi \le \pi)$$

$$\begin{aligned} \mathbf{e}_r &= \cos\theta\sin\psi\mathbf{i} + \sin\theta\sin\psi\mathbf{j} + \cos\psi\mathbf{k} \\ \mathbf{e}_\theta &= -\sin\theta\mathbf{i} + \cos\theta\mathbf{j} \\ \mathbf{e}_\psi &= \cos\theta\cos\psi\mathbf{i} + \sin\theta\cos\psi\mathbf{j} - \sin\psi\mathbf{k}. \end{aligned}$$

1. *gradient*:
$$\nabla f = \frac{\partial f}{\partial r}\mathbf{e}_r + \frac{1}{r\sin\psi}\frac{\partial f}{\partial \theta}\mathbf{e}_\theta + \frac{1}{r}\frac{\partial f}{\partial \psi}\mathbf{e}_\psi$$

2. *divergence* $(\mathbf{F} = F_r\mathbf{e}_r + F_\theta\mathbf{e}_\theta + F_\psi\mathbf{e}_\psi)$:
$$\text{div }\mathbf{F} = \underbrace{\frac{\partial}{\partial r}F_r + \frac{2}{r}F_r}_{\text{radial part}} + \frac{1}{r}\underbrace{\left[\frac{1}{\sin\psi}\frac{\partial}{\partial \theta}F_\theta + \frac{\partial}{\partial \psi}F_\psi + \cot\psi F_\psi\right]}_{\text{spherical part}}$$

3. *laplacian*:
$$\Delta f = \underbrace{\frac{\partial^2 f}{\partial r^2} + \frac{2}{r}\frac{\partial f}{\partial r}}_{\text{radial part}} + \frac{1}{r^2}\underbrace{\left\{\frac{1}{(\sin\psi)^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial \psi^2} + \cot\psi\frac{\partial f}{\partial \psi}\right\}}_{\text{spherical part (Laplace-Beltrami operator)}}$$

4. *curl*:
$$\text{rot }\mathbf{F} = \frac{1}{r^2\sin\psi}\begin{vmatrix} \mathbf{e}_r & r\mathbf{e}_\psi & r\sin\psi\mathbf{e}_\theta \\ \partial_r & \partial_\psi & \partial_\theta \\ F_r & rF_\psi & r\sin\psi F_z \end{vmatrix}.$$

C.2 Formulas

Gauss' formulas

In $\mathbb{R}^n$, $n \geq 2$, let:

- Ω be a bounded smooth domain and and $\boldsymbol{\nu}$ the outward unit normal on $\partial\Omega$;
- $\mathbf{u}, \mathbf{v}$ be vector fields of class $C^1\left(\overline{\Omega}\right)$;
- φ, ψ be real functions of class $C^1\left(\overline{\Omega}\right)$;
- $d\sigma$ be the area element on $\partial\Omega$.

1. $\int_\Omega \text{div } \mathbf{u} \, d\mathbf{x} = \int_{\partial\Omega} \mathbf{u} \cdot \boldsymbol{\nu} \, d\sigma$ (Divergence Theorem)

2. $\int_\Omega \nabla\varphi \, d\mathbf{x} = \int_{\partial\Omega} \varphi\boldsymbol{\nu} \, d\sigma$

3. $\int_\Omega \Delta\varphi \, d\mathbf{x} = \int_{\partial\Omega} \nabla\varphi \cdot \boldsymbol{\nu} \, d\sigma = \int_{\partial\Omega} \partial_\nu\varphi \, d\sigma$

4. $\int_\Omega \psi \, \text{div}\mathbf{F} \, d\mathbf{x} = \int_{\partial\Omega} \psi\mathbf{F} \cdot \boldsymbol{\nu} \, d\sigma - \int_\Omega \nabla\psi \cdot \mathbf{F} \, d\mathbf{x}$ (Integration by parts)

5. $\int_\Omega \psi\Delta\varphi \, d\mathbf{x} = \int_{\partial\Omega} \psi\partial_\nu\varphi \, d\sigma - \int_\Omega \nabla\varphi \cdot \nabla\psi \, d\mathbf{x}$ (Green's identity I)

6. $\int_\Omega (\psi\Delta\varphi - \varphi\Delta\psi) \, d\mathbf{x} = \int_{\partial\Omega} (\psi\partial_\nu\varphi - \varphi\partial_\nu\psi) \, d\sigma$ (Green's identity II)

7. $\int_\Omega \text{curl } \mathbf{u} \, d\mathbf{x} = - \int_{\partial\Omega} \mathbf{u} \times \boldsymbol{\nu} \, d\sigma$

8. $\int_\Omega \mathbf{u} \cdot \text{curl } \mathbf{v} \, d\mathbf{x} = \int_{\boldsymbol{\Omega}} \mathbf{v} \cdot \text{curl } \mathbf{u} \, d\mathbf{x} - \int_{\partial\Omega} (\mathbf{u} \times \mathbf{v}) \cdot \boldsymbol{\nu} \, d\sigma$.

Identities

1. div curl $\mathbf{u} = 0$
2. curl $\nabla\varphi = \mathbf{0}$
3. div $(\varphi\mathbf{u}) = \varphi \text{ div } \mathbf{u} + \nabla\varphi \cdot \mathbf{u}$
4. curl $(\varphi\mathbf{u}) = \varphi \text{ curl } \mathbf{u} + \nabla\varphi \times \mathbf{u}$
5. curl $(\mathbf{u} \times \mathbf{v}) = (\mathbf{v} \cdot \nabla)\mathbf{u} - (\mathbf{u} \cdot \nabla)\mathbf{v} + (\text{div } \mathbf{v})\mathbf{u} - (\text{div } \mathbf{u})\mathbf{v}$
6. div $(\mathbf{u} \times \mathbf{v}) = \text{curl}\mathbf{u} \cdot \mathbf{v} - \text{curl}\mathbf{v} \cdot \mathbf{u}$
7. $\nabla(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \times \text{curl } \mathbf{v} + \mathbf{v} \times \text{curl } \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{u}$
8. $(\mathbf{u} \cdot \nabla)\mathbf{u} = \text{curl}\mathbf{u} \times \mathbf{u} + \frac{1}{2}\nabla|\mathbf{u}|^2$
9. curl curl $\mathbf{u} = \nabla(\text{div } \mathbf{u}) - \Delta\mathbf{u}$.

References

Partial Differential Equations

E. DiBenedetto, *Partial Differential Equations.* Birkhäuser, 1995.

L. C. Evans, *Partial Differential Equations.* A.M.S., Graduate Studies in Mathematics, 1998.

A. Friedman, *Partial Differential Equations of parabolic Type.* Prentice-Hall, Englewood Cliffs, 1964.

D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order.* II edition, Springer-Verlag, Berlin Heidelberg, 1998.

R. B. Guenter and J. W. Lee, *Partial Differential Equations of Mathematical Physics and Integral Equations.* Dover Publications, Inc., New York, 1998.

F. John, *Partial Differential Equations* (4th ed.). Springer-Verlag, New York, 1982.

O. Kellog, *Foundations of Potential Theory.* Springer-Verlag, New York, 1967.

G. M. Lieberman, *Second Order Parabolic Partial Differential Equations.* World Scientific, Singapore, 1996.

J. L. Lions and E. Magenes, *Nonhomogeneous Boundary Value Problems and Applications.* Springer-Verlag, New York, 1972.

R. McOwen, *Partial Differential Equations: Methods and Applications.* Prentice-Hall, New Jersey, 1996.

M. Protter and H. Weinberger, *Maximum Principles in Differential Equations.* Prentice-Hall, Englewood Cliffs, 1984.

M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations.* Springer-Verlag, New York, 1993.

J. Rauch, *Partial Differential Equations.* Springer-Verlag, Heidelberg 1992.

J. Smoller, *Shock Waves and Reaction-Diffusion Equations.* Springer-Verlag, New York, 1983.

W. Strauss, *Partial Differential Equation: An Introduction.* Wiley, 1992.

D. V. Widder, *The Heat Equation.* Academic Press, New York, 1975.

Mathematical Modelling

A. J. Acheson, *Elementary Fluid Dynamics.* Clarendon Press-Oxford, 1990.

J. Billingham and A. C. King, *Wave Motion.* Cambridge University Press, 2000.

R. Courant and D. Hilbert, *Methods of Mathematical Phisics.* Vol. 1 e 2. Wiley, New York, 1953.

R. Dautray and J. L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology,* Vol. 1-5. Springer-Verlag, Berlin Heidelberg, 1985.

C. C. Lin and L. A. Segel. *Mathematics Applied to Deterministic Problems in the Natural Sciences.* SIAM Classics in Applied Mathematics, (4th ed.) 1995.

J. D. Murray, *Mathematical Biology.* Springer-Verlag, Berlin Heidelberg, 2001.

L. A. Segel, *Mathematics Applied to Continuum Mechanics.* Dover Publications, Inc., New York, 1987.

G. B. Whitham, *Linear and Nonlinear Waves.* Wiley-Interscience, 1974.

Analysis and Functional Analysis

R. Adams, *Sobolev Spaces.* Academic Press, New York, 1975.

H. Brezis, *Analyse Fonctionnelle.* Masson, 1983.

L. C. Evans and R. F. Gariepy, *Measure Theory and Fine properties of Functions.* CRC Press, 1992.

V. G. Maz'ya, *Sobolev Spaces.* Springer-Verlag, Berlin Heidelberg, 1985.

W. Rudin, *Principles of Mathematical Analysis* (3th ed.). Mc Graw-Hill, 1976.

W. Rudin, *Real and Complex Analysis* (2th ed). Mc Graw-Hill, 1974.

L. Schwartz, *Théorie des Distributions.* Hermann, Paris, 1966.

K. Yoshida, *Functional Analysis.* Springer-Verlag, Berlin Heidelberg, 1965.

Numerical Analysis

R. Dautray and J. L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology.* Vol. 4 and 6. Springer-Verlag, Berlin Heidelberg, 1985.

A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations.* Springer-Verlag, Berlin Heidelberg, 1994.

Stochastic Processes and Finance

M. Baxter and A. Rennie, *Financial Calculus: An Introduction to Derivative Pricing.* Cambridge U. Press, 1996.

L. C. Evans, *An Introduction to Stochastic Differential Equations,* Lecture Notes, http://math.berkeley.edu/~evans/

B. K. Øksendal, *Stochastic Differential Equations: An Introduction with Applications.* (4th ed.), Springer-Verlag, Berlin Heidelberg, 1995.

P. Wilmott, S. Howison and J. Dewinne, *The Mathematics of Financial Derivatives. A Student Introduction.* Cambridge U. Press, 1996.

Index

End of printing December 2007