Methods of Applied Mathematics III: Partial Differential Equations

Bernard Deconinck
Department of Applied Mathematics
University of Washington
Campus Box 352420
Seattle, WA, 98195, USA

April 25, 2020

Prolegomenon

These are the lecture notes for Amath 569: Methods of applied mathematics III: partial differential equations. This is the first year these notes are typed up, thus it is guaranteed that these notes are full of mistakes of all kinds, both innocent and unforgivable. Please point out these mistakes to me so they may be corrected for the benefit of your successors. If you think that a different phrasing of something would result in better understanding, please let me know.

These lecture notes are not meant to supplant the textbook used with this course. The main textbook is "Partial differential equations of mathematical physics and integral equations", by Ron Guenther and John Lee (Dover, 1996). Other recommended sources are Whitham's "Linear and nonlinear waves" (Wiley, 1999), and Evans' "Partial differential equations" (AMS, 1998).

Contents

Chapter 1

Motivation and important problems related to partial differential equations

1.1 Motivation

Most courses on partial differential equations start off by deriving some of the basic equations from the physical models from which they come. You have probably already seen such derivations before, so we will spend little time on this. Let's look at some of the basic partial differential equations we'll consider in this course.

a) $u_{tt} = c^2(u_{xx} + u_{yy} + u_{zz})$. This is the **wave equation**. It arises in all kinds of settings where signals propagate without undergoing any damping. Typically, you'll see a derivation of this equation in textbooks using a vibrating string setting.

b) $u_t = \sigma(u_{xx} + u_{yy} + u_{zz})$. This is the **heat equation**. It often arises in settings where dissipation is an important mechanism. A typical example is the redistribution of heat in some domain.

c) $u_{xx} + u_{yy} + u_{zz} = 0$. This is **Laplace's equation**. It can be derived in its own right, but it is also seen to govern the stationary solutions of the wave and heat equations.

I will not take you through these derivations. Rather, I'd like to discuss some of the important systems of partial differential equations, and what some outstanding problems are.

The Navier-Stokes equations

There are many incarnations of the Navier-Stokes equations. One of the most popular is

$$\begin{cases} \frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} &= \nu \Delta \boldsymbol{u} - \nabla p \\ \nabla \cdot \boldsymbol{u} &= 0, \end{cases}$$

where the three-dimensional vector $\boldsymbol{u}$ represents the velocity field of the flow, ν is the viscosity, and p is the pressure. These equations describe incompressible viscous flow. Proving that these equations have solutions, and establishing these solutions, is the topic of one of the Clay Institute million dollar prizes.

Maxwell's equations

These equations describe the dynamical response of the electric and magnetic fields to different media. They are a popular T-shirt topic.

$$\begin{aligned} \nabla \cdot \boldsymbol{D} &= \rho, \\ \nabla \times \boldsymbol{H} - \frac{\partial \boldsymbol{D}}{\partial t} &= \boldsymbol{J}, \\ \nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} &= 0, \\ \nabla \cdot \boldsymbol{B} &= 0, \end{aligned}$$

with the additional two equations $\boldsymbol{D} = \boldsymbol{\epsilon E}$ and $\boldsymbol{B} = \boldsymbol{\mu H}$. Here $\boldsymbol{D}$ is the electric displacement, ρ is the charge density, $\boldsymbol{H}$ is the magnetic field, $\boldsymbol{J}$ is the current density, $\boldsymbol{E}$ is the electric field, and $\boldsymbol{B}$ is the magnetic induction. Lastly, $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ represent the dielectric tensor and the magnetic permeability tensor respectively. Additional constitutive relationships give connections between ρ and $\boldsymbol{J}$ and the other field quantities.

Note that Maxwell's equations are linear. Any nonlinearity enters through the response of the fields to the medium.

The Schrödinger equation

The single-particle Schrödinger equation is given by

$$-\frac{\hbar^2}{2m}\Delta\psi + V(x)\psi = i\hbar\frac{\partial \psi}{\partial t}.$$

This equation is also linear. It has to be, due to the essence of the superposition principle in quantum mechanics.

The nonlinear Schrödinger equation

$$i\psi_t = -\psi_{xx} + \alpha|\psi|^2\psi.$$

This equation describes the slow modulation of a wave packet in a nonlinear medium. Higher-dimensional versions of this equation exist as well.

The Korteweg-deVries equation

$$u_t = 6uu_x + u_{xxx}.$$

This equation describes the propagation of long waves in a dispersive medium.

Almost everything that we will do in this course deals with linear partial differential equations. How can we get away with this?

- We'll deal with nonlinear equations in a different course: amath573 (or at least: we'll deal with *some* nonlinear partial differential equations in another course).
- If we want to understand nonlinear partial differential equations, we should understand linear partial differential equations first.
- For "small" solutions, we can linearize the nonlinear model, resulting in a linear set of equations.
- Pragmatic: there's so much more we can do with linear partial differential equations than with nonlinear ones. But even then, there is still lots we cannot do.

All courses on linear partial differential equations teach the principle of superposition as the main, if not the only, method of solution. This method will play a big role here as well, but we will treat it with a little more rigor. In its simplest form, this leads to Fourier transforms and Fourier series. These occur when our equations are linear and have constant coefficients.

If the geometry of the problem becomes a bit more complicated (spherical, cylindrical) or if the equation has non-constant coefficients, we are often led to the theory of special functions.

But, most often, we are dealing with problems that we cannot attack with analytical methods, so we need numerics or perturbation methods. Even so, especially in these cases, it is important to establish some decent intuition about what the solution behavior of an equation will be, so as to be able to check the validity of the numerical or perturbation results.

1.2 Well-posedness of partial differential equations

Given that we typically use partial differential equations to obtain information about physical or other application problem, it is important that the partial differential equation is well-posed.

What do we mean when we say a problem is well posed?

- There are a **sufficient number of initial conditions:** if there are time derivatives in the problem, an equal amount of initial conditions should be specified, just like for ordinary differential equations.

- If there are **boundaries**, **the conditions** at these boundaries must be **consistent**, such that solutions actually do exist. Note: this is hard to check in many practical problems.
- And most importantly, and typically hardest to check: the solution of the problem should be stable with respect to perturbations in the initial and boundary conditions.

Why is well-posedness important?

Since we're dealing with a physical problem, we would like these problems to (a) have a solution, (b) not to be too sensitive with respect to measured quantities such as initial and boundary conditions.

Example: Consider the heat equation

$$\begin{cases} u_t &= \alpha \Delta u \\ u(\boldsymbol{x}, 0) &= 1, \end{cases}$$

for $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$ and $\alpha > 0,\ \alpha \in \mathbb{R}$ Suppose we have a "final condition" $u(\boldsymbol{x}, 0)$, but we want to know where it came from. In other words, we want to solve the heat equation backwards. Thus, we'd have to solve

$$\begin{cases} u_t &= -\alpha \Delta u \\ u(\boldsymbol{x}, 0) &= 1, \end{cases}$$

Note that

$$u_n(\boldsymbol{x}, t) = 1 + \frac{1}{n} e^{3\alpha n^2 t} \sin(nx) \sin(ny) \sin(nz)$$

solves the initial-value problem

$$\begin{cases} u_t &= -\alpha \Delta u \\ u(\boldsymbol{x}, 0) &= 1 + \frac{1}{n} \sin(nx) \sin(ny) \sin(nz), \end{cases}$$

for $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$.

Thus, we have two initial-value problems that are arbitrarily close together, but we see that as $t > 0$, the solution of the second problem diverges exponentially from that of the first. We conclude that solutions of the backward heat equation are not stable with respect to perturbations on their initial and/or boundary conditions.

There's a physical reason for this: the $u = 1$ solution is an asymptotically stable solution of the heat equation, so many things are attracted to it. It is thus unreasonable to ask what initial disturbance produced it.

We will return to issues of well-posedness throughout these notes.

Chapter 2

Partial differential equations of the first order: the method of characteristics

2.1 First-order equations

In this section we look at partial differential equations of the form

$$\frac{\partial u}{\partial t} + c(u)\frac{\partial u}{\partial x} = F(x,t,u) \tag{2.1}$$

Such an equation is called **quasi-linear**, because it is linear in the derivatives of $u(x,t)$. We could have written down a slightly more general equation, by considering $c(u)$ to be an explicit function of x and t as well.

We will use the method of characteristics to solve equations like this. This method is more important than just for solving equations like (2.1), as it also allows us to classify partial differential equations of higher order. We'll get back to that later.

Let $u(x,0) = u_0(x)$ be the initial condition for (2.1), given on the whole line $x \in \mathbb{R}$, see Fig. 2.1. We want to know what happens to $u(x,t)$ as $t > 0$. Let's look at the homogeneous case first.

1. Homogeneous case

Our equation reads

$$\frac{\partial u}{\partial t} + c(u)\frac{\partial u}{\partial x} = 0. \tag{2.2}$$

Note that this can be rewritten as

$$\frac{du}{dt} = 0,$$

where

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{dx}{dt}\frac{\partial u}{\partial x},$$

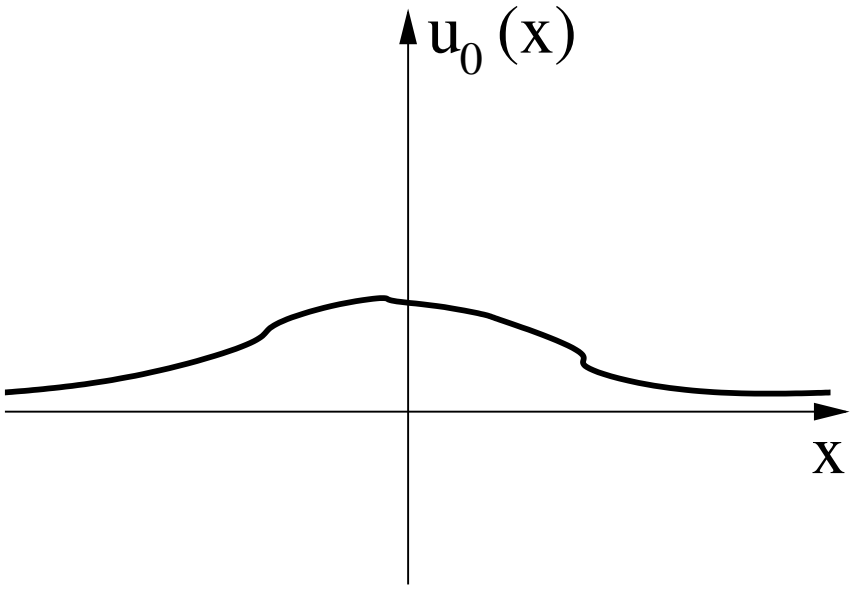

Figure 2.1: The initial condition $u = u_0(x)$ for equation (2.1)

along curves for which

$$\frac{dx}{dt} = c(u).$$

Such curves are called **characteristic curves**. How do we use this? Along such a characteristic curve C

$$\begin{aligned} & \frac{du}{dt} = 0 \\ \Rightarrow \quad & u = \text{constant along C} \\ \Rightarrow \quad & u = u_0(\xi), \end{aligned}$$

along the characteristic curve that starts at $x = \xi$ in the (x, t)-plane. What is the equation of this curve? Well, we have along this curve

$$\begin{aligned} & \frac{dx}{dt} = c(u(x,t)) \\ \Rightarrow \quad & \frac{dx}{dt} = c(u_0(\xi)) \\ \Rightarrow \quad & x = \xi + tc(u_0(\xi)). \end{aligned}$$

Thus, the solution of the equation (2.2) with initial condition $u(x, 0) = u_0(x)$ is given by

$$u = u_0(\xi),$$

where ξ is determined by the implicit equation

$$x = \xi + tc(u_0(\xi)).$$

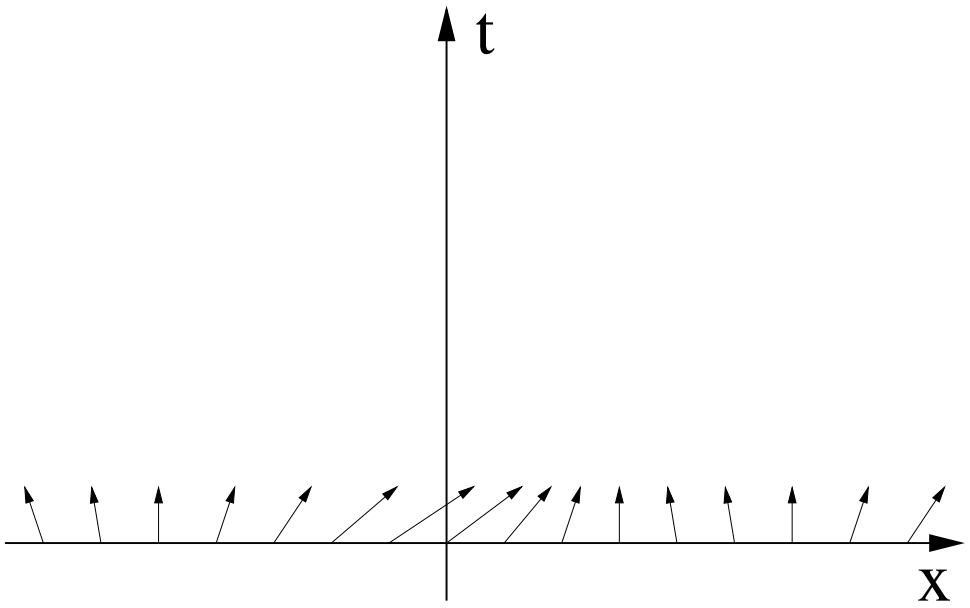

Figure 2.2: The initial condition $u(x,0) = u_0(x)$ being transported away from the x axis, in the directions of the arrows. (2.1)

This implicit equation should be solved for ξ, as a function of x and t.

Geometrically, it is clear what is happening, as is shown in Fig. 2.2. On every point of the x-axis, we can plot a vector with direction given by

$$\frac{dx}{dt} = c(u_0(\xi)),$$

which indicates how the initial condition is "transported" along the characteristic curves. For the equation we have considered, the characteristic curves are straight lines. This would not be the case if we allowed $c(u)$ to depend explicitly on x and/or t. For the straight-line case, there are different possibilities depending on $c(u)$ and $u_0(x)$:

- the characteristic lines could be parallel, or
- the characteristic lines could fan out, or
- the characteristic lines could cross.

Let's look at some examples.

Example: Consider the initial-value problem

$$\begin{cases} u_t + cu_x &= 0, \\ u(x,0) &= u_0(x), \end{cases}$$

where c is a constant. The characteristic curves are determined by

$$\frac{dx}{dt} = c \quad \Rightarrow \quad x = \xi + ct \quad \Rightarrow \quad \xi = x - ct.$$

Thus the characteristics are all straight lines with slope $1/c$ in the (x,t)-plane. Furthermore, $u(x,t)$ is constant along these characteristics. Thus, along characteristics,

$$u(x,t) = u_0(\xi) = u_0(x - ct),$$

which provides the explicit solution to our initial-value problem. Since the characteristics are all parallel, no problems arise and the solution exists for all time. From Fig. 2.3 and our calculations we see that the initial condition is merely transported at constant speed along the characteristics.

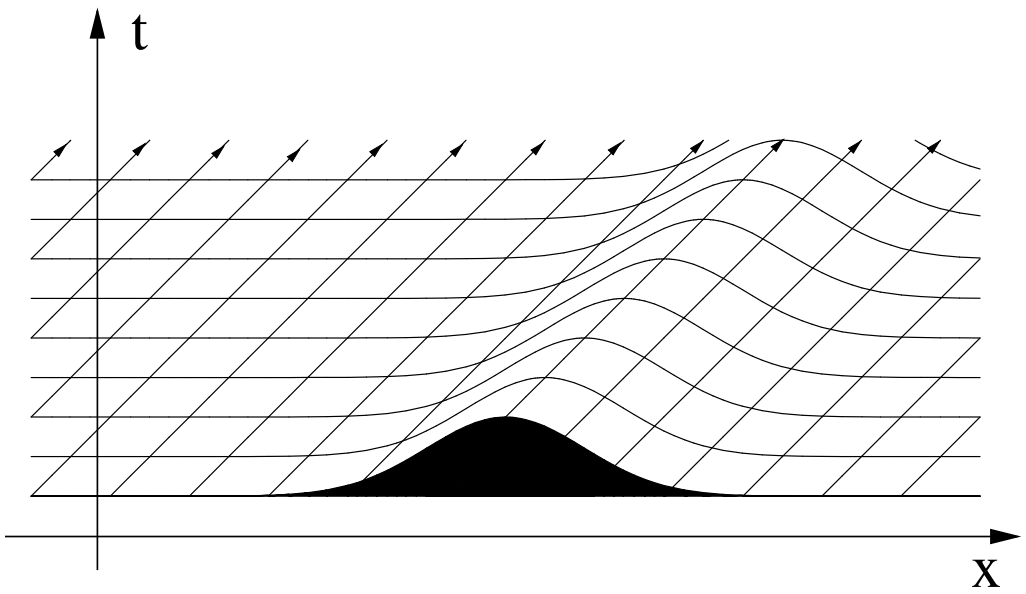

Figure 2.3: The characteristics for $u_t + cu_x = 0$, demonstrating the uniform translation of the initial condition over time. (2.1)

Example: A more complicated example is given by

$$\begin{cases} u_t + uu_x &= 0, \\ u(x,0) &= u_0(x) = \arctan x. \end{cases}$$

This partial differential equation is known as the (dissipationless) **Burgers equation**. It has played and still plays an important role as the simplest nonlinear partial differential equation, and as such is a laboratory for new techniques, numerical methods, *etc.*

Noting that $u(x,t) = u_0(\xi)$ is constant along characteristics, and that we have $c(u) = u$ for this equation, it follows that the characteristics are determined by

$$\begin{aligned} & \frac{dx}{dt} = u \\ \Rightarrow \quad & x = \xi + u_0(\xi)t \\ \Rightarrow \quad & x = \xi + t\,\arctan\xi \\ \Rightarrow \quad & t = \frac{x-\xi}{\arctan\xi}. \end{aligned}$$

These characteristic lines are fanning out, thus the solution spreads out. It is defined for all $t \geq 0$. It is given by

$$u(x,t) = u_0(\xi) = \arctan\xi,$$

where $\xi = \xi(x,t)$ is determined by the implicit relationship

$$x = \xi + t\,\arctan\xi.$$

In this case, we cannot write down an explicit solution, as we cannot solve $x = \xi + t\,\arctan\xi$ explicitly for ξ as a function of x and t. However, it is still straightforward to analyze the behavior of the solution. The solution at different times and the characteristics are shown in Fig. 2.4.

Example: Consider the system

$$\begin{cases} u_t + uu_x &= 0, \\ u(x,0) &= -\arctan x. \end{cases}$$

This is almost the same example, but now the sign of the initial condition is switched. Following what we did in the previous example, the solution is given by

$$u(x,t) = u_0(\xi) = -\arctan\xi,$$

where $\xi = \xi(x,t)$ is determined by the implicit relationship

$$x = \xi - t\,\arctan\xi.$$

As before, we cannot solve this explicitly for ξ, but we can analyze the behavior of the solution. The plot of the characteristics is shown in Fig. 2.5. We see there is a time t^* after which the solution obtained through the methods of characteristics is no longer valid: multiple characteristics cross, and in this wedge of overlapping characteristics, it is not possible to assign a unique value to $u(x,t)$.

We will return to this example soon, to resolve the issues we just encountered. When a situation with crossing characteristics is encountered, what happens for $t > t^*$, the time at

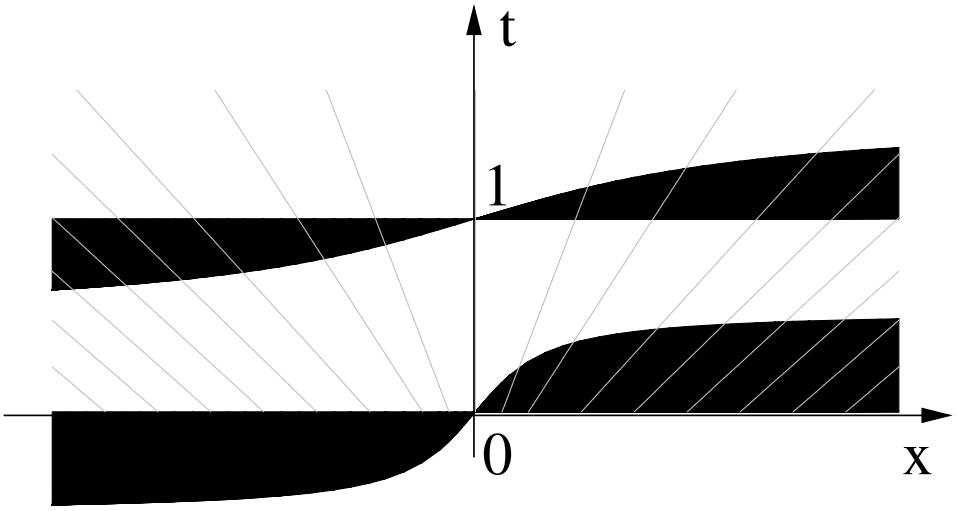

Figure 2.4: The characteristics for $u_t + uu_x = 0$, demonstrating the spreading of the initial condition $u_0 = \arctan x$ over time.

which they first cross? This is a question we'll look into later. For now, let's find a condition to determine the time of first crossing t^*. This time is known as the breaking time, as it is the time at which the wave first "breaks": the solution develops an infinite derivative in space x or time t (if one happens, the other one happens as well, because of the partial differential equation). We have

$$u_t = u_{0\xi}\,\xi_t, \quad u_x = u_{0\xi}\,\xi_x,$$

using the chain rule. These will become infinite when both of ξ_x and ξ_t become infinite, since u_0 is a well-behaved function. Taking x and t derivatives of the equation for the characteristics gives

$$\begin{aligned}
& x = \xi + tc(u_0(\xi)) \\
\Rightarrow \qquad & 1 = \xi_x + tc'(u_0(\xi))u_{0\xi}\xi_x \qquad \Rightarrow \qquad \xi_x = \frac{1}{1 + tc'(u_0(\xi))u_{0\xi}} \\
\Rightarrow \qquad & 0 = \xi_t + tc'(u_0(\xi))u_{0\xi}\xi_t + c(u_0(\xi)) \qquad \Rightarrow \qquad \xi_t = \frac{-c(u_0(\xi))}{1 + tc'(u_0(\xi))u_{0\xi}}.
\end{aligned}$$

These have the same denominators, so that the breaking time is determined by when these denominators are zero. Thus

$$t^* = -\frac{1}{c'(u_0(\xi^*))u_{0\xi}(\xi^*)},$$

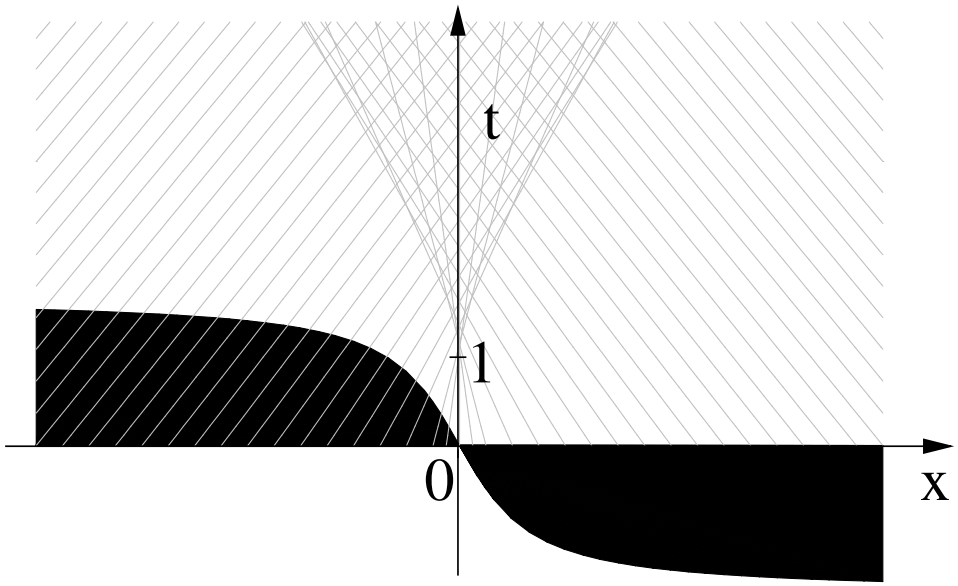

Figure 2.5: The crossing of the characteristics for $u_t + uu_x = 0$, demonstrating the steepening of the initial condition $u_0 = -\arctan x$ over time.

where ξ^* corresponds to those characteristics for which

$$c'(u_0)u_{0\xi} < 0,$$

(to ensure a positive breaking time), and

$$|c'(u_0)u_{0\xi}| \text{ is maximal,}$$

(to ensure t^* is the smallest time at which crossing occurs).

Example: Let's return to our previous example to try to determine the breaking time t^*. We have

$$\begin{aligned} c(u) = u \quad &\Rightarrow \quad c' = 1, \\ u_0 = -\arctan\xi \quad &\Rightarrow \quad u_{0\xi} = \frac{-1}{1+\xi^2}. \end{aligned}$$

Thus ξ^* is determined by

$$c'u_{0\xi} < 0 \quad \Rightarrow \quad \frac{-1}{1+\xi^2} < 0,$$

which is always true, and

$$\frac{1}{1+\xi^2} \text{ is maximal,}$$

which implies $\xi^* = 0$. Thus

$$t^* = -\frac{1}{c'(u_0(\xi^*))u_{0\xi}(\xi^*)} = -\frac{1}{-\frac{1}{1+(\xi^*)^2}} = 1 + (\xi^*)^2 = 1,$$

which agrees with Fig. 2.5.

2. Nonhomogeneous case

Now we're working with the equation

$$u_t + c(u, x, t)u_x = F(u, x, t).$$

Following what we did before, we get

$$\begin{cases} \dfrac{du}{dt} &= F(u, x, t), \\ \dfrac{dx}{dt} &= c(u, x, t). \end{cases}$$

This system of equations should be solved as seen fit, along the characteristic for which $x = \xi$ and $u = u_0$ at $t = 0$. After doing so, ξ is eliminated from this procedure, resulting in the final solution $u(x, t)$. Note that since now we're allowing c to depend explicitly on x and/or t, the characteristics are not necessarily straight lines.

Example: Consider the equation

$$u_t + \frac{x}{x+t}u_x = \frac{u+1}{x+t},$$

with

$$u(x, 0) = u_0(x) = x^2.$$

The characteristics are given by

$$\frac{dx}{dt} = \frac{x}{x+t},$$

and the solution along characteristics satisfies

$$\frac{du}{dt} = \frac{u+1}{x+t}.$$

From the first equation we get

$$\begin{aligned} & \frac{dt}{dx} = \frac{x+t}{x} = 1 + \frac{t}{x} \\ \Rightarrow \quad & \frac{dt}{dx} - \frac{1}{x}t = 1 \\ \Rightarrow \quad & \left(\frac{t}{x}\right)' = \frac{1}{x} \\ \Rightarrow \quad & \frac{t}{x} = \ln|x| + c. \end{aligned}$$

We want this to be the equation of the characteristic curve that passes through $x = \xi$ when $t = 0$, thus

$$0 = \ln|\xi| + c \quad \Rightarrow \quad c = -\ln|\xi|,$$

from which

$$t = x \ln\left|\frac{x}{\xi}\right|.$$

This is the equation for the characteristics. From Fig. 2.6, one sees that there will be problems with the solution straight away, as the characteristics have a singularity at $\xi = 0$, and multiple characteristics become tangent immediately, for any $t > 0$. We may still hope to construct a solution valid in other regions of the (x, t)-plane, if we don't necessarily think of this as an initial-value problem.

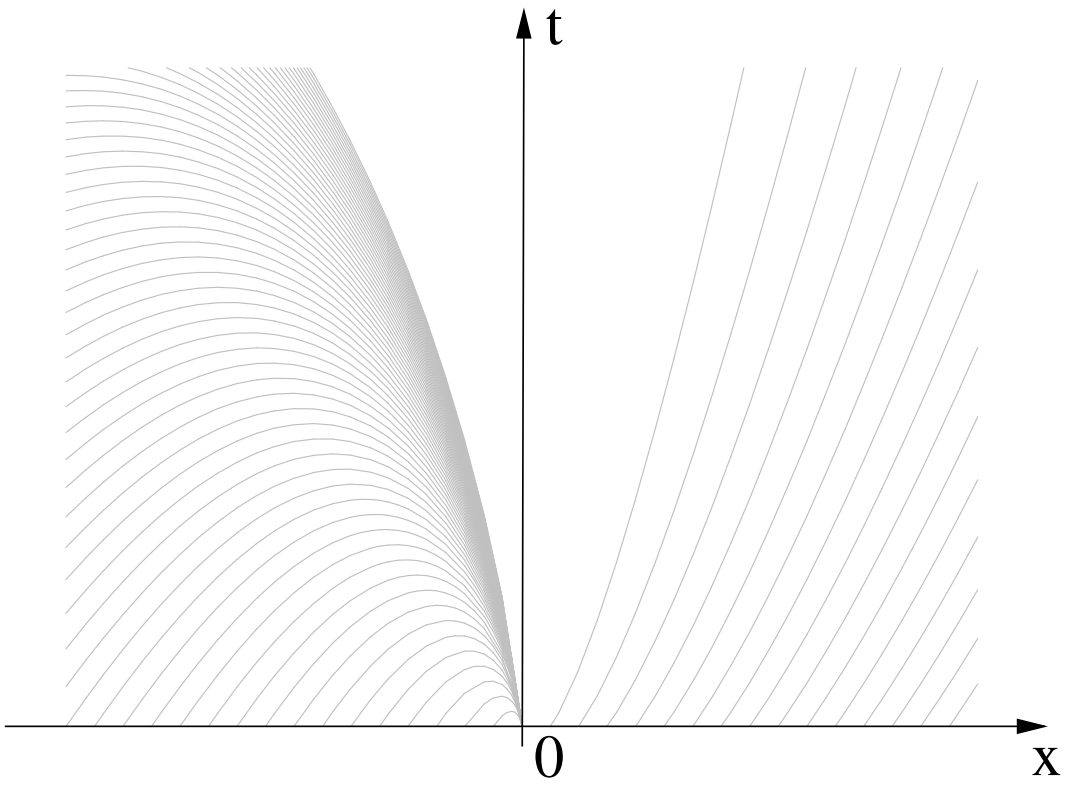

Figure 2.6: Some characteristics for the problem $u_t + u_x x/(x + t) = (u + 1)/(x + t)$.

The solution along these characteristics is determined by

$$\frac{du}{dt} = \frac{u+1}{x+t} = \frac{u+1}{x + x\ln|x/\xi|} = \frac{u+1}{x(1+\ln|x/\xi|)}$$

$$\Rightarrow \qquad \frac{du}{dx}\frac{dx}{dt} = \frac{u+1}{x(1+\ln|x/\xi|)}$$

$$
\begin{aligned}
\Rightarrow \quad & \frac{du}{dx}\frac{1}{dt/dx} = \frac{u+1}{x(1+\ln|x/\xi|)} \\
\Rightarrow \quad & \frac{du}{dx}\frac{1}{(1+t/x)} = \frac{u+1}{x(1+\ln|x/\xi|)} \\
\Rightarrow \quad & \frac{du}{dx}\frac{1}{1+\ln|x/\xi|} = \frac{u+1}{x(1+\ln|x/\xi|)} \\
\Rightarrow \quad & \frac{du}{dx} = \frac{u+1}{x} \\
\Rightarrow \quad & \frac{du}{dx} - \frac{1}{x}u = \frac{1}{x} \\
\Rightarrow \quad & \left(\frac{u}{x}\right)' = \frac{1}{x^2} \\
\Rightarrow \quad & \frac{u}{x} = -\frac{1}{x} + c \\
\Rightarrow \quad & u = -1 + cx.
\end{aligned}
$$

At $x = \xi$, we want $u = u_0$, so that

$$u_0 = -1 + c\xi \quad \Rightarrow \quad c = \frac{u_0+1}{\xi}.$$

This gives

$$
\begin{aligned}
& u = -1 + \frac{u_0+1}{\xi}x \\
\Rightarrow \quad & u = -1 + \frac{\xi^2+1}{\xi}x \\
\Rightarrow \quad & u = -1 + x\left(\xi + \frac{1}{\xi}\right).
\end{aligned}
$$

From the equation for the characteristics we get that

$$\frac{t}{x} = \ln\left|\frac{x}{\xi}\right| \quad \Rightarrow \quad e^{t/x} = \frac{x}{\xi} \quad \Rightarrow \quad \xi = xe^{-t/x}.$$

Note that we don't need absolute value signs anymore. The last equation implies that x and ξ have the same sign. Finally,

$$
\begin{aligned}
u &= -1 + x\left(xe^{-t/x} + \frac{1}{x}e^{t/x}\right) \\
&= -1 + x^2e^{-t/x} + e^{t/x}.
\end{aligned}
$$

Let me emphasize again that in general an explicit solution cannot be obtained using this procedure. We got lucky in the previous example, since we could solve the equation for the characteristics explicitly for ξ, which we could then use to eliminate ξ for the solution

formula for u. Even so, we can still obtain many important properties of the solution from this procedure[1]. In this particular case breaking starts immediately at $t = 0$, and a shock is formed.

2.2 What happens after shock formation?

Once a shock forms, it is clear that the equation

$$u_t + c(x, t, u)u_x = F(x, t, u)$$

is no longer valid. How do we modify it? This question has several answers, but none of them are determined by mathematics. At this point, the application that we are attempting to model needs to give us ideas to modify the model, so as to be able to come up with a unique solution. We'll see two ways to do this. Both may give analytical results. More importantly they are used in numerical methods to either avoid shock formation, or else to dictate the motion and location of shocks, once they form.

1. Introducing dissipation

The first method we'll see modifies the governing equation to prevent the formation of shocks. There are many ways of doing this, and typically one looks at the application being considered to figure out what the right way is to achieve this. Our equation is supposed to model some process associated with an application. When we get close to shock formation and derivatives get large, it is possible that other terms that we have previously omitted become important.

Example: We have found that the dissipationless Burgers equation

$$u_t + uu_x = 0$$

develops a shock if the initial condition has a profile as shown in Fig. 2.7. In general, any profile with negative derivative will lead to steepening. Now suppose that the problem we are modeling actually corresponds to the differential equation

$$u_t + uu_x = \epsilon u_{xx},$$

where $\epsilon > 0$ is very small. Because of the small coefficient ϵ we may have thought it was fine to ignore this second derivative term. This approximation will be valid, as long as derivatives in the solution are "small" (meaning "not large"). However, as we approach the shock, derivatives become very large and the term on the right does become significant.

Let's verify that this is indeed the case. Close to the shock time t^*, we have

$$u_x = u_0'(\xi)\xi_x = \frac{u_0'(\xi)}{1 + tu_0'(\xi)},$$

[1] and what else but this do we want an explicit solution formula for anyways?

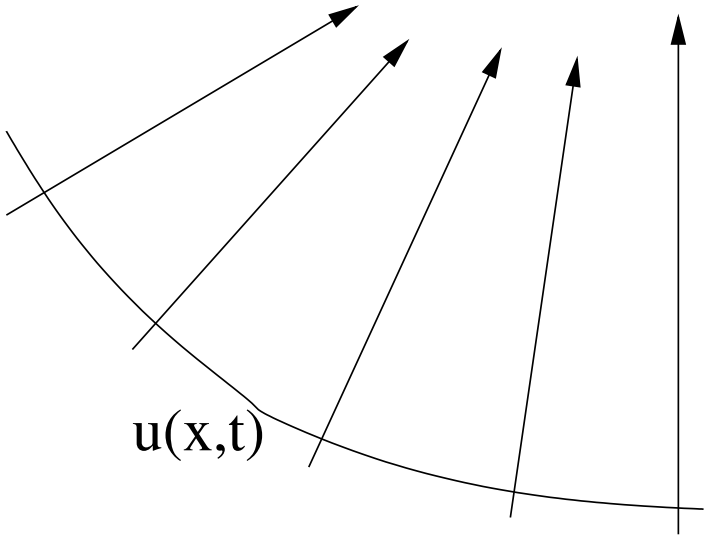

Figure 2.7: Part of an initial profile that will lead to shock formation in the dissipationless Burgers equation.

and the denominator approaches zero. Also,

$$u_{xx} = u_0''(\xi)\xi_x^2 + u_0'(\xi)\xi_{xx}.$$

To calculate ξ_{xx}, we return to the definition of the characteristic curve:

$$\begin{aligned}
& & x &= \xi + tu_0(\xi) \\
\Rightarrow & & 1 &= \xi_x + tu_0'(\xi)\xi_x \\
\Rightarrow & & 0 &= \xi_{xx} + tu_0'(\xi)\xi_{xx} + tu_0''(\xi)\xi_x^2 \\
\Rightarrow & & \xi_{xx}(1 + tu_0'(\xi)) &= -tu_0''(\xi)\xi_x^2 \\
\Rightarrow & & \xi_{xx} &= \frac{-tu_0''(\xi)\xi_x^2}{1 + tu_0'(\xi)}.
\end{aligned}$$

Using this, we obtain

$$\begin{aligned}
u_{xx} &= u_0''(\xi)\xi_x^2 + u_0'(\xi)\xi_{xx} \\
&= u_0''(\xi)\xi_x^2 - u_0'(\xi)\frac{tu_0''(\xi)\xi_x^2}{1 + tu_0'(\xi)} \\
&= u_0''(\xi)\xi_x^2 \left(1 - \frac{tu_0'}{1 + tu_0'}\right) \\
&= \frac{u_0''(\xi)\xi_x^2}{1 + tu_0'} \\
&= \frac{u_0''(\xi)}{(1 + tu_0')^3},
\end{aligned}$$

and we see that $u_{xx} \to \infty$ much faster than u_x or u_t. Thus, the right-hand side term cannot be ignored close to the shock formation.

Next, we'll show that taking this second-derivative term into account does indeed have the desired effect of arresting the shock formation. Near the shock we have

$$u_t = \epsilon u_{xx} - u u_x \sim \epsilon u_{xx},$$

Since u_{xx} is much larger than the product of uu_x just before the shock formation. Just before a shock is formed, the solution has a singular inflection point (*i.e.*, the second derivative changes sign, but by passing through infinity instead of zero). Above this inflection point $u_t \sim \epsilon u_{xx} < 0$, whereas below the inflection point $u_t \sim \epsilon u_{xx} > 0$. The combination of these is to counteract the steepening, as desired, see Fig. 2.8.

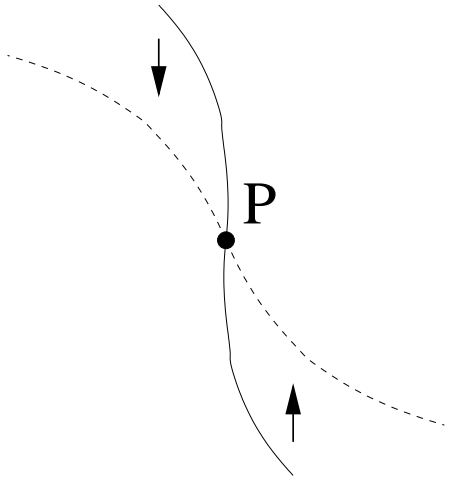

Figure 2.8: The effect of second-derivative dissipation on the near-formation of a shock near an inflection point P.

For other examples, and depending on the applications, different mechanisms may prevent shock formation. For instance for dispersive processes, the introduction of a third-derivative term may be more appropriate:

$$u_t + u u_x = \epsilon u_{xxx}.$$

The resulting partial differential equation is known as the Korteweg-deVries equation. It is one of the seminal equations of nonlinear wave theory and integrable systems, as treated in Amath573. Investigating the behavior of solutions as $\epsilon \to 0$ is nowhere near as straightforward as for the Burgers equation with dissipation. It is the topic of what is now known as Lax-Levermore theory, where one examines the zero-dispersion limit of wave equations with increasingly small dispersion. We will have more to say about dispersive equations in later chapters.

For yet different applications, one may have to resort to higher-order dissipation or dispersion, or different types of nonlinear terms.

2. Rankine-Hugoniot conditions

Another approach to the problem of shocks forming is to embrace them: instead of trying to prevent them, let's just figure out what to do with them once they form. Let's figure out where the shocks go, and how fast they move. The key to doing this is to work with an integral form of the partial differential equation. This integral form typically corresponds to the conservation law form of the application being modeled. Most partial differential equations arising in applications are originally derived in integral form, expressing the conservation of various quantities, such as mass, charge, momentum, energy, *etc.* This is true for Maxwell's equations in electro-magnetism, for the equations of continuum mechanics, including the Navier-Stokes equations.

The idea behind using an integral form of the partial differential equation is that integrals do not require continuous integrands to act on, whereas the arguments of derivatives should be smooth. Thus, using the integral form, we can deal with discontinuous solutions. Such solutions are also known as weak solutions of the differential equation.

Example: Again we consider the dissipationless Burgers equation

$$u_t + uu_x = 0.$$

We can rewrite this equation as

$$\frac{\partial}{\partial t}(u) + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = 0.$$

This form of the equation is called the conservation law form: integrating over the entire spatial domain, we find a conserved quantity for the partial differential equation if suitable boundary conditions are used. For instance

$$\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial t}(u) + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right)\right)dx = 0 \quad \Rightarrow \quad \frac{d}{dt}\int_{-\infty}^{\infty} u dx = 0,$$

where we've assumed that $u \to 0$ as $x \to \pm\infty$. Thus $\int_{-\infty}^{\infty} u dx$ is a constant of the motion.

Let's integrate the conservation law form over only part of space:

$$\int_{x_1}^{x_2} \frac{\partial}{\partial t}(u)dx + \frac{1}{2}u^2\Big|_{x=x_1}^{x=x_2} = 0$$

$$\Rightarrow \qquad \frac{d}{dt}\int_{x_1}^{x_2} u dx = \frac{1}{2}\left(u_1^2 - u_2^2\right),$$

where $u_1 = u(x_1, t)$ and $u_2 = u(x_2, t)$. As long as our solution has continuous derivatives, this integral equation, which is valid for all x_1 and x_2, is completely equivalent with the original partial differential equations form of the dissipationless Burgers equation.

Now suppose that the solution exhibits a shock at $x = s(t)$. Also, assume that $x_2 > s(t)$ and $x_1 < s(t)$. Lastly, let $u^- = \lim_{x<s(t)} u(x,t)$, and $u^+ = \lim_{x>s(t)} u(x,t)$. Thus u^- and u^+

are the values of u just past and in front of the shock. Then we get

$$\frac{d}{dt}\left(\int_{x_1}^{s(t)} u dx + \int_{s(t)}^{x_2} u dx\right) = \frac{1}{2}\left(u_1^2 - u_2^2\right)$$

$$\Rightarrow \qquad u^- s' - u^+ s' + \int_{x_1}^{s(t)} u_t dx + \int_{s(t)}^{x_2} u_t dx = \frac{1}{2}\left(u_1^2 - u_2^2\right).$$

Now we let $x_1 \to s$ and $x_2 \to s$. Then both of

$$\int_{x_1}^{s(t)} u_t dx \quad \text{and} \quad \int_{s(t)}^{x_2} u_t dx$$

vanish, as u_t is smooth both in front of and past the shock. We get

$$(u^- - u^+)s' = \frac{1}{2}\left(u^{-2} - u^{+2}\right)$$

$$\Rightarrow \qquad s' = \frac{1}{2}\frac{(u^- - u^+)(u^- + u^+)}{u^- - u^+}$$

$$\Rightarrow \qquad s' = \frac{1}{2}(u^- + u^+).$$

We find that the velocity of the shock is equal to the average of the solution amplitude in front of and past the shock.

How can we use this? For typical initial conditions the values of u^- and u^+ are not constant, and the shock velocity condition represents an ordinary differential equation for the shock position. However, we know at which time t^* and on which characteristic ξ^* the shock first forms. We may use our condition to move the shock approximately, using an Euler-like numerical method for instance.

Let's see how this works: at time t^*, a shock forms on characteristic ξ^*, from which we find the initial position $s(t^*)$ of the shock. Then we know the values of $u^- = u(s(t^*), t^*) = u^+$. Note that $u^-(t^*) = u^+(t^*)$ for smooth initial conditions. This is not the case if the initial conditions are discontinuous. Using this, we may propagate the shock with a small time step Δt:

$$s(t^* + \Delta t) = s(t^*) + \frac{1}{2}(u^- + u^+)\Delta t + \mathcal{O}((\Delta t)^2).$$

Next, we update the values of u^+ and u^-, using the values transported to the shock line along the new characteristics that now contribute to the shock. The introduction of the shock, splits the region where characteristics used to cross in two regions, one to the left of the shock line, one to the right. The characteristics can now safely be followed all the way to the shock line, without risking any multivaluedness. A schematic of this idea is presented in Fig. 2.9.

In general, for an equation of the form

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x} j = 0,$$

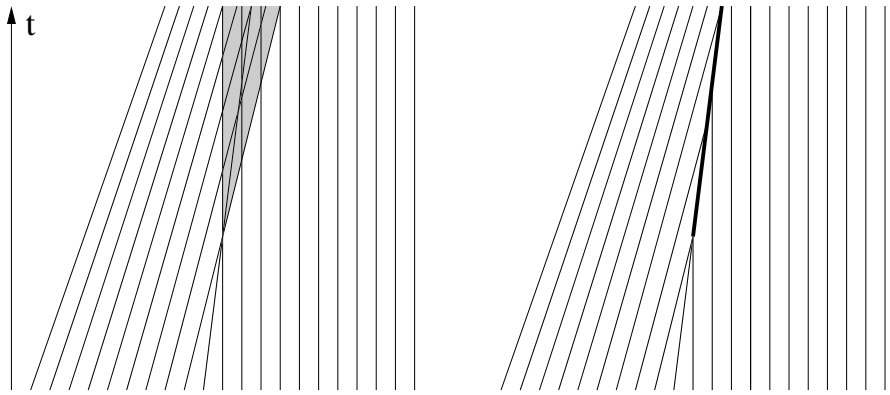

Figure 2.9: A shock region before and after the introduction of the shock. In the original shock region (left, filled in), the characteristics cross. After the shock is introduced (right, bold line), the characteristics are followed all the way to the shock.

the equation for the shock gives

$$\boxed{s' = \frac{\Delta j}{\Delta \rho},}$$

where $\Delta j = j^- - j^+$, and $\Delta \rho = \rho^- - \rho^+$. This follows a derivation entirely similar to the one for the dissipationless Burgers equation. This condition is known as the **Rankine-Hugoniot condition**.

A word of warning... Suppose we wish to introduce a shock in the characteristic picture for the dissipationless Burgers equation

$$u_t + uu_x = 0.$$

We could also, perversely, rewrite this equation as

$$uu_t + u^2 u_x = 0,$$

which has as its conservation law form the equation

$$\frac{\partial}{\partial t}\left(\frac{1}{2}u^2\right) + \frac{\partial}{\partial x}\left(\frac{1}{3}u^3\right) = 0.$$

Using the Rankine-Hugoniot condition with $\rho = u^2/2$ and $j = u^3/3$, we get

$$s' = \frac{2}{3}\frac{u^{-3} - u^{+3}}{u^{-2} - u^{+2}} = \frac{2}{3}\frac{u^{-2} + u^- u^+ + u^{+2}}{u^- + u^+}.$$

This is different from the result we have obtained before! So, when going from the differential form of the partial differential equation to the integral form (which is of course what we're

doing when we invoke the Rankine-Hugoniot condition), we might lose uniqueness. This is unsatisfying, but not unexpected. As previously announced, the application in question (physics, engineering, biology, *etc.*) will have to decide which integral formulation to use. The above calculation merely illustrates that the mathematics alone cannot tell us the right answer.

Let's summarize: typically an integral formulation is obtained first for many applications. Then one often switches to a differential formulation, because differential equations are easier to solve than integral equation. Or maybe we just have more experience with them. In any case, the limit process of going from an integral equation to a differential equation is straightforward and it results in a unique differential formulation. Then, when shocks are found in the differential equation, it is important to know which integral formulation the differential equation originated from, so as to be able to resolve the shock in a consistent manner.

2.3 Using characteristics to classify higher-order equations

In this section, we'll use our understanding of characteristics for first-order equations to classify equations of the second order. Specifically, we'll look at equations of the form

$$a(x,y)u_{xx} + 2b(x,y)u_{xy} + c(x,y)u_{yy} = f(x,y,u,u_x,u_y).$$

We've denoted the independent variables by x and y. One of them might be interpreted as a time-like variable t, depending on the results we'll find.

Using an invertible (at least locally) transformation, we will transform the above equation to one of the normal forms

1. $u_{xx} - u_{yy} = F$, or
2. $u_y = u_{xx} + F$, or
3. $u_{xx} + u_{yy} = F$,

where in all cases $F = F(x,y,u,u_x,u_y)$. In the first case, we'll call the equation hyperbolic. The independent variable x will get the interpretation of time t. The second case is called parabolic, and y plays the role of time t. The last case is that of an elliptic equation, and neither variable can be interpreted as time.

The classification we'll carry out is very similar to that of classifying conic sections in the plane. If we cared to, and had enough time and energy, we could deal with equations with a third independent variable z, which would be similar to classifying quadrics in three-dimensional space.

In order to achieve the classification above, we use a transformation on the independent variables

$$\begin{cases} \alpha &= \phi(x,y) \\ \beta &= \psi(x,y) \end{cases},$$

which we want to be invertible, at least locally. This implies that the Jacobian

$$J(x,y) = \det \begin{pmatrix} \dfrac{\partial \alpha}{\partial x} & \dfrac{\partial \alpha}{\partial y} \\ \\ \dfrac{\partial \beta}{\partial x} & \dfrac{\partial \beta}{\partial y} \end{pmatrix} \neq 0.$$

Using the chain rule, we see that the first derivatives are transformed according to

$$\begin{cases} u_x &= u_\alpha \alpha_x + u_\beta \beta_x \\ u_y &= u_\alpha \alpha_y + u_\beta \beta_y \end{cases}.$$

More chain ruling shows that the second derivatives transform as

$$\begin{cases} u_{xx} &= u_\alpha \alpha_{xx} + u_\beta \beta_{xx} + u_{\alpha\alpha}\alpha_x^2 + 2u_{\alpha\beta}\alpha_x\beta_x + u_{\beta\beta}\beta_x^2 \\ u_{yy} &= u_\alpha \alpha_{yy} + u_\beta \beta_{yy} + u_{\alpha\alpha}\alpha_y^2 + 2u_{\alpha\beta}\alpha_y\beta_y + u_{\beta\beta}\beta_y^2 \\ u_{xy} &= u_\alpha \alpha_{xy} + u_\beta \beta_{xy} + u_{\alpha\alpha}\alpha_x\alpha_y + u_{\beta\beta}\beta_x\beta_y + u_{\alpha\beta}(\alpha_x\beta_y + \alpha_y\beta_x) \end{cases}.$$

(Are you excited yet? We're just getting started!) We now use these expressions to transform the original second-order equation. We get

$$\begin{aligned} a(x,y)u_{xx} + 2b(x,y)u_{xy} + c(x,y)u_{yy} = \\ a(x,y)(u_\alpha \alpha_{xx} + u_\beta \beta_{xx} + u_{\alpha\alpha}\alpha_x^2 + 2u_{\alpha\beta}\alpha_x\beta_x + u_{\beta\beta}\beta_x^2) \\ +2b(x,y)(u_\alpha \alpha_{xy} + u_\beta \beta_{xy} + u_{\alpha\alpha}\alpha_x\alpha_y + u_{\beta\beta}\beta_x\beta_y + u_{\alpha\beta}(\alpha_x\beta_y + \alpha_y\beta_x)) \\ +c(x,y)(u_\alpha \alpha_{yy} + u_\beta \beta_{yy} + u_{\alpha\alpha}\alpha_y^2 + 2u_{\alpha\beta}\alpha_y\beta_y + u_{\beta\beta}\beta_y^2) = \\ f(x,y,u,u_\alpha\alpha_x + u_\beta\beta_x, u_\alpha\alpha_y + u_\beta\beta_y), \end{aligned}$$

So that

$$A(\alpha,\beta)u_{\alpha\alpha} + 2B(\alpha,\beta)u_{\alpha\beta} + C(\alpha,\beta)u_{\beta\beta} = F(\alpha,\beta,u,u_\alpha,u_\beta),$$

where

$$\begin{cases} A &= a\alpha_x^2 + 2b\alpha_x\alpha_y + c\alpha_y^2 \\ B &= a\alpha_x\beta_x + b(\alpha_x\beta_y + \alpha_y\beta_x) + c\alpha_y\beta_y \\ C &= a\beta_x^2 + 2b\beta_x\beta_y + c\beta_y^2 \end{cases},$$

and $F(\alpha,\beta,u,u_\alpha,u_\beta)$ contains all the other contributions, not explicitly written.

Example: Assume that $a = 0 = c$, so that the original equation reads

$$u_{xy} = \frac{f(x,y,u,u_x,u_y)}{2b(x,y)}.$$

We may assume that $b \neq 0$, since otherwise our equation is not of second order. Using the transformation

$$\begin{cases} x &= \alpha + \beta \\ y &= \alpha - \beta \end{cases},$$

(which is certainly invertible), the equation becomes (verify this)

$$u_{\alpha\alpha} - u_{\beta\beta} = F(\alpha, \beta, u, u_\alpha, u_\beta),$$

which is of the announced form of a hyperbolic equation.

The above example shows that we can assume that $(a, c) \neq (0, 0)$, since we have shown explicitly how to reduce that case to one of our normal forms. Since $a(x, y)$ and $c(x, y)$ appear symmetrically in the equation, let's pick $a(x, y) \neq 0$ in the region of (x, y)-space of interest (otherwise we switch the roles of x and y).

Because of the form of the equations for A and B given above, we examine the equation

$$a\varphi_x^2 + 2b\varphi_x\varphi_y + c\varphi_y^2 = 0.$$

If α and β can be chosen to satisfy this equation, then we have achieved $A = 0$ and $C = 0$, and the equation will have the standard form

$$u_{\alpha\beta} = \frac{F(\alpha, \beta, u, u_\alpha, u_\beta)}{2B(\alpha, \beta)}.$$

(using an invertible transformation, it is not possible to make B zero as well) The example above shows how this form is easily transformed to the standard form of a hyperbolic equation.

For starters, we divide by φ_y^2 to obtain

$$a\left(\frac{\varphi_x}{\varphi_y}\right)^2 + 2b\left(\frac{\varphi_x}{\varphi_y}\right) + c = 0,$$

and we see that all is determined by the quadratic equation

$$\boxed{az^2 + 2bz + c = 0} \quad \Rightarrow \quad a(z - z_1)(z - z_2) = 0,$$

where z_1 and z_2 are the roots of the quadratic.

1. Hyperbolic equations: $z_1 \neq z_2$, real

In this case

$$\begin{cases} \alpha_x &= z_1\alpha_y \\ \beta_x &= z_2\beta_y \end{cases}.$$

We can solve these equations using the method of characteristics. For α we have

$$\frac{d\alpha}{dx} = 0, \quad \frac{dy}{dx} = -z_1,$$

where the second equation defines the characteristic $y = y(x)$ passing through $y = y_0$ at $x = x_0$, and

$$\alpha = \alpha(x_0, y_0)$$

along this characteristic. Note that z_1 typically depends on x and y, since a, b and c do. We have similar equations for β. These equations for α and β are valid as long as x_0 and y_0 are in regions of the (x, y)-plane where z_1 and z_2 are real. Then in that region of the (x, y)-plane the equation is reduced to the standard form

$$2B(\alpha, \beta)u_{\alpha\beta} = F(\alpha, \beta, u, u_\alpha, u_\beta)$$

if the Jacobian condition holds:

$$\begin{aligned} & \det \begin{pmatrix} \alpha_x & \alpha_y \\ \beta_x & \beta_y \end{pmatrix} \neq 0 \\ \Rightarrow \quad & \det \begin{pmatrix} z_1\alpha_y & \alpha_y \\ z_2\beta_y & \beta_y \end{pmatrix} \neq 0 \\ \Rightarrow \quad & \alpha_y\beta_y(z_1 - z_2) \neq 0 \\ \Rightarrow \quad & \alpha_y\beta_y \neq 0, \end{aligned}$$

since $z_1 \neq z_2$ by assumption. Furthermore

$$\begin{aligned} B &= a\alpha_x\beta_x + b(\alpha_x\beta_y + \alpha_y\beta_x) + c\alpha_y\beta_y \\ &= az_1z_2\alpha_y\beta_y + b(z_1\alpha_y\beta_y + z_2\alpha_y\beta_y) + c\alpha_y\beta_y \\ &= \alpha_y\beta_y(az_1z_2 + b(z_1 + z_2) + c) \\ &= \alpha_y\beta_y\left(c - b\frac{2b}{a} + c\right) \\ &= 2\alpha_y\beta_y\frac{ac - b^2}{a} \\ &\neq 0. \end{aligned}$$

Thus, we can divide by B and the normal form of our equation becomes

$$u_{\alpha\beta} = \frac{F(\alpha, \beta, u, u_\alpha, u_\beta)}{2B(\alpha, \beta)},$$

valid in the region where $b^2 - ac > 0$. In this region the partial differential equation is called hyperbolic. We can think of the partial differential equation as transporting information along the characteristics

$$\begin{cases} \alpha_x &= z_1\alpha_y \\ \beta_x &= z_2\beta_y \end{cases}.$$

One consequence of this is that information contained in initial data for a hyperbolic problem travels at finite speed, dictated by the inverse slope of the characteristics. We'll have more to say about this when we study the **wave equation**

$$u_{tt} - c^2u_{xx} = 0,$$

which is the prototypical example of a second-order hyperbolic equation. Note that we have identified y with t.

2. Parabolic equations: $z_1 = z_2$, real

If $b^2 - ac \equiv 0$, then $z_1 = z_2 \in \mathbb{R}$. In this case the quadratic equation has only one distinct solution, and we can only define one characteristic. Clearly, we're not allowed to pick α and β both this way, as the resulting transformation $(x, y) \to (\alpha, \beta)$ would not be invertible. Choosing α to satisfy

$$\alpha_x = z_1 \alpha_y \quad \Rightarrow \quad \alpha_x = -\frac{b}{a}\alpha_y,$$

we may choose β to be anything that is linearly independent of our choice for α. Now we have

$$\begin{aligned} B &= a\alpha_x\beta_x + b(\alpha_x\beta_y + \alpha_y\beta_x) + c\alpha_y\beta_y \\ &= -b\alpha_y\beta_x + b\left(-\frac{b}{a}\alpha_y\beta_y + \alpha_y\beta_x\right) + c\alpha_y\beta_y \\ &= -\frac{b^2}{a}\alpha_y\beta_y + c\alpha_y\beta_y \\ &= \alpha_y\beta_y\frac{ac - b^2}{a} \\ &= 0, \end{aligned}$$

and our normal form is

$$u_{\beta\beta} = \frac{F(\alpha, \beta, u, u_\alpha, u_\beta)}{C(\alpha, \beta)}.$$

Equations that reduce to this form are called parabolic. The prototypical example of a parabolic equation is the **heat equation**

$$u_t = \sigma u_{xx},$$

with parameter $\sigma > 0$. Note that we have identified y with t. For the heat equation, the x-axis is a characteristic. This is often interpreted in the sense that initial data information for the heat equation propagates infinitely fast: any disturbance in $u(x, t_0)$ immediately impacts $u(x, t)$ for any $t > t_0$.

3. Elliptic equations: $z_2 = z_1^*$, complex

Lastly, we turn to the case where $b^2 - ac < 0$, and z_1 and z_2 are complex conjugate functions of x and y. Now, assuming that a, b and c are suitable analytic, the whole preceding calculation for the hyperbolic case may be repeated here. This leads to the normal form

$$u_{\alpha\beta} = G(\alpha, \beta, u, u_\alpha, u_\beta),$$

as before, valid whenever $b^2 - ac < 0$. However, now

$$\left\{ \begin{array}{rcl} \alpha_x &=& z_1\alpha_y \\ \beta_x &=& z_2\beta_y \end{array} \right. ,$$

and z_1 and z_2 are complex, leading to complex characteristics. To eliminate these complex quantities, we set

$$\left\{ \begin{array}{rcl} \alpha & = & \xi + i\eta \\ \beta & = & \xi - i\eta \end{array} \right. \quad \Rightarrow \quad \left\{ \begin{array}{rcl} \xi & = & (\alpha+\beta)/2 \\ \eta & = & (\alpha-\beta)/2i \end{array} \right. .$$

Note that these substitutions are consistent with our equations, since $z_1^* = z_2$. We get

$$\begin{aligned} u_{\alpha\beta} &= (u_\xi \xi_\alpha + u_\eta \eta_\alpha)_\beta \\ &= \frac{1}{2}\left(u_\xi - i u_\eta\right)_\beta \\ &= \frac{1}{2}\left(u_{\xi\xi}\xi_\beta + u_{\xi\eta}\eta_\beta - i u_{\eta\xi}\xi_\beta - i u_{\eta\eta}\eta_\beta\right) \\ &= \frac{1}{4}\left(u_{\xi\xi} + u_{\eta\eta}\right), \end{aligned}$$

which leads us to the standard form of an elliptic equation

$$u_{\xi\xi} + u_{\eta\eta} = 4G(\alpha, \beta, u, u_\alpha, u_\beta),$$

valid whenever $b^2 - ac < 0$. Since the characteristics are complex, we no longer have the notion of information being transported along them. A disturbance anywhere in the (ξ, η)-domain has an instantaneous impact on the entire domain. This is even worse than in the parabolic case! The prototypical equation of elliptic type is **Laplace's equation**

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

which we'll say more about later as well.

Example: Consider the equation

$$y u_{xx} - x u_{yy} = 0,$$

with $(x, y) \in \mathbb{R}^2$. Here we have

$$a = y, \quad c = -x, \quad b = 0.$$

Then

$$b^2 - ac = xy,$$

and the classification of the equation will be different in the four main quadrants. The boundaries of the different domains are determined by $b^2 - ac = 0$, which give the x and y axes, as shown in Fig. 2.10.

The hyperbolic region is the only region with (real) characteristics. They are determined by

$$\left\{ \begin{array}{rcl} \alpha_x & = & z_1 \alpha_y \\ \beta_x & = & z_2 \beta_y \end{array} \right. .$$

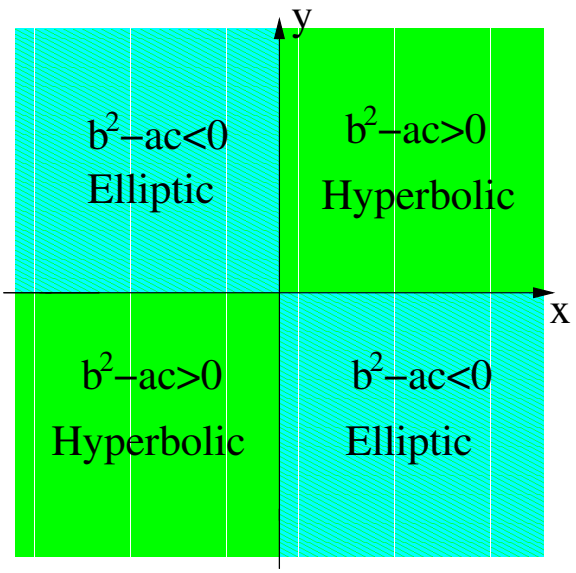

Figure 2.10: The different regions for the equation $yu_{xx} - xu_{yy} = 0$. It may be tempting to say that the equation is parabolic on the coordinate axis, but this is not correct. In order for partial derivatives to be defined at a point, we need an open region around that point. Thus the classification we've been going through only is valid in such open regions.

Here z_1 and z_2 are the solutions of

$$\begin{aligned} & az^2 + 2bz + c = 0 \\ \Rightarrow \quad & yz^2 - x = 0 \\ \Rightarrow \quad & z^2 = \frac{x}{y} \\ \Rightarrow \quad & z_1 = \sqrt{\frac{x}{y}}, \;\; z_2 = -\sqrt{\frac{x}{y}}, \end{aligned}$$

which are indeed real in the first and third quadrant.

- **First characteristic:** (in the first quadrant) We have

$$\begin{aligned} & \alpha_x = \sqrt{\frac{x}{y}}\alpha_y \\ \Rightarrow \quad & \frac{d\alpha}{dx} = 0, \quad \text{along} \quad \frac{dy}{dx} = -\sqrt{\frac{x}{y}} \\ \Rightarrow \quad & y^{1/2}dy = -x^{1/2}dx \\ \Rightarrow \quad & \frac{2}{3}y^{3/2} = -\frac{2}{3}x^{3/2} + c. \end{aligned}$$

Using the "initial condition" that $x = x_0$ implies $y = y_0$ gives

$$y = \left(x_0^{3/2} - x^{3/2} + y_0^{3/2} \right)^{2/3}.$$

This specifies a family of curves along which α is constant. On the other hand β will change along these curves. These curves play the role of coordinate lines for our new coordinates (α, β).

- **Second characteristic:** (in the first quadrant) A transverse (but not necessarily orthogonal) set of coordinate curves is given by the second set of characteristics. Repeating the same calculation with $z_2 = -\sqrt{x/y}$ for β gives

$$y = \left(x^{3/2} - x_0^{3/2} + y_0^{3/2} \right)^{2/3}.$$

Both families of characteristics are plotted in Fig. 2.11. A similar picture is valid for the third quadrant. Notice that the two families become tangent to each other on the original coordinate axes, reflecting the breakdown of the hyperbolic character of the partial differential equation.

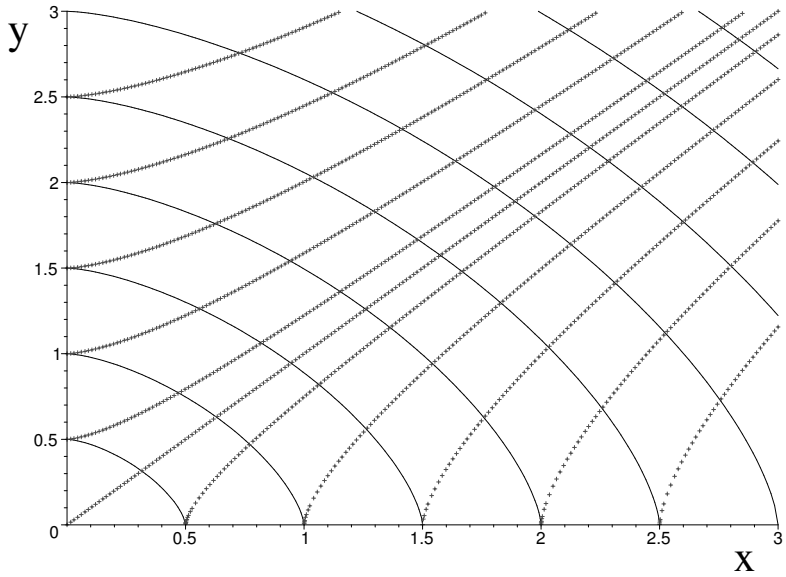

Figure 2.11: The two families of characteristic curves for $yu_{xx} - xu_{yy} = 0$. Curves of constant α are solid. Curves of constant β are displayed using fat dots.

4. A general classification

For completeness, we give some standard terminology here. Consider the second-order partial differential equation

$$Lu = F(\boldsymbol{x}, u, \boldsymbol{\nabla u}),$$

with

$$Lu = \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j},$$

and $\boldsymbol{x} = (x_1, \ldots, x_n)$.

- This partial differential equation is called **elliptic** in a certain region of space if its coefficient matrix $(a_{ij})_{i,j=1}^n$ is positive definite. Then the linear differential operator L is called elliptic as well.
- The partial differential equation

  $$u_t = Lu + F$$

  is called **parabolic** if the operator L is elliptic.
- The partial differential equation

  $$u_{tt} = Lu + F$$

  is called **hyperbolic** if the operator L is elliptic.

It should be stated that sometimes other conventions are used, for instance in such areas as relativity or field theory. In all cases, the classification is valid locally, in some open neighborhood of a point.

2.4 Exercises

1. Consider the initial-value problem

   $$\begin{cases} u_t + tuu_x = 0, \\ u(x,0) = \sin(x), x \in \mathbb{R}. \end{cases}$$

   Solve this initial-value problem for $t \geq 0$. Draw the characteristics. For how long is the solution valid?

2. Consider the initial-value problem

   $$\begin{cases} u_t + tuu_x = 0, \\ u(x,0) = \cos(x), x \in \mathbb{R}. \end{cases}$$

   Solve this initial-value problem for $t \geq 0$. Draw the characteristics. For how long is the solution valid?

3. Consider the initial-value problem

$$\begin{cases} u_t + 2xtu_x = 1, \\ u(x,0) = e^x, x \in \mathbb{R}. \end{cases}$$

Solve this initial-value problem for $t \geq 0$. Draw the characteristics. For how long is the solution valid?

4. Consider the initial-value problem

$$\begin{cases} u_t + xu_x = 1, \\ u(x,0) = e^x, x \in \mathbb{R}. \end{cases}$$

Solve this initial-value problem for $t \geq 0$. Draw the characteristics. For how long is the solution valid?

5. Consider the initial-value problem

$$\begin{cases} u_t + uu_x = 0, \\ u(x,0) = u_0(x), x \in \mathbb{R}. \end{cases}$$

Solve this initial-value problem for $t \geq 0$ with (a) $u_0(x) = -x$, (b) $u_0(x) = 1 - x^2$, (c) $u_0(x) = \sin x$. For each case, draw the characteristics. For how long is the solution valid?

6. The method of characteristics is not restricted to initial-value problems. (a) Consider the problem

$$\begin{cases} xu_x + yu_y = u + 1, \\ u(x,x^2) = x^2, (x,y) \in \mathbb{R}^2. \end{cases}$$

This is Example 2 from Section 2.1 of Guenther and Lee. Use our version of the method of characteristics to solve this problem. You can use either x or y as your "time-like" variable. Where is your solution defined? (b) Now consider the same problem, but with $u(x,x) = x^2$, instead of $u(x,x^2) = x^2$. Show that this problem cannot be solved using the method of characteristics.

Explanation: This is because the data from which you are to recover the solution is given on a characteristic curve (check this). Since the solution is transported along characteristic curves, giving data along a certain characteristic curve does not allow one to construct the solution along the other characteristics. In general, the method of characteristics will work when the data is given along a curve that crosses the characteristics.

7. We have seen an example where the breaking time t^* depends on the initial data. Now consider the initial-value problem

$$\begin{cases} (t-\alpha)u_t + xu_x = u, \\ u(x,0) = f(x), x \in \mathbb{R}, \alpha > 0. \end{cases}$$

Solve this initial-value problem for $t \geq 0$, *i.e.*, give an explicit solution for $u(x,t)$ in terms of the initial data $f(x)$. Draw the characteristics. For how long is the solution valid? As you see, in this problem the breaking time t^* is independent of the initial data $f(x)$.

8. The method of characteristics is easily extended to multidimensional problems. Consider the initial-value problem

$$\begin{cases} u_t + c_1(u)u_x + c_2(u)u_y = 0, \\ u(x,y,0) = u_0(x,y), (x,y) \in \mathbb{R}^2. \end{cases}$$

   - Generalize the method of characteristics to such two-dimensional equations: write down the equations for the characteristics and solve them if possible. Write down the exact solution for $u(x,y,t)$, in explicit form if possible, in implicit form otherwise.
   - Apply the above method to solve the two-dimensional transport equation $u_t + c_1 u_x + c_2 u_y = 0$, with initial condition $u(x,y,0) = u_0(x,y)$. Find the explicit solution for $u(x,y,t)$.
   - Use a slight variation of your method to solve the damped transport initial-value problem

$$\begin{cases} u_t + c_1 u_x + c_2 u_y + \alpha u = 0, \quad \alpha > 0 \\ u(x,y,0) = u_0(x,y), (x,y) \in \mathbb{R}^2. \end{cases}$$

9. Consider the IVP

$$\begin{cases} u_t + uu_x = 0, \\ u(x,0) = \phi(x), \end{cases}$$

   where $\phi(x)$ is determined by

$$\phi(x) = \begin{cases} 1, & x \leq 0 \\ 1-x, & 0 < x < 1 \\ 0, & x \geq 1. \end{cases}$$

   (a) Show that a shock forms at $t_* = 1$. (b) You know that different integral representations of the differential equation can lead to different shock dynamics. By multiplying the differential equation by u^α, $\alpha \in (-1, +\infty)$, find different shock positions for any $t > 1$, depending on α. (c) Show that by a suitable choice of α, the shock can be continued along any of the characteristics starting in $x_0 \in (0,1)$. (d) For any choice of α, draw the solution $u(x,t)$ at $t = 0$, $t = 1/2$, $t = 1$, $t = 2$, and $t = 5$.

10. Consider the IVP

$$\begin{cases} u_t + uu_x = 0, \\ u(x,0) = \phi(x), \end{cases}$$

where $\phi(x)$ is defined by

$$\phi(x) = \begin{cases} 1, & x < -1 \\ x^2, & -1 < x < 0 \\ 0, & 0 < x \end{cases}$$

Determine the solution of this problem in all regions of $x \in \mathbb{R}$ and $t > 0$, using the integral representation.

$$\frac{d}{dt}\int_a^b u dx + \frac{1}{2}u^2\Big|_a^b = 0.$$

Are there any shocks? If so, where and when do they occur? Also draw all relevant phases the solution goes through, and provide a picture of the characteristics. If there are shocks, include a picture of the characteristics without and with the shock lines.

11. **Shock waves.** It is not often that shock conditions like the ones we derived in class can be solved as a differential equation for the shock position, hence determining the shock position $s(t)$ at any time after its formation. But sometimes, this is possible.

    Consider the following initial-value problem:

$$u_t + u^2 u_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 1 & \text{for } x \le 2, \\ 0 & \text{for } x > 2. \end{cases}$$

    Solve this initial-value problem in the region of the (x,t)-plane ($t \ge 0$) where no shock develops. Next, assuming that

$$\frac{d}{dt}\int_{x_1}^{x_2} u dx + \int_{x_1}^{x_2} u^2 u_x dx = 0$$

    is the correct integral formulation to use, derive the shock condition governing the motion of the shock position $x = s(t)$. In this case, this condition can be solved for the shock position. Proceed to solve it, and hence find the position of the shock for any time after its formation. Give the solution to the initial-value problem, for any time $t \ge 0$. Illustrate this with some appropriately chosen snapshots of the solution $u(x,t)$.

12. **Shock waves.** Consider the following initial-value problem:

$$u_t + u u_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 1+x & \text{for } x \in (-1,0), \\ 1-x & \text{for } x \in (0,1), \\ 0 & \text{for } |x| > 1. \end{cases}$$

    Solve this initial-value problem in the region of the (x,t)-plane ($t \ge 0$) where no shock develops. Next, assuming that

$$\frac{d}{dt}\int_{x_1}^{x_2} u dx + \int_{x_1}^{x_2} u u_x dx = 0$$

is the correct integral formulation to use, derive the shock condition governing the motion of the shock position $x = s(t)$. In this case, this condition can be solved for the shock position. Proceed to solve it, and hence find the position of the shock for any time after its formation. Give the solution to the initial-value problem, for any time $t \geq 0$. Illustrate this with some appropriately chosen snapshots of the solution $u(x, t)$.

13. **Rarefaction waves.** In class we considered several examples with characteristics. In many cases, each point of the (x, t)-plane had exactly one characteristic going through it. In other case, the characteristic crossed, and more than one characteristic could pass through a given point. There are other possibilities: it is possible that some regions of the plane have no characteristics in them.

    (a) Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x, 0) = \begin{cases} 0 & \text{for } x \leq 0, \\ 1 & \text{for } x > 0. \end{cases}$$

    Graph the characteristics for this problem in the (x, t)-plane ($t \geq 0$) and show that there are no characteristics in a whole sector of the plane.

    (b) As with shocks, there are several ways to deal with this problem. We'll examine one here: consider the same differential equation as above, but with initial condition $u_\epsilon(x, 0) = 1/2 + \tanh(x/\epsilon)/2$, with $\epsilon > 0$. Show that in this case, there is exactly one characteristic in every area of the (x, t)-plane ($t \geq 0$) (a graphical "proof" is fine). What happens to the initial condition $u_\epsilon(x, 0)$ as $\epsilon \to 0$? What happens to the characteristics as $\epsilon \to 0$?

    (c) We now use the limit characteristics found above to solve the initial-value problem. It is clear what to do on the original characteristics (*i.e.*, the ones we had before we talked about this limit process), as we can propagate the initial condition along them. What do we do along the characteristics in the sector? The solution should be constant along them, but we don't know what initial condition to propate along them. In order to be constant along these characteristics, the solution $u(x, t)$ should be of the form $u(x, t) = g(x/t)$. Substitute this ansatz in the partial differential equation to find an equation for $g(z)$, $z = x/t$ (ignore the possibility $g' = 0$, which cannot result in a continuous solution. It can be shown that this choice can indeed be ignored).

    (d) Finally, write down the solution to the original initial-value problem obtained this way and plot it for various instances of time. This solution satisfies the original partial differential equation at all time, except at the corner points emanating from the original discontinuity.

14. **Rarefaction waves: another example.** Use the same process as in the previous problem to solve the initial-value problem

$$u_t + u^2 u_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x, 0) = \begin{cases} 0 & \text{for } x \leq 0, \\ 1 & \text{for } x > 0. \end{cases}$$

15. **Rarefaction waves and shock waves combined.** Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 0 & \text{for } x \leq 0, \\ 1 & \text{for } 0 < x < 1, \\ 0 & \text{for } x \geq 1. \end{cases}$$

(a) Find the solution of this problem in all areas of the (x,t) $(t \geq 0)$-plane where there is exactly one characteristic through each point.

(b) Using the same process as before, fill in the rarefaction region with characteristic lines, and write down the solution to the initial-value problem in this region.

(c) Assuming that the integral formulation

$$\frac{d}{dt}\int_{x_1}^{x_2} u dx + \int_{x_1}^{x_2} \left(\frac{1}{2}u^2\right)_x dx = 0$$

is the correct integral formulation corresponding to the partial differential equation, derive the shock speed condition for a shock moving on a path $x = s(t)$.

(d) Use this shock condition to find the location of the shock at any time $t > 0$. Note that there are two different regions for this.

(e) Knowing the path of the shock, give a complete plot of the characteristics in the (x,t)-plane $(t \geq 0)$, and describe the complete solution to the initial-value problem. Also provide snapshots of the solution $u(x,t)$ at relevant times.

16. Consider the IVP

$$\begin{cases} u_t + uu_x = 0, \\ u(x,0) = \phi(x), \end{cases}$$

where $\phi(x)$ is defined by

$$\phi(x) = \begin{cases} 0, & |x| > 1 \\ 1 - x^2, & |x| < 1 \end{cases}$$

Determine the solution of this problem in all regions of $x \in \mathbb{R}$ and $t > 0$, using the integral representation.

$$\frac{d}{dt}\int_a^b u dx + \frac{1}{2}u^2\Big|_a^b = 0.$$

Are there any shocks and rarefaction regions? If so, where and when do they occur? Resolve all of them. Also draw all relevant phases the solution goes through, and provide a picture of the characteristics. If there are shocks, include a picture of the characteristics without and with the shock lines.

17. **Rarefaction waves and shock waves combined: good times were had by all.** Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 0 & \text{for } x < 0, \text{ or } x > 3 \\ 3 & \text{for } 0 < x < 1, \\ 2 & \text{for } 1 < x < 2, \\ 1 & \text{for } 2 < x < 3. \end{cases}$$

Using rarefaction waves and the integral formulation you used in the previous problem, solve this problem for all $t > 0$.

18. **Rarefaction waves and shock waves combined: more good times.** Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 0 & \text{for } x < -3, \text{ or } x > 2 \\ 4 & \text{for } -3 < x < -2, \\ 1 & \text{for } -2 < x < 1, \\ 3 & \text{for } 1 < x < 2. \end{cases}$$

Using rarefaction waves and the integral formulation with density u and flux $u^2/2$, solve this problem for all $t > 0$.

19. **Rarefaction waves and shock waves combined: good times continued.** Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} 1 & \text{for } 1 < |x| < 2, \\ 4 & \text{for } |x| < 1, \\ 0 & \text{for } |x| > 2. \end{cases}$$

Using rarefaction waves and the integral formulation with density u and flux $u^2/2$, solve this problem for all $t > 0$.

20. **Rarefaction waves and shock waves combined: only try this problem if you're VERY bored.** In what follows $\beta > \alpha > 0$. Consider the following initial-value problem:

$$u_t + uu_x = 0, \quad x \in \mathbb{R}, \quad \text{and} \quad u(x,0) = \begin{cases} -\alpha x & \text{for } x \in (-1,0), \\ -\beta x & \text{for } x \in (0,1), \\ 0 & \text{for } |x| > 1. \end{cases}$$

Using rarefaction waves and the integral formulation with density u and flux $u^2/2$, solve this problem for all $t > 0$.

21. Consider the nonlinear partial differential equation

$$\phi_{tt}\phi_x^2 - 2\phi_{xt}\phi_x\phi_t + \phi_t^2\phi_{xx} = 0.$$

This equation is a special case of the so-called Monge-Ampère equation. Show that in terms of the auxiliary function $u = \phi_t/\phi_x$ this equation may be reduced to a first-order equation. Assuming initial conditions $\phi(x,0) = 1 + 2e^{3x}$ and $\phi_t(x,0) = 4e^{3x}$, find $\phi(x,t)$ for $x \in \mathbb{R}$ and $t \geq 0$.

Chapter 3

Fourier series and Fourier transforms

I am sure you are all familiar with aspects of Fourier analysis from previous courses. The main result I want to achieve here are a review of some of the main theorems, this time around with rigorous statements and proofs. Then later, I want to illustrate why this level of rigor can matter in an application.

3.1 Fourier series

Let's start with Fourier series. Suppose $f(x)$ can be expanded in a Fourier series:

$$f(x) = A + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right),$$

for $x \in [-L, L]$. Here $\{a_n, n = 1, \ldots, \infty\}$, $b_n, n = 1, \ldots, \infty$ are the sequences of Fourier expansion coefficients. Given a function $f(x)$, the first question we should answer is that of how we find the coefficients A, a_n and b_n, $n = 1, 2, \ldots$.

Let's do some formal calculations to see what the right answer should be.

First, note that the average value of all the terms on the right-hand side is 0, except for the one constant term A. Thus A is the average value of the function $f(x)$:

$$A = \frac{1}{2L} \int_{-L}^{L} f(x) dx.$$

The other coefficients are computed in a similar manner: let's multiply the equation by $\cos(m\pi x/L)$, or by $\sin(m\pi x/L)$. Then only one of the terms on the right-hand side has non-zero average. This is easily seen from the following **orthogonality conditions** (check

these!):

$$\int_{-L}^{L} \sin\frac{n\pi x}{L}\cos\frac{m\pi x}{L}dx = 0,$$
$$\int_{-L}^{L} \sin\frac{n\pi x}{L}\sin\frac{m\pi x}{L}dx = L\delta_{nm},$$
$$\int_{-L}^{L} \cos\frac{n\pi x}{L}\cos\frac{m\pi x}{L}dx = L\delta_{nm},$$

for all n, m integer. Here δ_{nm} is the Krönecker delta:

$$\delta_{nm} = \begin{cases} 0 & \text{if } n \neq m, \\ 1 & \text{if } n = m. \end{cases}$$

Using these relations on our expansion multiplied by either $\cos(m\pi x/L)$ or $\sin(m\pi x/L)$ gives

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi x}{L}dx$$
$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi x}{L}dx$$

for $n = 1, 2, \ldots$. Note that the first formula with $n = 0$ gives $A = a_0/2$.

Given a function $f(x)$, periodic on $[-L, L]$, and a series

$$S(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos\frac{n\pi x}{L} + b_n\sin\frac{n\pi x}{L}\right),$$

where $a_0, a_1, \ldots$, $b_1, b_2, \ldots$ are given by the preceding formulas, our main question for now will be to figure out under what conditions $S(x) = f(x)$. When these are not equal, what can be said?

1. The Riemann-Lebesgue lemma

The formulas for the Fourier coefficients involve definite integrals of a function multiplied by a trigonometric function. The frequency of this trigonometric function $\to \infty$ as $n \to \infty$. What can we say about the behavior of the Fourier coefficients for large n? This question is answered (at least to leading order) by the Riemann-Lebesgue lemma. This lemma is one of the truly outstanding results of advanced calculus, and we'll treat it here in its own right. We'll use it later on many occasions.

Lemma 1 (Riemann-Lebesgue) *Let $g(x)$ be continuous (except at at most a finite number of points) in $[a, b]$, and let*

$$\int_a^b |g(x)|dx < \infty.$$

Then

$$\lim_{\lambda\to\pm\infty}\int_a^b g(x)\left\{\begin{array}{c}\cos\lambda x\\ \sin\lambda x\end{array}\right. dx = 0.$$

Since $g(x)$ is integrable, the integral is bounded:

$$\left|\int_a^b g(x)\cos\lambda x dx\right| \le \int_a^b |g(x)||\cos\lambda x| dx \le \int_a^b |g(x)| dx < \infty,$$

and similar for the sine. The lemma states that this bounded integral approaches zero as $\lambda \to \infty$.

I'll give you three proofs, only one of which really deserves to be called a proof. But the first two provide some intuition of what is going on, and why the Riemann-Lebesgue lemma has to be true.

First proof: graphical. Since the integrand is a continuous function (otherwise we break the problem up in smaller integrands), we have a graph as shown in Fig. 3.1: the graph of a continuous function $g(x)$ is multiplied by a trigonometric function of high frequency. This trig function, for large enough λ, oscillates on a scale that is much faster than the scale on which $g(x)$ changes. As a consequence, the figure has the appearance of a slowly modulated trig function, where $g(x)$ plays the role of the modulation. The definite integral that we are interested in is the area between this rapidly oscillating function and the horizontal axis. For λ sufficiently large, any one of the positive areas is almost exactly canceled out by one of the neighboring negative areas. This heuristic statement becomes more accurate as $n \to \infty$, and it seems reasonable to believe the Riemann-Lebesgue lemma.

A numerical evaluation of

$$I(\lambda) = \int_0^2 g(x)\cos\lambda x dx$$

is shown in Fig. 3.2.

Second proof: handwaving integration by parts. Let's assume that $g(x)$ is differentiable. Then

$$\begin{aligned}\lim_{\lambda\to\pm\infty}\int_a^b g(x)\cos\lambda x dx &= \lim_{\lambda\to\pm\infty}\left(g(x)\frac{\sin\lambda x}{\lambda}\bigg|_a^b - \frac{1}{\lambda}\int_a^b g'(x)\sin\lambda x dx\right)\\ &= -\lim_{\lambda\to\pm\infty}\frac{1}{\lambda}\int_a^b g'(x)\sin\lambda x dx.\end{aligned}$$

Note that we've assumed that $\lim_{x\to a,b} g(x)$ are finite.

We see that integration by parts buys us one factor of λ in the denominator, with an integral of the same type. If in addition we assume that $\int_a^b |g'(x)|dx < \infty$, then we may conclude that $\int_a^b g(x)\cos\lambda x dx \to 0$, at a rate of λ^{-1}.

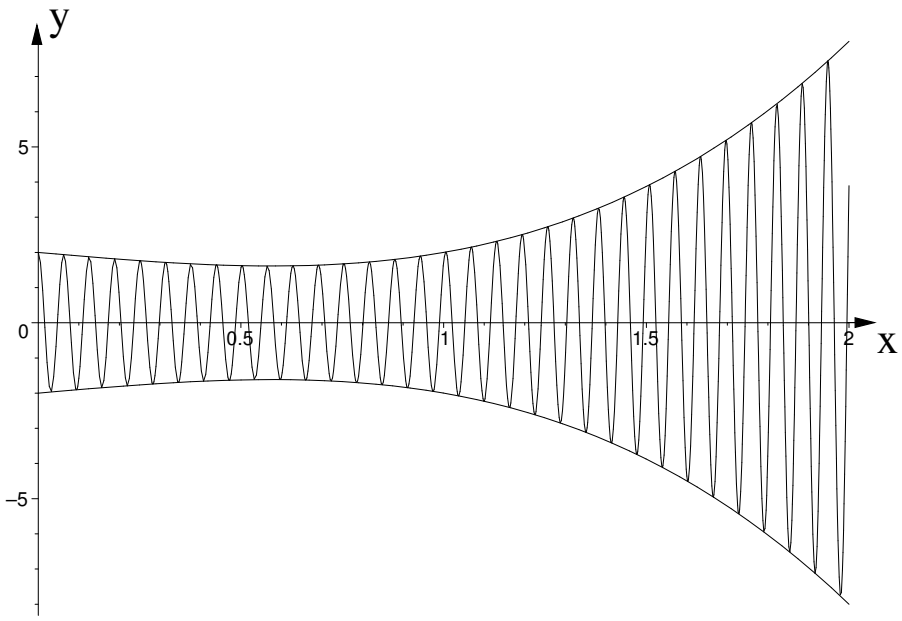

Figure 3.1: The function $y = g(x)\cos\lambda x$, with $g(x) = x^3 - x + 2$ and $\lambda = 100$. Also shown are $g(x)$ and $-g(x)$, which play the role of modulating amplitudes.

If $g(x)$ is twice differentiable, $\lim_{x\to a,b} g'(x)$ are finite, and $\int_a^b |g'(x)|dx < \infty$, then we can repeat this game, to find that $\int_a^b g(x)\cos\lambda x dx \to 0$, at a rate of λ^{-2}.

In general, if $g(x)$ is n times differentiable, suitable limits of various derivatives at the endpoints are finite, and all derivatives are integrable over $[a, b]$, then we conclude that $\int_a^b g(x)\cos\lambda x dx \to 0$, at a rate of λ^{-n}. Specifically, if $g(x)$ is analytic on $[a, b]$, $\int_a^b g(x)\cos\lambda x dx \to 0$ at a rate faster than any power of λ. This hints at the convergence being exponentially fast!

Third proof: and now for real! Let's look at

$$I = \int_a^b g(x)\sin\lambda x dx.$$

Let's assume that $g(x)$ is continuous. If it's not, we break the integral up in smaller parts, on which $g(x)$ is continuous. That way, the only discontinuities we have to worry about are at the endpoints. We'll investigate these later.

Let $x = t + \pi/\lambda$. Then $t = x - \pi/\lambda$, and

$$I = \int_{a-\pi/\lambda}^{b-\pi/\lambda} g\left(t + \frac{\pi}{\lambda}\right)\sin(\lambda t + \pi)dt$$

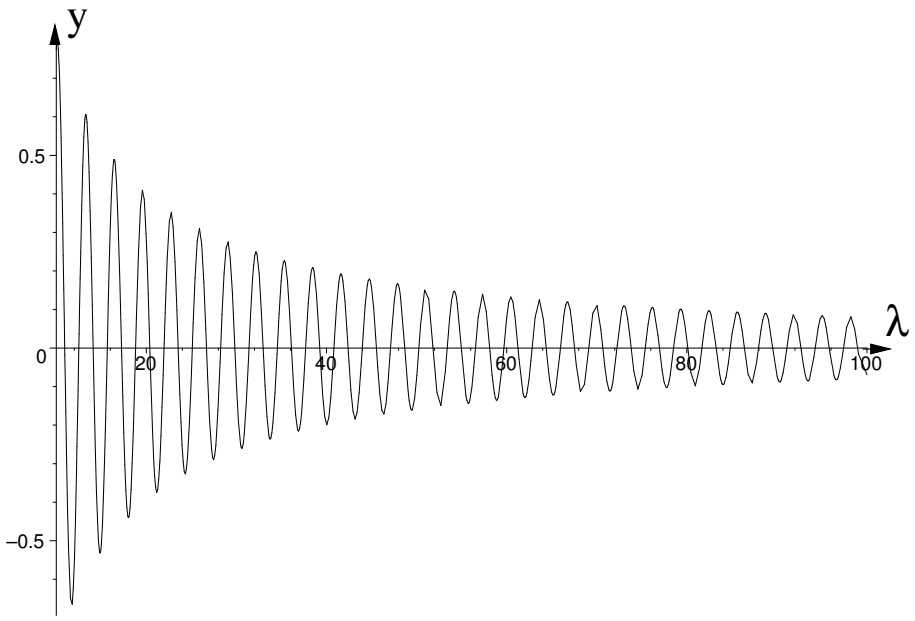

Figure 3.2: The decay of $y = \int_0^2 g(x)\cos\lambda x dx$, with $g(x) = x^3 - x + 2$, for $\lambda \in [10, 100]$.

$$= -\int_{a-\pi/\lambda}^{b-\pi/\lambda} g\left(t + \frac{\pi}{\lambda}\right)\sin\lambda t dt.$$

We now add this result to our definition of I, to obtain $2I$. Also, we rename t inside the integral to be x. This is allowed, since the integration variable is only a dummy variable. We get

$$\begin{aligned}
2I =& \int_a^b g(x)\sin\lambda x dx - \int_{a-\pi/\lambda}^{b-\pi/\lambda} g\left(x + \frac{\pi}{\lambda}\right)\sin\lambda x dx \\
=& -\int_{a-\pi/\lambda}^{a} g\left(x + \frac{\pi}{\lambda}\right)\sin\lambda x dx + \\
&+ \left(\int_a^{b-\pi/\lambda} g(x)\sin\lambda x dx - \int_a^{b-\pi/\lambda} g\left(x + \frac{\pi}{\lambda}\right)\sin\lambda x dx\right) + \\
&+ \int_{b-\pi/\lambda}^{b} g(x)\sin\lambda x dx \\
=& I_1 + I_2 + I_3 \\
\Rightarrow \quad |2I| \leq& |I_1| + |I_2| + |I_3|,
\end{aligned}$$

where I_1, I_2 and I_3 are the first, second (the stuff inside the parentheses) and third term of the previous line.

We'll now proceed to show that each of I_1, I_2 and I_3 can be made arbitrarily small. For I_1 and I_3 this is due to our integration being over a tiny interval. For I_2, however, this is due to the cancelation that we discussed in our graphical proof. Indeed, the transformation $x = t + \pi/\lambda$ is exactly the linear shift that moves us over half a period, where the integrand has the opposite sign, and cancellations can be expected. This is illustrated in Fig. 3.3.

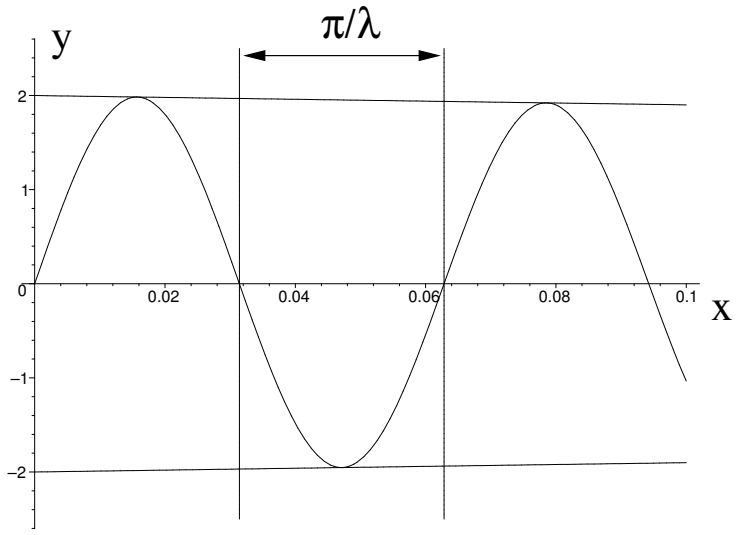

Figure 3.3: Shifting by π/λ moves you one half-period of the trig function. Shown are the function $y = g(x)\sin\lambda x$, with $g(x) = x^3 - x + 2$ and $\lambda = 100$. Also shown are $g(x)$ and $-g(x)$, which play the role of modulating amplitudes.

Let's look at I_3.

$$\begin{aligned} |I_3| &= \left| \int_{b-\pi/\lambda}^{b} g(x)\sin\lambda x dx \right| \\ &\leq \int_{b-\pi/\lambda}^{b} |g(x)||\sin\lambda x| dx \\ &\leq \int_{b-\pi/\lambda}^{b} |g(x)| dx \\ &\leq M\frac{\pi}{\lambda}, \end{aligned}$$

where

$$M = \max_{x\in[a,b]} |g(x)|.$$

Entirely similar, we obtain

$$|I_1| \leq M\frac{\pi}{\lambda}.$$

Lastly, we look at I_2. In doing so, we'll assume that λ is sufficiently large, so that $b - \pi/\lambda > a$. Then

$$\begin{aligned} |I_2| &= \left| \int_a^{b-\pi/\lambda} (g(x) - g(x+\pi/\lambda)) \sin \lambda x dx \right| \\ &\leq \left(b - \frac{\pi}{\lambda} - a \right) \max_{x\in[a,b-\pi/\lambda]} |g(x) - g(x+\pi/\lambda)| \,. \end{aligned}$$

Putting these three parts together, we get

$$\begin{aligned} |2I| \leq & \frac{2\pi M}{\lambda} + (b-a) \max_{x\in[a,b-\pi/\lambda]} |g(x) - g(x+\pi/\lambda)| \\ \leq & \frac{2\pi M}{\lambda} + (b-a) \max_{x\in[a,b]} |g(x) - g(x+\pi/\lambda)| \,. \end{aligned}$$

Since a continuous function on a bounded closed interval is uniformly continuous, this last maximum goes to zero as $\lambda \to \infty$. Thus

$$\lim_{\lambda\to\infty} I = 0,$$

which is what we had to prove. The cases with $\lambda \to -\infty$ or with $\cos \lambda x$ are all done similarly, and require no more work.

Now we'll look at what needs to be done if there are discontinuities at the endpoints of the interval. Notice that the preceding part did not require us to use that

$$\int_a^b |g(x)| dx < \infty.$$

We'll use this now: the existence of this integral allows us to control how bad the singularities in $g(x)$ at the endpoints a and b are allowed to be. Since the integral converges, we have that for any $\epsilon > 0$, we can find a $\delta > 0$ such that

$$\int_a^{a+\delta} |g(x)| dx < \frac{\epsilon}{2}, \quad \text{and} \quad \int_{b-\delta}^b |g(x)| dx < \frac{\epsilon}{2}.$$

Then (using our newly established result for the continuous case)

$$\begin{aligned}
\int_a^b g(x)\sin\lambda x dx &= \left(\int_a^{a+\delta} + \int_{a+\delta}^{b-\delta} + \int_{b-\delta}^{b}\right) g(x)\sin\lambda x dx \\
\Rightarrow \qquad \lim_{\lambda\to\infty}\int_a^b g(x)\sin\lambda x dx &= \lim_{\lambda\to\infty}\int_a^{a+\delta} g(x)\sin\lambda x dx + \lim_{\lambda\to\infty}\int_{a+\delta}^{b-\delta} g(x)\sin\lambda x dx + \\
&\quad \lim_{\lambda\to\infty}\int_{b-\delta}^{b} g(x)\sin\lambda x dx \\
&= \lim_{\lambda\to\infty}\int_a^{a+\delta} g(x)\sin\lambda x dx + \lim_{\lambda\to\infty}\int_{b-\delta}^{b} g(x)\sin\lambda x dx \\
\Rightarrow \qquad \left|\lim_{\lambda\to\infty}\int_a^b g(x)\sin\lambda x dx\right| &\le \lim_{\lambda\to\infty}\left|\int_a^{a+\delta} g(x)\sin\lambda x dx\right| + \lim_{\lambda\to\infty}\left|\int_{b-\delta}^{b} g(x)\sin\lambda x dx\right| \\
&\le \lim_{\lambda\to\infty}\int_a^{a+\delta} |g(x)| dx + \lim_{\lambda\to\infty}\int_{b-\delta}^{b} |g(x)| dx \\
&\le \lim_{\lambda\to\infty} 2\frac{\epsilon}{2} \\
&= \epsilon.
\end{aligned}$$

Since ϵ can be chosen arbitrarily small, we get

$$\lim_{\lambda\to\infty}\int_a^b g(x)\sin\lambda x dx = 0,$$

under the conditions stated in the theorem. This completes the proof of the Riemann-Lebesgue lemma. ■

2. The partial sum lemma

Let's look at Fourier series with $L = \pi$, which simplifies some of the expressions a bit. We now have

$$S(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos nx + b_n\sin nx\right),$$

so that the N-th partial sum S_N is

$$S_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N}\left(a_n\cos nx + b_n\sin nx\right).$$

Lemma 2 (Partial sum lemma) *Let $S_N(x)$ be the N-th partial sum of the 2π-periodic Fourier series obtained from the function $f(x)$,* i.e.,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx dx, \quad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx dx,$$

for $n = 0, 1, \ldots$. Then

$$S_N(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x+u)\frac{\sin(2N+1)\frac{u}{2}}{\sin\frac{u}{2}}du.$$

Proof. The proof consists of a straightforward calculation.

$$\begin{aligned}
S_N(x) &= \frac{a_0}{2} + \sum_{n=1}^{N}\left(a_n\cos nx + b_n\sin nx\right)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)dt + \sum_{n=1}^{N}\left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos ntdt\cos nx + \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin ntdt\sin nx\right)\\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\frac{1}{2} + \sum_{n=1}^{N}(\cos nt\cos nx + \sin nt\sin nx)\right]dt\\
&= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\frac{1}{2} + \sum_{n=1}^{N}\cos n(t-x)\right]dt.
\end{aligned}$$

Let's examine the part of the integrand that is independent of $f(t)$:

$$\begin{aligned}
\frac{1}{2} + \sum_{n=1}^{N}\cos n(t-x) &= \frac{1}{2} + \sum_{n=1}^{N}\Re\left[e^{in(t-x)}\right]\\
&= \frac{1}{2} + \Re\left[\sum_{n=1}^{N}\left(e^{i(t-x)}\right)^n\right]\\
&= \frac{1}{2} + \Re\left[e^{i(t-x)}\sum_{n=0}^{N-1}\left(e^{i(t-x)}\right)^n\right]\\
&= \frac{1}{2} + \Re\left[e^{i(t-x)}\frac{1-e^{i(t-x)N}}{1-e^{i(t-x)}}\right]\\
&= \frac{1}{2} + \Re\left[\left(\frac{e^{i(t-x)/2}\left(e^{i(t-x)N}-1\right)}{e^{i(t-x)/2}-e^{-i(t-x)/2}}\right)\right]\\
&= \frac{1}{2} + \Re\left[\frac{e^{i(t-x)(N+\frac{1}{2})}-e^{i(t-x)/2}}{2i\sin\frac{t-x}{2}}\right]\\
&= \frac{1}{2} - \frac{1}{2\sin\frac{t-x}{2}}\Re\left[i\left(e^{i(t-x)(N+\frac{1}{2})}-e^{i(t-x)/2}\right)\right]\\
&= \frac{1}{2} + \frac{1}{2\sin\frac{t-x}{2}}\left(\sin(2N+1)\frac{t-x}{2} - \sin\frac{t-x}{2}\right)\\
&= \frac{1}{2}\frac{\sin(2N+1)\frac{t-x}{2}}{\sin\frac{t-x}{2}}.
\end{aligned}$$

Ah, the wonders of complex numbers! In the above $\Re$ denotes the real part of the expression.

Substituting this back in to the expression for $S_N(x)$, we get

$$S_N(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin(2N+1)\frac{t-x}{2}}{\sin\frac{t-x}{2}}dt.$$

Notice that the integrand is periodic with period 2π. As the integral is over an entire period, it is not relevant what its upper and lower integration bounds are. Letting $u = t - x$, we get

$$S_N(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x+u)\frac{\sin(2N+1)\frac{u}{2}}{\sin\frac{u}{2}}du,$$

where we did not bother with the bounds of integration, because of our observation above. This is what we had to prove. ■.

The partial sum lemma provides an elegant, compact form for a partial sum of a Fourier series. The form even appears in a format that tempts us to use the Riemann-Lebesgue lemma as $N \to \infty$. That's not allowed though: the limit as $u \to 0$ of the integrand is (2N+1)f(x), which is not of the form suitable for the Riemann-Lebesgue lemma. Nevertheless, the partial sum lemma will be useful later on (otherwise it wouldn't be much of a lemma).

3. The Weierstrass approximation theorem

You may have seen the Weierstrass approximation theorem before, probably in a numerical analysis class. The theorem is usually phrased in its polynomial form, so that it states that every continuous function on an interval $[a, b]$ can be approximated arbitrarily close by a polynomial (the Weierstrass polynomial) so that the absolute error of the approximation does not exceed $\epsilon > 0$ anywhere in $[a, b]$. The degree of the Weierstrass polynomial depends on ϵ, of course.

In this section, we state and prove the theorem in its trigonometric form.

Theorem 1 (Weierstrass approximation) *Let $\epsilon > 0$ be arbitrary, and $f(x)$ is a 2π-periodic function, continuous on $[-\pi, \pi]$. There exists a trigonometric polynomial $T(x)$ such that $|f(x) - T(x)| < \epsilon$, for all $x \in [-\pi, \pi]$.*

Note that a trigonometric polynomial is any expression of the form

$$\frac{\alpha_0}{2} + \sum_{k=1}^{n}(\alpha_k \cos kx + \beta_k \sin kx).$$

In other words, a trigonometric polynomial is an expression that has a truncated Fourier series.

Proof. Define

$$T(x) = c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) f(x+u)du.$$

Note that $T(x)$ is a trigonometric polynomial: let $x + u = t$, then

$$T(x) = c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{t-x}{2}\right) f(t)dt.$$

Using the addition and subtraction formulas for cosine, this may be reduced to an expression with powers of $\cos x$ and $\sin x$. These powers can come out of the integral, which only cares about t. Thus $T(x)$ is a combination of powers of sines and cosines, with constant coefficients. These powers of sine and cosine are now reduced to sines and cosines of integer multiples of x, using the multi-angle formulas. Thus, indeed, $T(x)$ is a trigonometric polynomial.

Now, we choose c_n so that

$$c_n \int_{-\pi}^{\pi} \cos^{2n}\frac{u}{2} du = 1.$$

With a little effort and a lot of calculus, the integral may be done. You find

$$\begin{aligned} c_n &= \frac{2 * 4 * 6 * \cdots * 2n}{2\pi * 1 * 3 * \cdots * (2n-1)} \\ &= \frac{2}{3} * \frac{4}{5} * \cdots * \frac{2n-2}{2n-1} * \frac{2n}{2\pi} \\ &\leq \frac{n}{\pi}. \end{aligned}$$

Then

$$\begin{aligned} |T(x) - f(x)| &= \left| c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) f(x+u)du - f(x) \right| \\ &= \left| c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) f(x+u)du - c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) f(x)du \right| \\ &= \left| c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) \left(f(x+u) - f(x)\right) du \right| \\ &\leq c_n \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) \left|f(x+u) - f(x)\right| du \end{aligned}$$

As is seen from Fig. 3.4, the kernel function $\cos^{2n}(u/2)$ is sharply peaked at zero, but goes to zero rapidly elsewhere, for sufficiently high n. Because of this, we split the integral up in three parts: a central part over a small interval where the kernel is not small, and two outlying parts where the kernel is small. We'll use different techniques to control these different parts.

$$\begin{aligned} |T(x) - f(x)| \leq c_n \left(\int_{-\pi}^{-\delta} + \int_{\pi}^{\delta} \right) \cos^{2n}\left(\frac{u}{2}\right) |f(x+u) - f(x)| du + \\ c_n \int_{-\delta}^{\delta} \cos^{2n}\left(\frac{u}{2}\right) |f(x+u) - f(x)| du, \end{aligned}$$

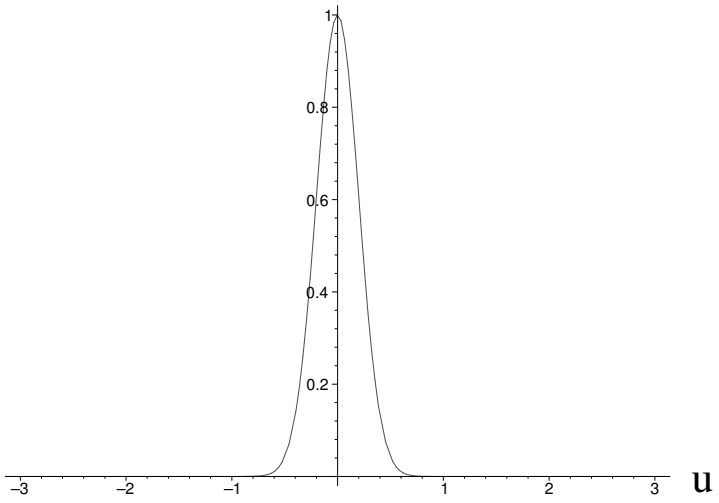

Figure 3.4: The kernel $\cos^{2n}(u/2)$ with $n = 50$.

where δ is a small positive constant, to be chosen as convenient.

Let's look at the central part. Since the kernel of the integral is not small here (in fact, it's near 1), we have to get what we need from the factor $|f(x+u) - f(x)|$ in the integrand. Indeed, this is possible: for any ϵ, we can choose δ so that for $|u| < \delta$ we have $|f(x+u) - f(x)| < \epsilon/2$, since $f(x)$ is continuous on a closed interval (thus it's uniformly continuous). Then

$$\begin{aligned} c_n \int_{-\delta}^{\delta} \cos^{2n}\left(\frac{u}{2}\right) |f(x+u) - f(x)| du &\leq c_n \frac{\epsilon}{2} \int_{-\delta}^{\delta} \cos^{2n}\left(\frac{u}{2}\right) du \\ &\leq c_n \frac{\epsilon}{2} \int_{-\pi}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) du \\ &= \frac{\epsilon}{2}. \end{aligned}$$

Next we look at the the other, noncentral parts. Because $f(x)$ is uniformly continuous, there exists a constant M such that

$$|f(x)| \leq M.$$

Then

$$\begin{aligned} c_n \left(\int_{-\pi}^{-\delta} + \int_{\delta}^{\pi} \right) \cos^{2n}\left(\frac{u}{2}\right) |f(x+u) - f(x)| du &\leq c_n \left(\int_{-\pi}^{-\delta} + \int_{\delta}^{\pi} \right) \cos^{2n}\left(\frac{u}{2}\right) 2M du \\ &\leq 2Mc_n 2 \int_{\delta}^{\pi} \cos^{2n}\left(\frac{u}{2}\right) du \end{aligned}$$

$$\leq \frac{4Mn}{\pi}\int_{\delta}^{\pi}\cos^{2n}\left(\frac{\delta}{2}\right)du$$
$$\leq 4Mn\cos^{2n}\left(\frac{\delta}{2}\right),$$

where the next-to-last step is based on $\cos(u/2)$ being a decreasing function on $[0,\pi]$. Notice that

$$n\cos^{2n}\left(\frac{\delta}{2}\right)\to 0,$$

as $n\to\infty$. Thus we can choose n so large that

$$4Mn\cos^{2n}\left(\frac{\delta}{2}\right)\leq\frac{\epsilon}{2}.$$

Putting everything together gives

$$|T(x)-f(x)|\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

This finishes the proof of the Weierstrass approximation theorem. ■

4. Least-square approximation: Bessel's inequality, Parseval's identity

Theorem 2 (Least-square approximation) *Let $f(x)$ be square integrable on $[-\pi,\pi]$,* i.e., *$\int_{-\pi}^{\pi}f^2(x)dx$ is finite. Then the n-th partial sum of the Fourier series provides the best least-square approximation to $f(x)$ of all trigonometric polynomials of order n.*

Proof. Let

$$T_n=\frac{\alpha_0}{2}+\sum_{k=1}^{n}\left(\alpha_k\cos kx+\beta_k\sin kx\right),$$

a trigonometric polynomial of order n. Then

$$\int_{-\pi}^{\pi}\left[f(x)-T_n(x)\right]^2dx=\int_{-\pi}^{\pi}f^2(x)dx-2\int_{-\pi}^{\pi}f(x)T_n(x)dx+\int_{-\pi}^{\pi}T_n^2(x)dx.$$

We now look at these last two terms individually.

The second term gives

$$\begin{aligned}\int_{-\pi}^{\pi}f(x)T_n(x)dx&=\int_{-\pi}^{\pi}f(x)\left(\frac{\alpha_0}{2}+\sum_{k=1}^{n}\left(\alpha_k\cos kx+\beta_k\sin kx\right)\right)dx\\&=\frac{\alpha_0}{2}\int_{-\pi}^{\pi}f(x)dx+\sum_{k=1}^{n}\alpha_k\int_{-\pi}^{\pi}f(x)\cos kxdx+\sum_{k=1}^{n}\beta_k\int_{-\pi}^{\pi}f(x)\sin kxdx\\&=\frac{\pi}{2}\alpha_0a_0+\pi\sum_{k=1}^{n}\left(\alpha_ka_k+\beta_kb_k\right),\end{aligned}$$

where a_k and b_k (k=0, 1, ...) are the Fourier coefficients corresponding to $f(x)$.

The last term gives

$$\begin{aligned}
&\int_{-\pi}^{\pi} T_n^2(x)dx \\
&= \int_{-\pi}^{\pi} \left(\frac{\alpha_0}{2} + \sum_{k=1}^{n} (\alpha_k \cos kx + \beta_k \sin kx) \right) \left(\frac{\alpha_0}{2} + \sum_{m=1}^{n} (\alpha_m \cos mx + \beta_m \sin mx) \right) dx \\
&= 2\pi \frac{\alpha_0^2}{4} + \pi \sum_{k=1}^{n} \alpha_k^2 + \pi \sum_{k=1}^{n} \beta_k^2 \\
&= \frac{\pi \alpha_0^2}{2} + \pi \sum_{k=1}^{n} \left(\alpha_k^2 + \beta_k^2 \right),
\end{aligned}$$

and we've used the orthogonality of the cosines and sines, as seen at the beginning of this chapter.

Putting things together, we get

$$\begin{aligned}
&\frac{1}{\pi} \int_{-\pi}^{\pi} \left[f(x) - T_n(x) \right]^2 dx \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x)dx - \alpha_0 a_0 - 2 \sum_{k=1}^{n} (\alpha_k a_k + \beta_k b_k) + \frac{a_0^2}{2} + \sum_{k=1}^{n} \left(\alpha_k^2 + \beta_k^2 \right) \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x)dx + \frac{(a_0 - \alpha_0)^2}{2} - \frac{a_0^2}{2} + \sum_{k=1}^{n} \left[(\alpha_k - a_k)^2 + (\beta_k - b_k)^2 \right] - \sum_{k=1}^{n} (a_k^2 + b_k^2).
\end{aligned}$$

Some of the terms in this expression depend on $f(x)$ only (the first, third, and last term) and are not affected by changes in $T_n(x)$. The other terms are all positive, since they are sums of squares. These terms do change as $T_n(x)$ is changed. It follows that the expression is minimized when these terms are. But these terms can all be made to vanish, exactly by choosing $\alpha_k = a_k$ and $\beta_k = b_k$, for $k = 0, 1, \ldots$. Doing this makes

$$T_n(x) = S_n(x),$$

the n-th partial sum of the Fourier series of $f(x)$, proving the theorem. ■

Thus the optimal least-square approximation to any square-integrable periodic function on a finite interval in the family of trigonometric polynomials of order n is given by the n-th partial sum of the Fourier series of $f(x)$. Note that we did not need to impose continuity of $f(x)$.

The least-square approximation theorem has a few important consequences. Letting $T_n(x) = S_n(x)$, we get

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \left[f(x) - S_n(x) \right]^2 dx = \frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x)dx - \frac{1}{2} a_0^2 - \sum_{k=1}^{n} (a_k^2 + b_k^2).$$

Taking the limit as $n \to \infty$, we obtain

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)dx \geq \frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}\left(a_k^2 + b_k^2\right).$$

This inequality is known as **Bessel's inequality**. The reason that we can't write an equality (without introducing more conditions on $f(x)$, see below) is that we can't conclude without extra work that the least-square error term will vanish in the limit.

Using the Weierstrass approximation theorem, we can do better than Bessel's inequality. Of course, this requires us to impose that $f(x)$ is continuous on $[-\pi, \pi]$. We get

$$\int_{-\pi}^{\pi} \left[f(x) - S_n(x)\right]^2 dx \leq \int_{-\pi}^{\pi} \left[f(x) - T(x)\right]^2 dx$$

where $T(x)$ is the Weierstrass polynomial of order n. The inequality holds because the least-square error among all trigonometric polynomials of order n is minized by $S_n(x)$. But we know that for a sufficiently large n we have

$$|f(x) - T(x)| < \epsilon,$$

uniformly in x, by the Weierstrass approximation theorem. This results in

$$\int_{-\pi}^{\pi} \left[f(x) - S_n(x)\right]^2 dx \leq 2\pi\epsilon^2.$$

Since ϵ is arbitrary, this implies

$$\lim_{n\to\infty}\int_{-\pi}^{\pi} \left[f(x) - S_n(x)\right]^2 dx = 0,$$

provided $f(x)$ is continuous and square integrable. This allows us to strengthen Bessel's inequality to obtain **Parseval's relation**: if $f(x)$ is continuous and 2π periodic, then

$$\boxed{\frac{1}{\pi}\int_{-\pi}^{\pi} f^2(x)dx = \frac{1}{2}a_0^2 + \sum_{k=1}^{\infty}\left(a_k^2 + b_k^2\right).}$$

Parseval's relation has a few immediate but important consequences:

- If $a_k \equiv b_k \equiv 0$ for all k, then $f(x) \equiv 0$.
- If $a_k \equiv \hat{a}_k$, $b_k \equiv \hat{b}_k$ for all k, then $f(x) \equiv \hat{f}(x)$. We'll use this often, when we equate coefficients of two different Fourier series, one known, the other one to be determined.

Parseval's relation is interpreted as a **completeness** relationship: when using Fourier modes to approximate functions in a least-square sense, we're not missing anything.

5. Some theorems from real analysis

Before we move on to one of the main results of Fourier analysis (Dirichlet's theorem; next section), we need to recall some results from real analysis.

- If the functions $\{f_n(x),\ n = 1, 2, \ldots\}$ are all continuous, and $f_n(x) \to f(x)$ uniformly in x as $n \to \infty$, then $f(x)$ is continuous.

  Another way to phrase this result is: the uniform limit of a sequence of continuous functions is continuous.

- If $g_n(x) \to g(x)$, uniformly on $x \in [a, b]$ as $n \to \infty$, then

$$\lim_{n\to\infty} \int_a^b g_n(x)dx = \int_a^b g(x)dx.$$

  Or, rephrased: the limit of the integral is the integral of the limit.

These results can be restated in terms of series, by identifying $f_n(x)$ with partial sums (check this!):

- If the functions $\{g_n(x),\ n = 1, 2, \ldots\}$ are all continuous, and $\sum_{k=1}^{n} g_k(x)$ is uniformly convergent, then $\sum_{k=1}^{\infty} g_k(x)$ is continuous.

- Let $\sum_{k=1}^{\infty} g_k(x)$ be uniformly convergent on $x \in [a, b]$, then

$$\lim_{n\to\infty} \int_a^b \sum_{k=1}^{n} g_k(x)dx = \int_a^b \lim_{n\to\infty} \sum_{k=1}^{n} g_k(x)dx,$$

  or

$$\sum_{k=1}^{\infty} \int_a^b g_k(x)dx = \int_a^b \sum_{k=1}^{\infty} g_k(x)dx.$$

  In other words, under these conditions, we can integrate a series term by term.

An important consequence of these results is the following theorem, which gives us conditions under which we can differentiate a series term by term.

Theorem 3 (Term-by-term differentiation) *If $\{g_n(x),\ n = 1, 2, \ldots\}$ are continuously differentiable on an interval I, and (i) $\sum_{k=1}^{\infty} g_k(a)$ converges at some $a \in I$, (ii) $\sum_{k=1}^{\infty} g_k'(x)$ is uniformly convergent on I, then*

$$\left(\sum_{k=1}^{\infty} g_k(x)\right)' = \sum_{k=1}^{\infty} g_k'(x)$$

on I.

Proof. Consider

$$\sum_{k=1}^{\infty} g_k'(x).$$

This series is uniformly convergent, thus we may integrate it term by term:

$$\int_a^x \sum_{k=1}^{\infty} g_k'(t)dt = \sum_{k=1}^{\infty} \int_a^x g_k'(t)dt = \sum_{k=1}^{\infty} (g_k(x) - g_k(a)) = \sum_{k=1}^{\infty} g_k(x) - \sum_{k=1}^{\infty} g_k(a),$$

since the latter series converges. Thus

$$\sum_{k=1}^{\infty} g_k(x) = \int_a^x \sum_{k=1}^{\infty} g_k'(t)dt + \sum_{k=1}^{\infty} g_k(a).$$

Taking a derivative of both sides we get

$$\left(\sum_{k=1}^{\infty} g_k(x) \right)' = \sum_{k=1}^{\infty} g_k'(x),$$

which is what we had to prove. ■

The last theorem from real analysis we need is the **Weierstrass M-test**: suppose that for all n we have $|g_n(x)| \leq M_n$, for $x \in S$, some interval. Here M_n is independent of x. If the series $\sum_{k=1}^{\infty} M_k$ is convergent, then the series $\sum_{k=1}^{\infty} g_k(x)$ is absolutely and uniformly convergent on S.

6. Dirichlet's theorem

Now we're armed and ready to go. Let's give a first answer to the question of convergence of a Fourier series.

Theorem 4 (Fourier series convergence) *If the Fourier series $S(x)$ of a 2π-periodic continuous function $f(x)$ converges uniformly for all $x \in [-\pi, \pi]$, then $S(x) = f(x)$.*

Proof. Let

$$S(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos kx + b_k \sin kx \right),$$

the Fourier series obtained from $f(x)$. Then clearly $S(x)$ is 2π-periodic, and $S(x)$ is continuous, as the uniform limit of a sequence of continuous functions (the partial sums).

Now consider $S(x) \cos nx$. This series is also uniformly convergent, so it can be integrated term by term. Similarly with $S(x) \sin nx$. This way we find that the Fourier coefficients of $S(x)$ are identical to those of $f(x)$. Since both of $f(x)$ and $S(x)$ are continuous and periodic, and their Fourier coefficients are equal, we have that $f(x) = S(x)$, as a consequence of Parseval's relation. ■

As immediate consequence of the Weierstrass M-test is the following.

Corollary 1 *Let $f(x)$ be 2π-periodic and continuous. Then its Fourier series is absolutely and uniformly convergent to $f(x)$ if $\sum_{k=1}^{\infty}|a_k|$ and $\sum_{k=1}^{\infty}|b_k|$ are convergent.*

Next we come to Dirichlet's theorem, which is one of the most important results in Fourier analysis.

Theorem 5 (Dirichlet) *Let $f(x)$ be 2π-periodic and piecewise smooth. Then its Fourier series converges for each $x \in [-\pi, \pi]$: for each fixed x*

$$\lim_{n\to\infty} S_n(x) = S(x) = \frac{f(x^-) + f(x^+)}{2},$$

where

$$f(x^-) = \lim_{h>0} f(x-h), \quad f(x^+) = \lim_{h>0} f(x+h).$$

Further, if $f(x)$ is continuous, then $S(x)$ converges absolutely and uniformly on any closed subinterval.

Proof. According to the partial sum lemma

$$S_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin(2n+1)\frac{u}{2}}{\sin\frac{u}{2}} f(x+u)du.$$

Using $f(x) = 1$ in this identity gives

$$\begin{aligned} 1 &= \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{\sin(2N+1)\frac{u}{2}}{2\sin\frac{u}{2}}du \\ \Rightarrow \qquad \frac{1}{2} &= \frac{1}{\pi}\int_{0}^{\pi} \frac{\sin(2N+1)\frac{u}{2}}{2\sin\frac{u}{2}}du, \end{aligned}$$

since the integrand is even. Thus

$$\frac{f(x_0^-) + f(x_0^+)}{2} = \frac{1}{\pi}\int_0^{\pi} \frac{\sin(2N+1)\frac{u}{2}}{2\sin\frac{u}{2}} \left[f(x_0^-) + f(x_0^+)\right] du.$$

Also,

$$\begin{aligned} S_n(x_0) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin(2n+1)\frac{u}{2}}{\sin\frac{u}{2}} f(x_0+u)du \\ &= \frac{1}{\pi}\int_0^{\pi} \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} f(x_0+u)du + \frac{1}{\pi}\int_{-\pi}^{0} \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} f(x_0+u)du \\ &= \frac{1}{\pi}\int_0^{\pi} \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} f(x_0+u)du - \frac{1}{\pi}\int_{\pi}^{0} \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} f(x_0-u)du \\ &= \frac{1}{\pi}\int_0^{\pi} \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} \left[f(x_0+u) + f(x_0-u)\right] du. \end{aligned}$$

Combining this with our previous result, we get

$$\begin{aligned} & S_n(x_0) - \frac{f(x_0^-) + f(x_0^+)}{2} \\ &= \frac{1}{\pi} \int_0^\pi \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} \left[f(x_0+u) + f(x_0-u) - f(x_0^-) - f(x_0^+) \right] du \\ &= \frac{1}{\pi} \int_0^\pi \frac{\sin(2n+1)\frac{u}{2}}{2\sin\frac{u}{2}} u \left[\frac{f(x_0+u) - f(x_0^+)}{u} + \frac{f(x_0-u) - f(x_0^-)}{u} \right] du. \end{aligned}$$

If we use the Riemann-Lebesgue lemma at this point we get

$$\lim_{n\to\infty} \left(S_n(x_0) - \frac{f(x_0^-) + f(x_0^+)}{2} \right) = 0,$$

from which the desired result follows immediately. The question remains whether we can use the Riemann-Lebesgue lemma. In order to use it, we need the function

$$\frac{u}{2\sin\frac{u}{2}} \left[\frac{f(x_0+u) - f(x_0^+)}{u} + \frac{f(x_0-u) - f(x_0^-)}{u} \right]$$

to be piecewise continuous and integrable. Let's check.

- **Piecewise continuous:** since $f(x)$ is assumed to be piecewise continuous, there's not much too check here. The only problem might be caused by $u = 0$, but

$$\begin{aligned} & \lim_{u\to 0} \frac{u}{2\sin\frac{u}{2}} \left[\frac{f(x_0+u) - f(x_0^+)}{u} + \frac{f(x_0-u) - f(x_0^-)}{u} \right] \\ & \qquad = \begin{cases} 0 \text{ if } f(x) \text{ is continuous and smooth at } x = x_0, \\ \text{finite, with different left- and right-limits otherwise.} \end{cases} \end{aligned}$$

- **Integrability:** Since $f(x)$ is piecewise smooth on $[-\pi, \pi]$, it may have discontinuities at some points, but its left- and right-limits are always defined and finite. Thus $f(x)$ is integrable. This means that, as before, $u = 0$ is our only point of contention. Notice that we only need to consider $u > 0$, since the integration is from $u = 0$ to $u = \pi$. Then, as $u \to 0$,

$$\frac{f(x_0+u) - f(x_0^+)}{u} \to f'(x_0^+),$$

and

$$\frac{f(x_0-u) - f(x_0^-)}{u} \to -f'(x_0^-).$$

Thus this part of the integrand does not cause a problem near the lower bound $u = 0$. Neither does the factor $u/(2\sin(u/2))$, which approaches 1 near the lower bound. Thus the integrand is integrable.

Thus the Riemann-Lebesgue lemma applies, proving the first part of Dirichlet's theorem.

To prove the second part of the theorem, assume that $f(x)$ is 2π-periodic, continuous and piecewise smooth. It suffices to show that

$$\sum_{k=1}^{\infty} |a_k| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} |b_k| < \infty.$$

This establishes absolute and uniform convergence by the Weierstrass M-test.

We have

$$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx \\ &= \frac{1}{\pi} \left(f(x) \frac{\sin nx}{n} \Big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f'(x) \frac{\sin nx}{n} dx \right) \\ &= -\frac{1}{n\pi} \int_{-\pi}^{\pi} f'(x) \sin nx dx \\ &= -\frac{1}{n} \beta_n, \end{aligned}$$

where β_n is a Fourier coefficient of $f'(x)$. Using

$$ab \leq \frac{1}{2} \left(a^2 + b^2 \right)$$

we have

$$|a_n| = \frac{1}{n} |\beta_n| \leq \frac{1}{2} \left(\frac{1}{n^2} + |\beta_n|^2 \right),$$

or

$$\begin{aligned} \sum_{n=1}^{\infty} |a_n| &\leq \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n^2} + \frac{1}{2} \sum_{n=1}^{\infty} |\beta_n|^2 \\ &= \frac{\pi^2}{12} + \frac{1}{2} \sum_{n=1}^{\infty} |\beta_n|^2 \\ &< \infty, \end{aligned}$$

by Bessel's inequality, since

$$\frac{1}{\pi} \int_{\pi}^{\pi} f'^2(x) dx$$

is finite.

The proof for $\sum_{n=1}^{\infty} |b_n| < \infty$ is similar. If there are points of discontinuity in $f'(x)$ at isolated points, the interval has to be split up in parts on which the function is continuous, as in the proof of the Riemann-Lebesgue lemma. ■.

3.2 Fourier integrals and transforms

In this section we'll introduce Fourier transforms in much the same way that we introduced Fourier series. Since many of the results are quite similar, we won't present the results here with the same level of rigor. Few proofs will be given.

Given a function $f(x)$, the Fourier transform of $f(x)$, for $x \in \mathbb{R}$ is defined as

$$\mathcal{F}[f(x)](\omega) = \hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx,$$

whenever this integral exists. The existence of the integral requires

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty.$$

A few notes:

- The above definition of the Fourier transform is one of many you'll find in the literature. They're all equivalent, but minus signs, factors of 2π, *etc.* may be in different places. I'm using the above definition, which may be the most commonly used.
- From the Riemann-Lebesgue lemma (or an extension of it to the whole real line), it follows that $\hat{f}(\omega) \to 0$ as $\omega \to \pm\infty$, provided $f(x)$ is continuous, or has at most a finite number of jump discontinuities.

Theorem 6 (Inverse of the Fourier transform) *Let $f(x)$ be continuous, piecewise smooth, and absolutely integrable. Then*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} \hat{f}(\omega) d\omega.$$

We won't prove this theorem, but a few notes are in order.

- If $f(x)$ is only piecewise continuous, then

$$\frac{f(x^+) + f(x^-)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} \hat{f}(\omega) d\omega.$$

  This can be thought of as a Fourier series analog of Dirichlet's theorem.

- The integral defining the inverse Fourier transform in this theorem is supposed to be interpreted in a Cauchy Principal Value sense. Note that this is not necessary for the original integral defining the Fourier transform itself.

Some more theorems, all without proof.

Theorem 7 *Let $f(x)$ be piecewise continuous and absolutely integrable. Then $\hat{f}(\omega)$ is continuous on $\mathbb{R}$.*

Theorem 8 *Let $f(x), f'(x), f''(x), \ldots$ be continuous and absolutely integrable. Then*

$$\mathcal{F}\left[f^{(k)}(x)\right](\omega) = (i\omega)^k \mathcal{F}\left[f(x)\right](\omega),$$

for all k for which these conditions are satisfied. Here $f^{(k)}(x)$ denotes the k-th derivative of $f(x)$.

1. Extensions to higher dimensions

In $\mathbb{R}^n$, we have

$$\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^n} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}} f(\boldsymbol{x}) d\boldsymbol{x},$$

where $\boldsymbol{x} \in \mathbb{R}^n$, and $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$. This form is valid whenever the integral exists. Its inversion is given by

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}} \hat{f}(\boldsymbol{\omega}) d\boldsymbol{\omega},$$

for $f(\boldsymbol{x})$ continuous, piecewise smooth, and absolutely integrable.

We won't go into the details of Fourier transforms to the same extent as we did for Fourier series. As stated before, the proofs are very similar, with the role of uniform convergence of infinite series taken over by uniform convergence of improper integrals. Let's state the theorems from real analysis that are required to do these proofs, as they'll be useful later.

2. More theorems from real analysis

Theorem 9 (Weierstrass M-test for improper integrals) *Suppose that*

$$F(x) = \int_{-\infty}^{\infty} g(x, y) dy$$

exists for each $x \in S \subset \mathbb{R}$, and that for $x \in S$, $y \in \mathbb{R}$ we have

$$|g(x, y)| \leq M(y), \quad \textit{with} \quad \int_{-\infty}^{\infty} M(y) dy < \infty.$$

Then $F(x)$ is absolutely and uniformly convergent.

Theorem 10 *Assume that $g(x, y)$ is defined on $(x, y) \in [a, b] \times \mathbb{R}$, and that $g(x, y)$ is piecewise continuous as a function of y. If*

$$F(x) = \int_{-\infty}^{\infty} g(x, y) dy$$

is uniformly convergent for all $x \in [a, b]$, then $F(x)$ is continuous for $x \in [a, b]$.

Theorem 11 (Switching the order of integration) *Under the circumstances of the previous theorem (uniform convergence of $F(x)$, piecewise continuity in y) we have that*

$$\int_a^b \left(\int_{-\infty}^{\infty} g(x,y)dy \right) dx = \int_{-\infty}^{\infty} \left(\int_a^b g(x,y)dx \right) dy.$$

Using a similar proof to that for switching infinite sums and derivatives we can prove a similar theorem for improper integrals and differentiation.

Theorem 12 (Switching derivatives and improper integration) *Suppose that $g(x,y)$ and $g_x(x,y)$ are continuous for $-\infty \leq \alpha < x < \beta \leq \infty$, $y \in \mathbb{R}$, then*

$$\frac{d}{dx} \int_{-\infty}^{\infty} g(x,y)dy = \int_{-\infty}^{\infty} \frac{\partial g}{\partial x}(x,y)dy,$$

for $x \in (a,b)$, provided both integrals exist, and provided the one on the right is uniformly convergent on bounded, closed subintervals of (α, β).

Example: As an example of what these theorems can do for us, we'll calculate

$$I = \int_0^{\infty} \frac{\sin y}{y} dy.$$

We'll use the different theorems many times over when we start solving partial differential equations in the following chapters, but this example may keep the dogs down for now.

Usually, this integral is done using methods of complex analysis, especially the residue theorem. We'll do it a different way here.

Consider

$$L(x) = \int_0^{\infty} e^{-xy} \frac{\sin y}{y} dy,$$

for $x \geq 0$. We wish to compute $L(0)$. In terms of the preceding theorems we have

$$g(x,y) = e^{-xy} \frac{\sin y}{y},$$

and

$$g_x(x,y) = -e^{-xy} \sin y.$$

Both of these are continuous on the intervals considered. Thus, for $x > 0$ (the above theorem does not allow us to include $x = 0$) we have

$$L'(x) = -\int_0^{\infty} e^{-xy} \sin y dy = -\frac{1}{1+x^2}.$$

This last integral is easily done using standard calculus methods, such as good ol' integration by parts. Another way is to use Laplace transforms.

Thus $L(x) = c - \arctan x$, for some constant c. Next, we have that

$$|L(x)| \leq \int_0^\infty e^{-xy} dy = \frac{1}{x}.$$

Since $1/x \to 0$ as $x \to \infty$, we have that $|L(x)| \to 0$ as $x \to \infty$. Thus, evaluating $L(x) = c - \arctan x$ as $x \to \infty$, we get that $c = \pi/2$. In summary

$$L(x) = \frac{\pi}{2} - \arctan x.$$

Since $L(x)$ is continuous for $x = 0$, we have that

$$\int_0^\infty \frac{\sin y}{y} dy = L(0) = \lim_{x \to 0} \left(\frac{\pi}{2} - \arctan x \right) = \frac{\pi}{2},$$

which is the desired result.

3.3 An example of why rigor may matter

Let's look at an example where the standard methods involving blindfolded term-by-term differentiation *etc* fail.

Consider the following conducting wave slab problem:

$$E_{yy} + E_{zz} + K^2 E = i\omega\mu_0 J.$$

Here E is the x-component of the electric field and J is a volume source, such as a current, prescribed inside the area of the slab, see Fig. 3.5. This problem describes the linear dynamics of the electric field in a planar wave slab, used for communication purposes such as antennas.

Assume that the volume source J is given as

$$J = \sum_{k=1}^{\infty} B_k(z) \sin\left(\frac{k\pi y}{L} \right).$$

As indicated in Fig. 3.5, the electric field E vanishes on the edges of the wave conductor, except at perhaps a few places where there is a hole in the surrounding (insulating) apperture. At those places we may assume the field is known. For now we'll consider the case of point sources on the boundary, for instance in the form of a Dirac delta function.

Because of the boundary conditions and the prescribed form of J, it is reasonable to expand E in a Fourier sine series:

$$E = \sum_{k=1}^{\infty} E_k(z) \sin\left(\frac{k\pi y}{L} \right),$$

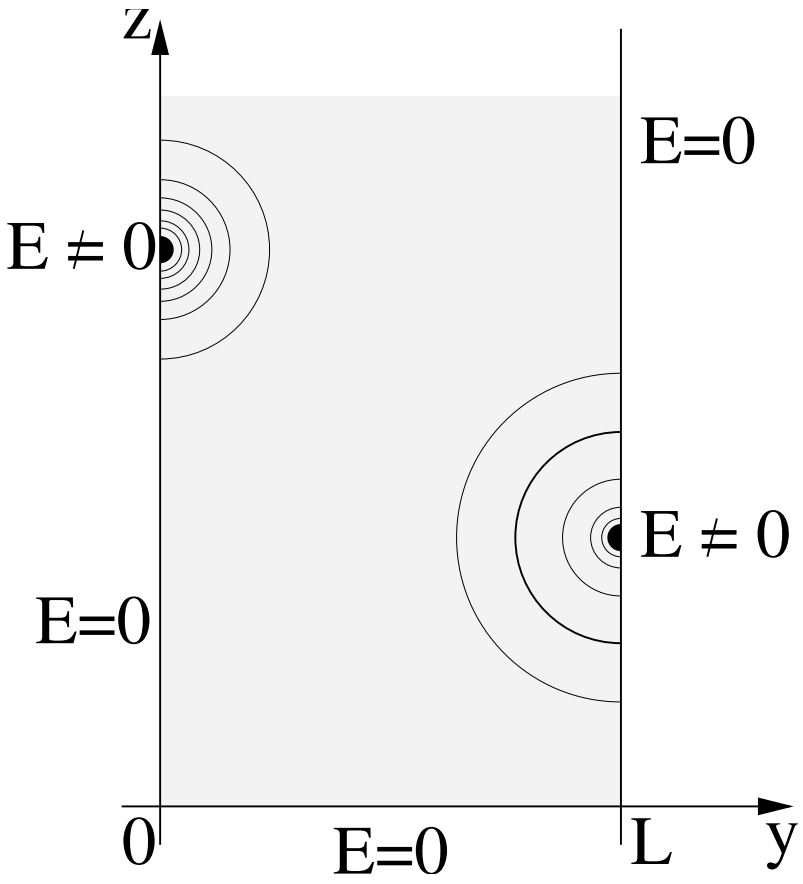

Figure 3.5: The wave slab set up.

for $0 < x < L$. A typical way to proceed would be to take as many derivatives of this expression as are required for the partial differential equation. In our case, we'd compute

$$E_{zz} = \sum_{k=1}^{\infty} E_k''(z) \sin\left(\frac{k\pi y}{L}\right),$$
$$E_{yy} = -\sum_{k=1}^{\infty} E_k(z) \left(\frac{k\pi}{L}\right)^2 \sin\left(\frac{k\pi y}{L}\right).$$

There are several problems with this ansatz.

- **Physical:** With these assumptions E_{yy} on the boundary $z = 0$ or $z = L$ is always zero. With the presence of holes in the apperture, this is not realistic. A delta function disturbance will typically lead to discontinuous derivatives.
- **Mathematical:** even if the Fourier series for E converges, we see that the Fourier coefficients for E_{yy} are multiplied by $(k\pi/L)^2$, thus it is not obvious that the Fourier

series for E_{yy} converges, and the term-by-term differentiation may not be justified. The above implies that as a consequence of the holes in the apperture the Fourier expansion for the electric field may not converge as rapidly as we'd want, with the consequence of the term-by-term differentiation failing.

To fix this problem, we propose a new, free and easy Fourier series for E_{yy}:

$$E_{yy} = \sum_{k=1}^{\infty} F_k(z) \sin\left(\frac{k\pi y}{L}\right).$$

Our goal is now to find a relationship between the Fourier coefficients $\{F_k(z)\}$ of E_{yy} and the Fourier coefficients $\{E_k(z)\}$ of E. By definition of the Fourier coefficients we have

$$\begin{aligned}
F_k(z) &= \frac{2}{L}\int_0^L E_{yy} \sin\left(\frac{k\pi y}{L}\right) dy \\
&= \frac{2}{L}\left(E_y \sin\left(\frac{k\pi y}{L}\right)\Big|_{y=0}^{y=L} - \frac{k\pi}{L}\int_0^L E_y \cos\left(\frac{k\pi y}{L}\right) dy \right) \\
&= -\frac{2k\pi}{L^2}\int_0^L E_y \cos\left(\frac{k\pi y}{L}\right) dy \\
&= -\frac{2k\pi}{L^2}\left(E \cos\left(\frac{k\pi y}{L}\right)\Big|_{y=0}^{y=L} + \int_0^L E\frac{k\pi}{L} \sin\left(\frac{k\pi y}{L}\right) dy \right) \\
&= -\frac{2k\pi}{L^2}\left(E(L,z)(-1)^k - E(0,z)\right) - \frac{k^2\pi^2}{L^2}E_k(z).
\end{aligned}$$

We see that this relationship is different from the one we find using term-by-term differentiation. In that case we find

$$F_k(z) = -\frac{k^2\pi^2}{L^2}E_k(z).$$

The extra terms are all due to the boundary effects at the holes in the apperture, as expected. Since the electric field there is supposed to be known, the relation

$$F_k(z) = -\frac{2k\pi}{L^2}\left(E(L,z)(-1)^k - E(0,z)\right) - \frac{k^2\pi^2}{L^2}E_k(z)$$

completely specifies the Fourier series for E_{yy}.

Our partial differential equation becomes

$$\begin{aligned}
&\sum_{k=1}^{\infty} \sin\left(\frac{k\pi y}{L}\right)\left[E_k'' + K^2 E_k - \frac{k^2\pi^2}{L^2}E_k - \frac{2k\pi}{L^2}\left(E(L,z)(-1)^k - E(0,z)\right)\right] \\
&= \sum_{k=1}^{\infty} B_k \sin\left(\frac{k\pi y}{L}\right) i\omega\mu_0.
\end{aligned}$$

Here the prime denotes differentiation with respect to z. Using the orthogonality of the different sines, we equate the coefficients of like terms. This gives

$$E_k'' + \left(K^2 - \frac{k^2\pi^2}{L^2}\right) E_k = i\omega\mu_0 B_k + \frac{2k\pi}{L^2}\left(E(L,z)(-1)^k - E(0,z)\right)$$

The first term on the right-hand side is due to the volume source in the problem, whereas the other terms on the right are a consequence of the boundary sources in the holes of the insulating apperture. This is the first instance of a few that we'll encounter in this course of the boundary conditions entering the governing equations as source terms.

The above ordinary differential equation is now solved as a superposition of a homogeneous and a particular solution. In this example the term-by-term differentiation caused wrong results, but the difficulties were relatively easy to circumvent. The important part is to be aware of a potential problem like this one.

3.4 Exercises

1. We have seen that the Fourier series of $f(x)$ can be termwise differentiated if $f(x)$ is continuous and piecewise smooth. Apart from functions whose Fourier series has a finite number of terms, can you think of any other functions in the mathematical compendium whose Fourier series can be differentiated ad exhaustiam? Essentially this question is asking to write down an $f(x)$ whose Fourier coefficients decay faster than any power of n, as taking derivatives multiplies these Fourier coefficients by powers of n. Fourier series are quite different from Taylor series in this way: we know many Taylor series whose derivatives have Taylor series that are equally well converging. Examples are exponentials, hyperbolics, trig, and others in their domain of convergence. Usually the worst thing that happens when we take a termwise derivative of a Taylor series is that the domain of convergence becomes a bit smaller. In the case of a Fourier series it is most common that the domain of convergence completely dissapears if we take too many termwise derivatives. So: do we know any other "nice" periodic functions other than the trig functions?

2. Invent a function defined by a Fourier series with an infinite number of terms that is infinitely differentiable.

3. **About the Riemann-Lebesgue lemma** The Riemann-Lebesgue lemma can easily be extended to improper integrals, using essentially the same assumptions as in Lemma 1, with $a = -\infty$ and/or $b = \infty$.

   (a) Consider $I(\alpha) = \int_{-\infty}^{\infty} e^{-x^2/2}\cos(\alpha x)dx$. The R-L lemma states that $\lim_{\alpha\to\infty} I(\alpha) = 0$. (a) What's the best estimate you can get for how fast $I(\alpha)$ approaches 0? (b) Calculate $I(\alpha)$, by obtaining a first-order differential equation for $I(\alpha)$, and solving it. Having found $I(\alpha)$, how fast does it approach 0? (c) Calculate the Fourier transform of $e^{-x^2/2}$.

(b) The requirement of absolute integrability of $f(x)$ is essential for the R-L lemma, as it is based on lots of cancelations due to the sign changes of the sin and/or cos functions. Consider the Fresnel integrals of geometric optics:

$$\int_{-\infty}^{\infty} \cos(x^2)dx = \int_{-\infty}^{\infty} \sin(x^2)dx = \sqrt{\frac{\pi}{2}}.$$

One might use a similar argument as that for the cancelations in the R-L lemma to argue the convergence of these integrals. Assuming the above result, show that

$$\int_{-\infty}^{\infty} \sin(x^2)\cos(\alpha x)dx = \sqrt{\pi}\cos\left(\frac{\alpha^2+\pi}{4}\right),$$

which does not converge to zero as $\alpha \to \infty$. What is the Fourier transform of $\sin(x^2)$?

Chapter 4

Wave propagation in one spatial dimension

In this chapter we investigate the one-dimensional wave equation

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

As we've seen before, this is the prototype of a second-order hyperbolic equation. For definiteness, we'll assume that $c > 0$.

4.1 d'Alembert's solution of the wave equation

1. Light-cone variables

Let's introduce the characteristic variables

$$\begin{cases} \alpha &= x + ct \\ \beta &= x - ct \end{cases}.$$

A similar calculation like the ones done before gives: (chain rule galore)

$$\begin{aligned} u_x &= u_\alpha + u_\beta \\ u_{xx} &= u_{\alpha\alpha} + 2u_{\alpha\beta} + u_{\beta\beta} \\ u_t &= c\left(u_\alpha - u_\beta\right) \\ u_{tt} &= c^2\left(u_{\alpha\alpha} - 2u_{\alpha\beta} + u_{\beta\beta}\right). \end{aligned}$$

Substituting these results in the one-dimensional wave equation gives

$$\boxed{u_{\alpha\beta} = 0.}$$

This is often called the light-cone form of the wave equation. This terminology originates from the theory of relativity, where c denotes the velocity of light. In this form, the equation

is easily integrated. We get

$$\begin{aligned} & u(\alpha,\beta) = A(\alpha) + B(\beta) \\ \Rightarrow \qquad & u(x,t) = A(x+ct) + B(x-ct). \end{aligned}$$

Here A and B are arbitrary functions of their arguments. Just like the general solution of a second-order ordinary differential equation depends on two arbitrary constants, we see that the general solution of this second-order partial differential equation with two independent variables depends on two arbitrary functions of one variable.

Note that $A(x+ct)$ represents a wave traveling to the left with velocity $-c$, whereas $B(x-ct)$ represents a wave going to the right with velocity c. We see that the wave equation splits given initial data up in a left-going and a right-going part. This is illustrated in Fig. 4.1.

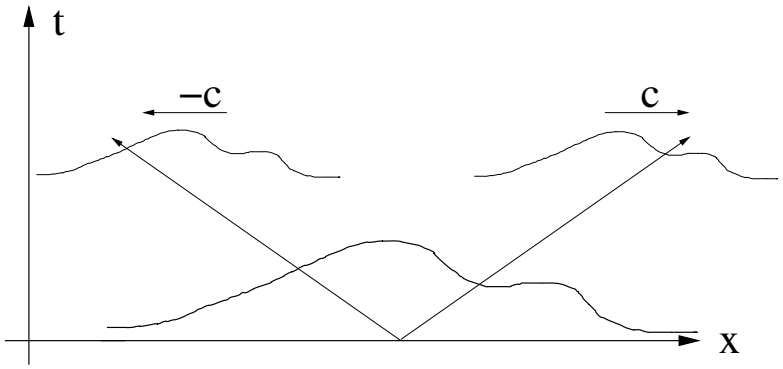

Figure 4.1: The wave equation splits given initial data up in a left-going and a right-going part.

2. The initial-value problem

Let's consider the initial-value problem

$$\left\{ \begin{array}{rcl} u_{tt} & = & c^2 u_{xx} \\ u(x,0) & = & U(x), \quad u_t(x,0) = V(x) \end{array} \right. ,$$

where $x \in \mathbb{R}$. Using the above general solution, we get

$$\begin{aligned} u(x,t) &= A(x+ct) + B(x-ct) \\ u_t(x,t) &= cA'(x+ct) - cB'(x-ct), \end{aligned}$$

where the $'$ denotes differentiation with respect to the argument. Evaluating these at $t=0$, we obtain

$$\begin{aligned} U(x) &= A(x) + B(x) \\ V(x) &= cA'(x) - cB'(x). \end{aligned}$$

Integrating this last equation gives

$$A(x) - B(x) = \frac{1}{c}\int_0^x V(z)dz + c_1,$$

where c_1 is constant. Adding to this the equation for $A(x) + B(x)$, we get

$$\begin{aligned} 2A(x) &= U(x) + \frac{1}{c}\int_0^x V(z)dz + c_1 \\ \Rightarrow \qquad A(x) &= \frac{1}{2}U(x) + \frac{1}{2c}\int_0^x V(z)dz + \frac{c_1}{2}. \end{aligned}$$

Similarly, by subtracting the two equations we get $B(x)$. Alternatively,

$$\begin{aligned} B(x) &= U(x) - A(x) \\ &= \frac{1}{2}U(x) - \frac{1}{2c}\int_0^x V(z)dz - \frac{c_1}{2} \\ &= \frac{1}{2}U(x) + \frac{1}{2c}\int_x^0 V(z)dz - \frac{c_1}{2}. \end{aligned}$$

Putting these results together results in

$$\begin{aligned} u(x,t) &= A(x+ct) + B(x-ct) \\ &= \frac{1}{2}U(x+ct) + \frac{1}{2c}\int_0^{x+ct} V(z)dz + \frac{c_1}{2} + \frac{1}{2}U(x-ct) + \frac{1}{2c}\int_{x-ct}^0 V(z)dz - \frac{c_1}{2} \\ &= \frac{U(x+ct)+U(x-ct)}{2} + \frac{1}{2c}\int_{x-ct}^{x+ct} V(z)dz, \end{aligned}$$

which is **d'Alembert's solution** of the wave equation with initial data specified on the real line. It generalizes to any number of spatial dimensions without much trouble. If follows from D'Alembert's solution that the wave equation with initial data defined on the entire real line is well-posed.

3. The cone of influence and the cone of dependence

From the characteristic picture discussed in Chapter 2, but also from the preceding calculation, it is clear that the influence of any point (x_0, y_0) is restricted to a cone emanating from this point with slopes $\pm c$, as shown in Fig. 4.2. This cone is called the **cone of influence** of (x_0, t_0): no matter the size of the disturbance at (x_0, t_0), its effect will never be felt outside of its cone of influence.

Similarly, the point (x_0, t_0) does not depend on any disturbances outside of its **cone of dependence**, given by the same slopes $\pm c$, but for $t < t_0$. The cone of dependence of (x_0, t_0) is the collection of all points that have (x_0, y_0) in their cone of influence. This is also illustrated in Fig. 4.2.

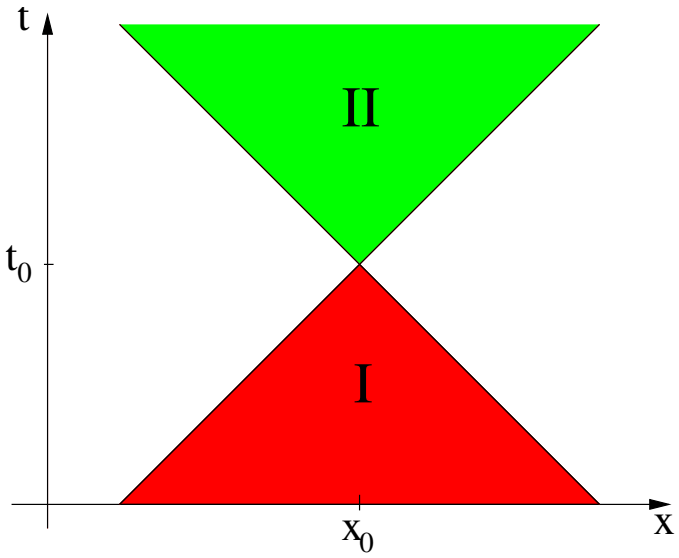

Figure 4.2: The cone of dependence (region I) and the cone of influence (region II) of the point (x_0, t_0).

4.2 Homogeneous initial-value problems

Let's look at an example where we impose boundary conditions on a finite-size domain $x \in [0, L]$, for some positive length L. Such an example is given by considering the wave equation as describing a string with fixed ends, with given initial profile.

$$\begin{cases} u_{tt} = c^2 u_{xx} & 0 < x < L,\ t > 0 \\ u(x,0) = f(x) & 0 \le x \le L \\ u_t(x,0) = g(x) & 0 \le x \le L \\ u(0,t) = 0 = u(L,t) & t \ge 0 \end{cases}.$$

This problem is an initial-boundary-value problem, since it involves both initial and boundary conditions. It is also a homogeneous problem because there is no forcing in the partial differential equation (making $u = 0$ a solution, albeit not a very interesting one), and the boundary conditions are zero as well.

1. Separation of variables

We'll solve this problem using the standard **separation of variables** method. The idea is to look for simple solutions that we will superimpose later.

Let's try to find solutions of the form

$$u(x,t) = X(x)T(t).$$

Substitution in the partial differential equation gives

$$T''X = c^2X''T \quad \Rightarrow \quad \frac{T''}{c^2T} = \frac{X''}{X}.$$

The left-hand side of this equation is a function of t alone, whereas the right-hand side depends only on x. These two sides can only be equal if both sides are constant. Thus

$$\frac{T''}{c^2T} = \frac{X''}{X} = K,$$

where K is a constant. This implies the two ordinary differential equations

$$\begin{aligned} X'' - KX &= 0, \quad \text{and} \\ T'' - c^2KT &= 0. \end{aligned}$$

Let's look at the first equation first, since the boundary conditions look simpler than the initial conditions. Using the boundary conditions, we impose

$$X(0) = 0, \quad \text{and} \quad X(L) = 0.$$

So, we have to solve the ordinary differential equation boundary-value problem

$$\begin{cases} X'' - KX &= 0 \\ X(0) = 0, & X(L) = 0. \end{cases}$$

This is an eigenvalue problem for the function $X(x)$, with eigenvalue K.

There are several ways to establish that K must be negative to have nontrivial solutions to this problem. The most obvious one is to go through all the possibilities separately ($K > 0$, $K = 0$, and $K < 0$) and see that only the last option gives nontrivial (*i.e.*, nonzero) solutions of the eigenvalue problem. However, that approach requires that we are able to solve the ordinary differential equation for its general solution. Below we use an approach that does not require that we know a single solution. In that sense it is more general. It is the first hint of what we'll refer to as an **energy method** (due to its similarities with physical systems, where the computed quantity corresponds to the total energy of the system).

Multiply the equation by X to get

$$\begin{aligned} & XX'' - KX^2 = 0 \\ \Rightarrow \quad & \int_0^L \left(XX'' - KX^2\right) dx = 0 \\ \Rightarrow \quad & -\int_0^L \left(X'^2 + KX^2\right) dx = 0 \\ \Rightarrow \quad & \int_0^L \left(X'^2 + KX^2\right) dx = 0, \end{aligned}$$

where we've used integration by parts and the boundary conditions for $X(x)$.

- $K > 0$: if $K > 0$ then all terms in the integrand are positive or zero, and we need $X = 0$, $X' = 0$ to satisfy the equality. Thus in this case there are no nontrivial solutions.
- $K = 0$**:** in this case we need $X' = 0$ to satisfy the integral equality. Thus $X(x)$ is constant. In order to satisfy the boundary conditions, this constant has to be zero. Once again, there are no nontrivial solutions.
- $K = -\lambda^2 < 0$**, with** $\lambda \in \mathbb{R}^+$**:** in this case the integral equality can ostensibly be satisfied. We will continue working with this possibility and see what we get.

Our boundary-value problem becomes

$$\begin{cases} X'' + \lambda^2 X &= 0 \\ X(0) = 0, & X(L) = 0. \end{cases}$$

The general solution of the differential equation is

$$X = c_1 \cos \lambda x + c_2 \sin \lambda x.$$

Imposing the boundary conditions gives

$$\begin{aligned} X(0) = 0 \qquad &\Rightarrow \qquad c_1 = 0, \\ X(L) = 0 \qquad &\Rightarrow \qquad c_2 \sin \lambda L = 0. \end{aligned}$$

Since we don't want $c_2 = 0$, we require

$$\sin \lambda L = 0,$$

which **quantizes the eigenvalues**: the allowed λ values are

$$\lambda_n = \frac{n\pi}{L}, \quad n \in \mathbb{N}.$$

Note that $n = 0$ results in the trivial solution with $K = 0$, not surprisingly. Negative values of n lead to the same value of K and can be ignored.

The eigenfunctions corresponding to these eigenvalues

$$K_n = -\lambda_n^2 = -\frac{n^2\pi^2}{L^2}, \quad n = 1, 2, \ldots$$

are

$$X_n(x) = \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, \ldots,$$

where we have chosen $c_2 = 1$, since multiples of eigenfunctions remain eigenfunctions.

The moment is right to look at the time equation: we have

$$T_n'' + \left(\frac{c^2 n^2 \pi^2}{L^2}\right) T_n = 0, \quad n = 1, 2, \ldots.$$

We've given $T_n(t)$ a subindex n, to indicate that a different function will be obtained depending on which eigenvalue we use. We have

$$T_n(t) = \alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right), \quad n = 1, 2, \ldots,$$

where α_n and β_n are arbitrary constants.

In summary, we find the special class of solutions, one for each eigenvalue,

$$u_n(x,t) = X_n(x)T_n(t) = \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right)\right), \quad n = 1, 2, \ldots.$$

Since the wave equation is linear, we get a more general solution by superimposing all the special solutions we just found. You saw this coming a mile away, right?

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} u_n(x,t) \\ &= \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right)\right). \end{aligned}$$

In order to determine the constants α_n and β_n, we get to (finally) use the initial conditions.

At $t = 0$, we get for $u(x, 0)$

$$u(x,0) = \sum_{n=1}^{\infty} \alpha_n \sin\left(\frac{n\pi x}{L}\right) = f(x),$$

by the first initial conditions. Next we need to take a derivative of $u(x,t)$ with respect to t, and evaluate this at $t = 0$. Let's not worry for now about taking the derivative inside the sum. That's the topic of the next section. We get

$$u_t(x,0) = \sum_{n=1}^{\infty} \beta_n \frac{cn\pi}{L} \sin\left(\frac{n\pi x}{L}\right) = g(x).$$

Both series are Fourier sine series, requiring us to use odd extensions of the given initial conditions. The constants α_n and β_n are nothing but Fourier coefficients, and we have

$$\alpha_n = \frac{2}{L}\int_0^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx,$$

and

$$\begin{aligned} \beta_n \frac{cn\pi}{L} &= \frac{2}{L}\int_0^L g(x) \sin\left(\frac{n\pi x}{L}\right) dx \\ \Rightarrow \qquad \beta_n &= \frac{2}{n\pi c}\int_0^L g(x) \sin\left(\frac{n\pi x}{L}\right) dx, \end{aligned}$$

where these formulas for α_n and β_n are valid for $n = 1, 2, \ldots$.

At this point it appears we have solved the problem posed to us: we have a solution that we found by explicitly enforcing the boundary conditions and the initial conditions. However, we have only obtained a formal solution, as we have freely switched sums and derivatives, as is inherent to the separation of variables procedure. In the next section, we'll figure out under what conditions this formal solution does indeed solve our given initial-boundary-value problem.

Remark: the formal solution does not appear to be consistent with d'Alembert's solution. We've seen that *any* solution of the wave equation should be of the form $A(x + ct) + B(x - ct)$, independent of initial or boundary conditions. Indeed, the formal solution obtained above can be written as a linear combination of terms of this form by using trigonometric identities expressing the products of trigonometric functions as sums and differences of the same (the so-called Simpson formulas).

2. Justification

There's a number of things we need to justify: can we switch infinite sums and derivatives? Can we use Fourier expansions for $f(x)$ and $g(x)$? We'll now justify the above solution process by taking the solution we obtained formally, and showing under what conditions it is actually a solution.

We have

$$u(x,t) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right)\right)$$

In order to substitute this expression into the differential equation, we need to take partial derivatives of it, both with respect to x and with respect to t. This requires us to switch infinite sums and derivatives. If allowed, we'd get

$$u_x = \sum_{n=1}^{\infty} \frac{n\pi}{L} \cos\left(\frac{n\pi x}{L}\right)\left(\alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right)\right),$$
$$u_{xx} = -\sum_{n=1}^{\infty} \left(\frac{n\pi}{L}\right)^2 \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_n \cos\left(\frac{cn\pi t}{L}\right) + \beta_n \sin\left(\frac{cn\pi t}{L}\right)\right).$$

These differentiations are valid provided the obtained series are uniformly convergent. By the Weierstrass M-test, it suffices to find a constant C such that

$$(*) \begin{cases} \left|\dfrac{n^2\pi^2\alpha_n}{L^2}\right| \leq \dfrac{C}{n^2}, \\ \left|\dfrac{n^2\pi^2\beta_n}{L^2}\right| \leq \dfrac{C}{n^2}, \end{cases}$$

for $n = 1, 2, \ldots$. Note that if these conditions are satisfied that the series of u_x is automatically uniformly convergent, as its terms would approach zero as C/n^3. The calculation of u_t and u_{tt} leads to similar series. You should also verify that a similar requirement leads to the validity of the the Fourier expansions of $f(x)$ and $g(x)$.

Next, we wish to translate the conditions $(*)$ into conditions on the input data of the problem, namely $f(x)$ and $g(x)$. If we manage to do this, we will have obtained a criterion for the solvability of our original problem that will only involve easily verifiable conditions on $f(x)$ and $g(x)$!

Let's look at α_n. We have ($n = 1, 2, \ldots$ in what follows)

$$\alpha_n = \frac{2}{L}\int_0^L f(x)\sin\left(\frac{n\pi x}{L}\right)dx.$$

If we impose the compatibility conditions (requiring the initial conditions to match the boundary conditions at the edges) $f(0) = 0$, $f(L) = 0$, then

$$\begin{aligned}\alpha_n &= \frac{2}{L}\frac{L}{n\pi}\int_0^L f'(x)\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{2}{n\pi}\int_0^L f'(x)\cos\left(\frac{n\pi x}{L}\right)dx\end{aligned}$$

Assuming that $f'(x)$ is integrable has already bought us one factor of $1/n$. We're getting closer! If we impose the additional conditions $f''(0) = 0$, $f''(L) = 0$, we get

$$\begin{aligned}\alpha_n &= -\frac{2L}{n^2\pi^2}\int_0^L f''(x)\sin\left(\frac{n\pi x}{L}\right)dx \\ &= -\frac{2L^2}{n^3\pi^3}\int_0^L f'''(x)\cos\left(\frac{n\pi x}{L}\right)dx \\ &= \frac{2L^3}{n^4\pi^4}\int_0^L f^{(iv)}(x)\sin\left(\frac{n\pi x}{L}\right)dx.\end{aligned}$$

This immediately gives rise to

$$\begin{aligned}|\alpha_n| &\le \frac{2L^3}{(n\pi)^4}\int_0^L |f^{(iv)}(x)|dx \\ \Rightarrow \qquad \left|\frac{n^2\pi^2\alpha_n}{L^2}\right| &\le \frac{C}{n^2} = \frac{2L}{\pi^2 n^2}\int_0^L |f^{(iv)}(x)|dx,\end{aligned}$$

thus it suffices to use

$$C = \frac{2L}{\pi^2}\int_0^L |f^{(iv)}(x)|dx.$$

A similar calculation places restrictions on $g(x)$. All this leads to the following theorem.

Theorem 13 *Let $f(x)$ have a fourth derivative which is absolutely integrable on $[0, L]$, and let $g(x)$ have a third derivative which is absolutely integrable on $[0, L]$, such that*

$$\begin{cases} f(0) = f(L) = f''(0) = f''(L) = 0 \\ g(0) = g(L) = g''(0) = g''(L) = 0. \end{cases}$$

Then the formal solution obtained using separation of variables solves the initial-boundary-value problem

$$\begin{cases} u_{tt} = c^2 u_{xx} & 0 < x < L,\ t > 0 \\ u(x,0) = f(x) & 0 \le x \le L \\ u_t(x,0) = g(x) & 0 \le x \le L \\ u(0,t) = 0 = u(L,t) & t \ge 0 \end{cases}.$$

As announced, the theorem list conditions on the input functions that are easy to check. If these conditions are satisfied, we may proceed with our regular methods to find a formal solution. The theorem then guarantees the validity of this formal solution.

3. Well-posedness

We could study well-posedness for the same initial-boundary-value problem as in the previous section, but what would be the fun of that? Let's look at a slightly more general problem instead. Consider the problem

$$\begin{cases} u_{tt} + ku_t = c^2 u_{xx} & 0 < x < L,\ t > 0 \\ u(x,0) = f(x) & 0 \le x \le L \\ u_t(x,0) = g(x) & 0 \le x \le L \\ u(0,t) = 0 = u(L,t) & t \ge 0 \end{cases}.$$

Here $k > 0$, representing the damping in a string. This problem can be solved in an almost identical way as the previous one, using separation of variables. Let's consider its well-posedness instead.

Existence of a solution

As stated, using separation of variables, a solution to this problem can be constructed. Perhaps we have to impose some regularity conditions on $f(x)$ and $g(x)$ to make this rigorous. So be it. Thus a solution exists.

Uniqueness of a solution

Having established that a solution exists, let's prove that it is unique. Proving uniquess of a solution is almost always done the same way: by contradiction. We'll assume there's another solution, which is of course assumed to be different. If $u_1(x,t)$ and $u_2(x,t)$ represent our two distinct solutions, what is the problem that is satisfied by $v = u_1 - u_2$?

- **Partial differential equation:** we have

$$\begin{aligned} v_{tt} + kv_t &= (u_1 - u_2)_{tt} + k(u_1 - u_2)_t \\ &= u_{1tt} + ku_{1t} - (u_{2tt} + ku_{2t}) \\ &= c^2 u_{1xx} - c^2 u_{2xx} \\ &= c^2 (u_1 - u_2)_{xx} \\ &= c^2 v_{xx}, \end{aligned}$$

  for $0 < x < L$ and $t > 0$.

- **Initial conditions:** this is easier. We get

$$v(x,0) = (u_1 - u_2)(x,0) = u_1(x,0) - u_2(x,0) = f(x) - f(x) = 0,$$

  and

$$v_t(x,0) = (u_{1t} - u_{2t})(x,0) = u_{1t}(x,0) - u_{2t}(x,0) = g(x) - g(x) = 0.$$

- **Boundary conditions:** similarly,

$$v(0,t) = (u_1 - u_2)(0,t) = u_1(0,t) - u_2(0,t) = 0,$$

  and

$$v(L,t) = (u_1 - u_2)(L,t) = u_1(L,t) - u_2(L,t) = 0.$$

In summary, we find that $v(x,t) = u_1(x,t) - u_2(x,t)$ satisfies the initial-boundary-value problem

$$\begin{cases} v_{tt} + kv_t = c^2 v_{xx} & 0 < x < L,\ t > 0 \\ v(x,0) = 0 & 0 \le x \le L \\ v_t(x,0) = 0 & 0 \le x \le L \\ v(0,t) = 0 = v(L,t) & t \ge 0 \end{cases}.$$

This is a somewhat easier problem than the original one, since the initial conditions are homogeneous. We'll now show that the only solution of this problem is the zero solution $v(x,t) = 0$, so that $u_1(x,t) = u_2(x,t)$, proving uniqueness of the solution of the original problem. We can't use separation of variables to show that $v = 0$. The whole point is to argue that the solution produced by separation of variables is the only solution. We need to rule out that another method might produce another solution.

We'll use an energy argument instead. Integration by parts is our main weapon[1].

$$
\begin{aligned}
& & v_{tt} + kv_t &= c^2 v_{xx} \\
\Rightarrow & & v_t v_{tt} + kv_t^2 &= c^2 v_{xx} v_t \\
\Rightarrow & & \int_0^L \left(v_t v_{tt} + kv_t^2\right) dx &= c^2 \int_0^L v_{xx} v_t dx \\
\Rightarrow & & \int_0^L \left(\frac{\partial}{\partial t}\left(\frac{1}{2}v_t^2\right) + kv_t^2\right) dx &= -c^2 \int_0^L v_x v_{xt} dx \\
\Rightarrow & & \int_0^L \left(\frac{\partial}{\partial t}\left(\frac{1}{2}v_t^2\right) + kv_t^2\right) dx &= -\frac{c^2}{2} \int_0^L \frac{\partial}{\partial t}\left(v_x^2\right) dx \\
\Rightarrow & & \frac{d}{dt}\int_0^L \left(\frac{1}{2}v_t^2 + \frac{c^2}{2}v_x^2\right) dx &= -k \int_0^L v_t^2 dx \\
\Rightarrow & & \frac{dE}{dt} &= -k \int_0^L v_t^2 dx \\
\Rightarrow & & \frac{dE}{dt} &\leq 0 \\
\Rightarrow & & E(t) &\leq E(0).
\end{aligned}
$$

Here

$$E(t) = \int_0^L \left(\frac{1}{2}v_t^2 + \frac{c^2}{2}v_x^2\right) dx$$

can be interpreted as the total energy of the system, comprised of both kinetic (first term) and potential (second term) energy. We have just shown that the energy is always decreasing as a consequence of the damping, implying that the energy at any time $t > 0$ is less than or equal to its initial amount.

At $t = 0$, we have $v(x, 0) = 0$, from which $v_x(x, 0) = 0$. The other initial condition states that $v(x, 0) = 0$. Thus,

$$E(0) = \int_0^L \left(\frac{1}{2}v_t^2(x, 0) + \frac{c^2}{2}v_x^2(x, 0)\right) dx = 0.$$

Thus

$$E(t) \equiv 0,$$

for all $t \geq 0$. Since the integrand of the energy integral is always non-negative, this is possible only if

$$\frac{1}{2}v_t^2 + \frac{c^2}{2}v_x^2 \equiv 0 \quad \Rightarrow \quad v_x \equiv 0, v_t \equiv 0 \quad \Rightarrow \quad v(x, t) = \text{constant} = 0,$$

where the last equality follows from the boundary or initial conditions. Thus

$$u_1(x, t) = u_2(x, t),$$

for all $x \in [0, L]$ and all $t \geq 0$. Thus the solution of the original problem is unique.

[1] No! Integration by parts *and* the boundary conditions are our main weapons!

No sensitive dependence on the input data

The last thing on the list to establish well-posedness is to verify that the solution depends smoothly on the input functions $f(x)$, $g(x)$. Suppose that $u_1(x,t)$ is the solution corresponding to the functions $f_1(x)$ and $g_1(x)$, whereas $u_2(x,t)$ is the solution corresponding to the functions $f_2(x)$ and $g_2(x)$. We have to show that if f_1 and f_2 are close in some sense, and if g_1 and g_2 are close in perhaps another sense, then u_1 and u_2 are close, in perhaps a third sense, for all $t > 0$. Part of this discussion will be to figure out what we mean by close for all these functions. Figuring this out is always an issue with problems like this. Ideally, the flow of the calculation will dictate what sense of close to use. This is indeed the case below.

When we studied the uniqueness of a solution, we were able to avoid the specific form of the solution, resulting in a more general argument. We can't do that here, and we need to rely on the specific solution form quite a bit. To this end, let's do this for the case $k = 0$, where we know the solution from before. You should check that including $k \neq 0$ only results in exponentially damped terms, and the absence of sensitive dependence on the input data is actually easier to show.

Using the solution formula with $k = 0$, we get

$$\begin{aligned} u_1 - u_2 &= \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_{1n}\cos\left(\frac{cn\pi t}{L}\right) + \beta_{1n}\sin\left(\frac{cn\pi t}{L}\right)\right) - \\ &\qquad \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left(\alpha_{2n}\cos\left(\frac{cn\pi t}{L}\right) + \beta_{2n}\sin\left(\frac{cn\pi t}{L}\right)\right) \\ &= \sum_{n=1}^{\infty} \sin\left(\frac{n\pi x}{L}\right)\left((\alpha_{1n} - \alpha_{2n})\cos\left(\frac{cn\pi t}{L}\right) + (\beta_{1n} - \beta_{2n})\sin\left(\frac{cn\pi t}{L}\right)\right) \\ \Rightarrow \quad |u_1 - u_2| &\leq \sum_{n=1}^{\infty} \left(|\alpha_{1n} - \alpha_{2n}| + |\beta_{1n} - \beta_{2n}|\right), \end{aligned}$$

where

$$\alpha_{1n} - \alpha_{2n} = \frac{2}{L}\int_0^L (f_1 - f_2)\sin\left(\frac{n\pi x}{L}\right) dx,$$

and

$$\beta_{1n} - \beta_{2n} = \frac{2}{n\pi c}\int_0^L (g_1 - g_2)\sin\left(\frac{n\pi x}{L}\right) dx.$$

Let's assume that solutions exist corresponding to the input data $f_1(x)$ and $g_1(x)$, as well as to the $f_2(x)$ and $g_2(x)$. Otherwise the above examinations don't make much sense. Assuming the existence of solutions, we get to impose the assumptions of our theorem, namely that $f_1(x)$, $f_2(x)$ ($g_1(x)$, $g_2(x)$) has four (three) absolutely integrable derivatives. Furthermore, these functions and their second derivatives vanish as the endpoints $x = 0$ and

$x = L$. Using these assumptions we obtain

$$\begin{aligned}
\alpha_{1n} - \alpha_{2n} &= \frac{2}{L}\left(\frac{L}{n\pi}\right)\int_0^L (f_1' - f_2')\cos\left(\frac{n\pi x}{L}\right)dx \\
&= -\frac{2}{L}\left(\frac{L}{n\pi}\right)^2\int_0^L (f_1'' - f_2'')\sin\left(\frac{n\pi x}{L}\right)dx \\
\Rightarrow \qquad |\alpha_{1n} - \alpha_{2n}| &\leq \frac{2L}{n^2\pi^2}\int_0^L |f_1'' - f_2''|dx.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\beta_{1n} - \beta_{2n} &= \frac{2}{n\pi c}\int_0^L (g_1 - g_2)\sin\left(\frac{n\pi x}{L}\right)dx \\
&= \frac{2}{n\pi c}\left(\frac{L}{n\pi}\right)\int_0^L (g_1' - g_2')\cos\left(\frac{n\pi x}{L}\right)dx \\
&= \frac{2L}{n^2\pi^2 c^2}\int_0^L (g_1' - g_2')\cos\left(\frac{n\pi x}{L}\right)dx \\
\Rightarrow \qquad |\beta_{1n} - \beta_{2n}| &\leq \frac{2L}{n^2\pi^2 c^2}\int_0^L |g_1' - g_2'|dx.
\end{aligned}$$

Putting these results together, we get

$$\begin{aligned}
|u_1 - u_2| &\leq \sum_{n=1}^{\infty}\left(|\alpha_{1n} - \alpha_{2n}| + |\beta_{1n} - \beta_{2n}|\right) \\
&\leq \sum_{n=1}^{\infty}\left(\frac{2L}{n^2\pi^2}\int_0^L |f_1'' - f_2''|dx + \frac{2L}{n^2\pi^2}\int_0^L |f_1'' - f_2''|dx\right) \\
&= \frac{2L}{\pi^2}\int_0^L |f_1'' - f_2''|dx\sum_{n=1}^{\infty}\frac{1}{n^2} + \frac{2L}{\pi^2 c^2}\int_0^L |g_1' - g_2'|dx\sum_{n=1}^{\infty}\frac{1}{n^2} \\
&= \frac{L}{3}\int_0^L |f_1'' - f_2''|dx + \frac{L}{3c^2}\int_0^L |g_1' - g_2'|dx.
\end{aligned}$$

We see that $u_1 - u_2$ will stay small, as long as

$$\int_0^L |f_1'' - f_2''|dx \quad \text{and} \quad \int_0^L |g_1' - g_2'|dx$$

stay small. One way to accomplish this is to ensure that

$$F_2 = \max_{x\in[0,L]} |f_1'' - f_2''|$$

and

$$G_1 = \max_{x\in[0,L]} |g_1' - g_2'|$$

remain small. Then

$$|u_1 - u_2| \leq \frac{L}{3} L F_2 + \frac{L}{3c^2} L G_1 = \frac{L^2}{3}\left(F_2 + \frac{1}{c^2} G_1\right)$$

remains small for all $t > 0$.

In summary, we have established that the original initial-boundary-value problem is well-posed, provided that the input functions are so that the formal solution obtained using separation of variables is indeed a solution.

Note that demanding that F_2 and G_1 are small also implies that

$$F_1 = \max_{x\in[0,L]} |f_1' - f_2'|, \quad F_0 = \max_{x\in[0,L]} |f_1 - f_2|,$$

and

$$G_0 = \max_{x\in[0,L]} |g_1 - g_2|$$

are small, provided that $\delta = |f_1'(0) - f_2'(0)|$ is small. Indeed, if

$$\begin{aligned}
& -\epsilon < f_1'' - f_2'' < \epsilon \\
\Rightarrow \quad & -\int_0^x \epsilon dx < \int_0^2 (f_1'' - f_2'') dx < \int_0^x \epsilon dx \\
\Rightarrow \quad & -\epsilon x < (f_1' - f_2') - (f_1'(0) - f_2'(0)) < \epsilon x \\
\Rightarrow \quad & -\epsilon L < (f_1' - f_2') - (f_1'(0) - f_2'(0)) < \epsilon L \\
\Rightarrow \quad & -\epsilon L - \delta < f_1' - f_2' < \epsilon L + \delta \\
\Rightarrow \quad & |f_1' - f_2'| < \epsilon L + \delta \\
\Rightarrow \quad & \max_{x\in[0,L]} |f_1' - f_2'| < \epsilon L + \delta,
\end{aligned}$$

which is small. Similar calculations show that $|f_1 - f_2|$ and $|g_1 - g_2|$ are small as well.

4.3 Exercises

1. Find a formal solution for the following IVP with boundary conditions:

$$\begin{cases}
u_{tt} = u_{xx} + u_{yy}, \quad 0 < x < a, \ 0 < y < b, \\
u(x, y, 0) = \phi(x, y), \ u_t(x, y, 0) = \varphi(x, y), \quad 0 \leq x \leq a, \ 0 \leq y \leq b, \\
u(0, y, t) = u(a, y, t) = 0, \quad 0 \leq y \leq b, \\
u(x, 0, t) = u(x, b, t) = 0, \quad 0 \leq x \leq a.
\end{cases}$$

Chapter 5

General remarks about linear evolution equations

An evolution equation is any partial differential equation (vector or scalar) that may be solved for the time derivative of the dependent variables. We'll only look at autonomous evolution equations, which do not depend on the time or spatial variables explicitly. Thus, in one spatial dimension such equations may be written as

$$\boldsymbol{u}_t = \boldsymbol{F}(\boldsymbol{u}, \boldsymbol{u}_x, \ldots, \boldsymbol{u}_{Nx}),$$

where N is the highest order of the x-derivatives. Generalizing this to multiple space dimensions is straightforward. In this chapter, we'll look at linear evolution equations in some detail. Such equations are of the form

$$\boldsymbol{u}_t = \boldsymbol{L}(\boldsymbol{u}, \boldsymbol{u}_x, \ldots, \boldsymbol{u}_{Nx}),$$

where L is linear in its arguments. In other words

$$\boldsymbol{u}_t = \sum_{k=0}^{N} \alpha_k \boldsymbol{u}_{kx},$$

and the α_k (k=0,1, ..., N) are constants. Here $\boldsymbol{u}$ could be a vector function, in which case $\boldsymbol{F}$ and $\boldsymbol{L}$ are vectors with the same dimension as $\boldsymbol{u}$. In what follows, we'll consider the partial differential equation with vanishing boundary conditions as $x \to \pm\infty$:

$$\lim_{x \to \pm\infty} u = 0.$$

We'll impose similar conditions on whatever number of derivatives we need, in order for the solution formulas below to make sense. A similar treatment could be used for periodic boundary conditions. In that case, we'd find solutions expressed as Fourier series, whereas now we'll need Fourier transforms.

5.1 The linear dispersion relationship

In what follows, we work with the scalar case for simplicity, unless otherwise stated. Since our linear evolution equation has constant coefficients, we may look for fundamental solutions of the form

$$u = Ae^{ikx-i\omega t},$$

where k and ω are constants, to be determined. Then

$$\begin{aligned} u_t &= -i\omega u, \\ u_x &= iku, \\ u_{xx} &= (ik)^2 u, \\ &\dots \end{aligned}$$

Substituting this into the partial differential equation, we obtain

$$\begin{aligned} -i\omega u &= L(u, iku, (ik)^2 u, \dots, (ik)^N u) \\ &= uL(1, ik, (ik)^2, \dots, (ik)^N), \end{aligned}$$

since L is homogeneous and linear. Thus

$$\omega = \omega(k) = iL(1, ik, (ik)^2, \dots, (ik)^N).$$

We get an algebraic relationship between k and ω. This is the **linear dispersion relationship**. For any k, it gives ω as a function of k. A physical interpretation of the dispersion relationship is that it gives the time response of the system for a given spatial disturbance. You may want to work out what you get in the vector case. You'll see that there are as many branches of the time response $\omega(k)$, as the order in time of the vector partial differential equation.

5.2 Superposition

We have found that the fundamental solutions are of the form

$$u_k = A(k)e^{ikx-i\omega(k)t},$$

where we have shown the dependence on k explicitly. The most general solution to be obtained from all these fundamental solutions is a linear superposition of all of them. Since there is no restriction on k, other than that it's real, we get

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx-i\omega(k)t}dk.$$

The factor 2π is chosen entirely for convenience, but could be omitted. In order to determine $A(k)$, we evaluate the above at $t = 0$. We get

$$u(x,0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx}dk = \mathcal{F}^{-1}[A(k)](x)$$
$$\Rightarrow \qquad A(k) = \mathcal{F}[u(x,0)](k),$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and the inverse Fourier transform. We see that knowing the dispersion relationship of a linear evolution equation allows us to solve the initial-value problem for this equation for all initial data for which the Fourier transform is defined. The schematic of this solution technique is shown in Fig. 5.1. As before, you should try to see what happens to the vector case.

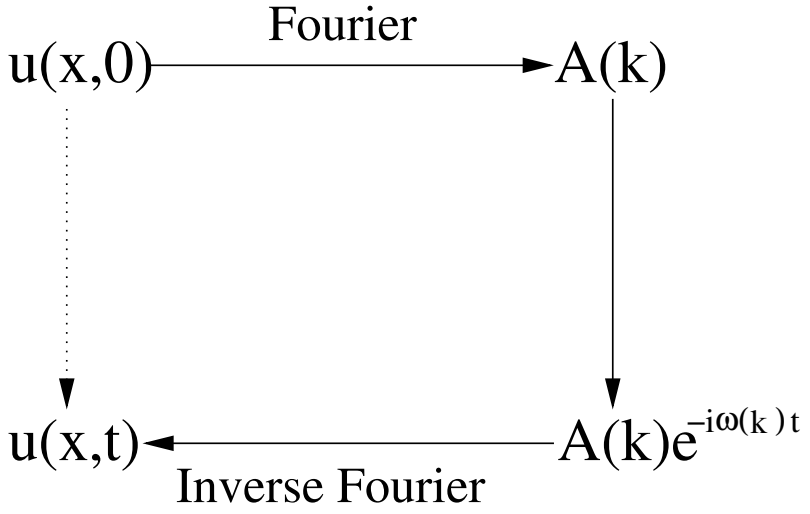

Figure 5.1: The schematic of using Fourier transforms for solving initial-value evolution partial differential equations. (2.1)

Thus, if our initial condition is absolutely integrable (so that its Fourier transform is defined), we have found a unique solution of the initial-value problem for our given evolution equation. What can be concluded from this? Specifically, what conclusions follow from the knowledge of the dispersion relationship? The answer is "quite a few". We'll show in the next section that nearly all important information about the equation is contained in the linear dispersion relationship.

5.3 Consequences of the dispersion relationship

The dispersion relationship is defined as

$$\omega(k) = iL(1, ik, (ik)^2, \ldots, (ik)^N).$$

Thus typically, it is a complex-valued function even though $k \in \mathbb{R}$. Let's denote

$$\omega(k) = \omega_R(k) + i\omega_I(k),$$

where $\omega_R(k)$ and $\omega_I(k)$ are real-valued functions.

a) If $\omega_I(k_0) > 0$ for some real k_0, then this corresponds to a mode that is exponentially growing in time. Indeed, the fundamental solution corresponding to k_0 is

$$A(k_0)e^{ik_0x - i\omega_0 t} = A(k_0)e^{ik_0x - i\omega_R(k_0)t}e^{\omega_I(k_0)t},$$

which is exponentially growing in t. This implies that the problem is unstable. The most unstable mode is the one that maximizes $\omega_I(k)$.

b) If $\omega_I \to \infty$ for any value of k (including $k \to \pm\infty$), then the problem is ill-posed since the exponential growth is unbounded, and any perturbation in the system parameters leads to unlimited growth.

c) If $\omega_I(k) < 0$ for all real k, then the problem is **dissipative**, and the equation is asymptotically stable.

d) By parseval's relation:

$$\begin{aligned}\int_{-\infty}^{\infty} |u^2|dx &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{u}(k)|^2 dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |A(k)|^2 e^{2\omega_I(k)t} dk,\end{aligned}$$

since

$$\hat{u}(k) = A(k)e^{ik_0x - i\omega(k)t}.$$

Let's call $\int_{-\infty}^{\infty} |u^2|dx$ the **energy**. Then the energy is conserved only if $\omega(k)$ is real for real k. In this case there is no growth or decay as a consequence of loss of energy. We'll show shortly that decay is still possible, but a different process will determine it.

e) If there is a term

$$\partial_x^n u$$

in the differential equation, then its contribution to the dispersion relationship is

$$i(ik)^n = k^n i^{n+1}.$$

If $n > 1$ is odd, then this is real. Such a term is called **dispersive**. If $n > 1$ is even, such a term results in an imaginary contribution to $\omega(k)$, and the term is dissipative or amplifying, depending on its sign.

There are more consequences, but they require a bit more work. Let's give them their own section.

5.4 Asymptotics of the solution

If there is growth or decay due to dissipation, the asymptotics of the solution is dominated by this behavior. If these two processes are absent, how can we determine the behavior of the solution as $t \to \infty$ if it is either purely propagating, or dispersive?

Assume that $\omega = \omega(k) = \omega_R(k)$ is real for real k. Then the solution of the problem is given by (see above)

$$\left\{ \begin{array}{rcl} u(x,t) & = & \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx-i\omega(k)t}dk, \\ A(k) & = & \int_{-\infty}^{\infty} u(x,0)e^{-ikx}dx. \end{array} \right.$$

Note that if we were dealing with periodic boundary conditions (so that the above would be a Fourier series instead of a Fourier integral), no asymptotic state would be possible: there is no place for the energy (which is conserved) to go. However, in an infinite domain the energy can dissapear to infinity. We'll show that this is indeed what happens.

The speed of a single wave is

$$c_p(k) = \frac{\omega(k)}{k},$$

which is referred to as the **phase velocity**. If $c_p(k)$ is nonconstant for different k, the problem is dispersive. This will happen if more than first derivatives in x are present in the equation. A fundamental solution then moves according to

$$A(k)e^{ik(x-c_p(k)t)}.$$

If the problem is not dispersive then

$$c_p = \frac{\omega(k)}{k} = c = \text{constant},$$

and $\omega = ck$. Then there is only propagation and, indeed,

$$u_t = cu_x.$$

The problem is dispersive if $\omega''(k) \neq 0$. Then different waves travel at different speeds.

Let's look briefly at the method of **stationary phase** to examine the long-time behavior of such a solution. Consider

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx-i\omega(k)t}dk.$$

For large t, the phases $kx - \omega(k)t$ corresponding to wavenumber k and wavenumber $k + \Delta k$ will be very different:

$$\begin{array}{rl} k: & kx - \omega(k)t \\ k+\Delta k: & (k+\Delta k)x - \omega(k+\Delta k)t \approx kx + x\Delta k - \omega(k)t - \omega'(k)t\Delta k. \end{array}$$

The difference is

$$(x - \omega'(k)t)\Delta k,$$

which is large for large t. This leads to destructive interference of such neighboring waves, unless $x \approx \omega'(k)t$, in which case phases remain approximately constant. In such directions no destructive phase interference occurs. This is the direction in which information propagates.

More explicitly, consider a case where $A(k)$ is sharply peaked at $k = k_0$, and decays rapidly to zero away from $k = k_0$. We have

$$\begin{aligned} u(x,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx-i\omega(k)t}dk \\ &= \frac{e^{ik_0x-i\omega(k_0)t}}{2\pi}\int_{-\infty}^{\infty} A(k)e^{i(k-k_0)x-i(\omega(k)-\omega(k_0))t}dk \\ &= \frac{e^{ik_0x-i\omega(k_0)t}}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ix\Delta k-i(\omega'(k_0)\Delta k+\mathcal{O}((\Delta k)^2))t}dk \\ &= \frac{e^{ik_0x-i\omega(k_0)t}}{2\pi}\int_{-\infty}^{\infty} A(k)e^{it\Delta k(x/t-\omega'(k_0)+\mathcal{O}(\Delta k))}dk. \end{aligned}$$

By a suitable modification of the Riemann-Lebesgue lemma, this integral $\to 0$ as $t \to \infty$, unless

$$x \approx \omega'(k_0)t = c_g(k_0)t.$$

The velocity $c_g(k_0) = \omega'(k)$ is called the **group velocity**. It is the velocity at which the information corresponding to wavenumber k_0 propagates. The equation

$$x = c_g(k_0)t + o(t)$$

determines an approximate ray along which the wave number k_0 dominates. This is illustrated in Fig. 5.2.

Thus different wave numbers dominate along different rays. After a sufficiently long time, this will result in dispersive decay, as opposed to dissipative decay of the solution.

The energy contribution from modes with wavenumbers between $k - \Delta k$ and $k + \Delta k$ is

$$\frac{1}{2\pi}\int_{k-\Delta k}^{k+\Delta k} |A(k)|^2 dk.$$

As Fig. 5.2 demonstrates this energy is spreading over a region that is growing linearly in t. Since this energy is conserved, with density $|u(x,t)|^2$, we expect

$$|u(x,t)|^2 \sim \frac{1}{t} \quad \Rightarrow \quad u(x,t) \sim \frac{1}{t^{1/2}}.$$

Thus the amplitude of the solution should decay as a the inverse of a square root. This estimate can be made rigorous, using the method of stationary phase or the method of steepest decent. We've barely hinted at the method of stationary phase here.

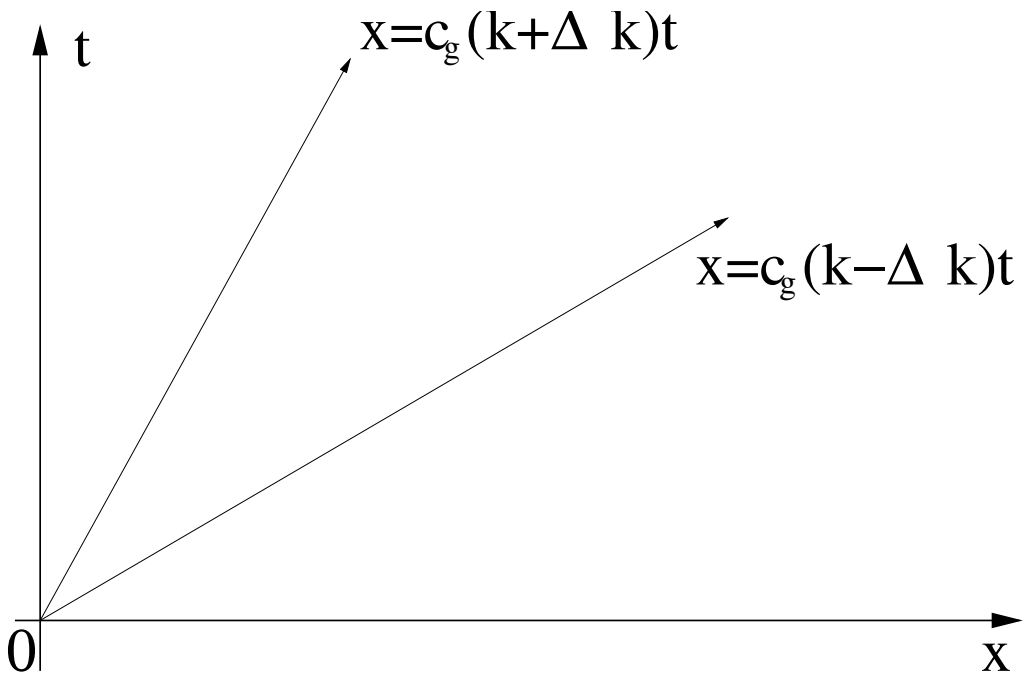

Figure 5.2: A simple cartoon illustrating the idea of different wavenumbers dominating along different directions in the (x, t)-plane. (2.1)

5.5 Examples

Example: Consider the equation

$$u_t = cu_{xx}.$$

Its dispersion relationship is

$$-i\omega = c(ik)^2 \quad \Rightarrow \quad \omega = -ick^2.$$

Its imaginary part is

$$\omega_I = -ck^2.$$

We see that the problem is ill-posed if $c < 0$. If $c > 0$ (heat equation) then the energy is given by

$$\int_{-\infty}^{\infty} |u|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |A(k)|^2 e^{-2ck^2 t} dk \quad \to 0,$$

as $t \to \infty$.

Example: Consider the equation

$$u_t = u_{xxx},$$

known as the linear Korteweg-deVries equation. Its dispersion relationship is

$$-i\omega = (ik)^3 = -ik^3 \quad \Rightarrow \quad \omega = k^3.$$

The equation is dispersive, without dissipation or amplification. Its phase velocity is

$$c_p = \frac{\omega}{k} = k^2,$$

and its group velocity is

$$c_g = \omega' = 3k^2 = 3c_p.$$

Example: Consider the linear Schrödinger equation

$$i\phi_t = \phi_{xx}.$$

This is not a partial differential equation with real coefficients, so it doesn't really fit into our framework. We'll be flexible for once. The dispersion relationship is given by

$$i(-i\omega) = (ik)^2 \quad \Rightarrow \quad \omega = -k^2.$$

As in the previous example, this is real-valued for real k and the equation is purely dispersive, without dissipative loss. We have

$$c_p = -k, \quad c_g = -2k.$$

5.6 Exercises

1. Some partial differential equations admit self-similar solutions of the form

   $$u(x,t) = t^{-p} f(\eta), \quad \eta = x/t^q,$$

   where p and q are constant, and $f(\eta)$ satisfies an ordinary differential equation. In other words, the partial differential equation is invariant under the transformation

   $$T_b(x,t,u) = (b^q x, bt, b^{-p} u),$$

   where b is scalar. The set of all such transformations forms a group, and is the simplest nontrivial subgroup of the symmetry group of the differential equation. Such self-similar solutions often lie outside the function space of interest: often they are not integrable, *etc.* Nevertheless, in many problems, the asymptotic ($t \to \infty$) solution becomes approximately self-similar locally, modulated by a slowly varying function that depends on the initial data. Thus we see that the time evolution of the initial data is separated in two distinct parts: (1) local behavior, determined entirely by the differential equation, through the self-similar solution. This part is not influenced by the initial data, and (2) a slow modulation of this local behavior, determined by the initial data.

   Consider the linear Schrödinger equation

   $$i\psi_t + \psi_{xx} = 0.$$

(a) What are the conditions on p and q for this equation to have solutions of the form $u = t^{-p} f(\eta)$, with $\eta = x/t^q$ that satisfy an ordinary differential equation? (b) Solve this differential equation with $p = 1/2$. (c) Show that the solution you have obtained of the linear Schrödinger equation cannot be captured by Fourier transform methods.

Chapter 6

The heat equation

Let's turn our attention to parabolic problems, more specifically the heat equation

$$u_t = au_{xx},$$

with one spatial dimension, or

$$u_t = a\Delta u,$$

in an arbitrary number of dimensions. Here $\Delta = \partial_x^2 + \partial_y^2$ in two dimensions. This definition of the Laplacian generalizes to three and more dimensions in an obvious manner.

6.1 A homogeneous initial-boundary-value problem

Consider the problem

$$\begin{cases} u_t &= au_{xx}, & t > 0, 0 < x < L, \\ u(0,t) &= 0, & t \geq 0, \\ u(L,t) &= 0, & t \geq 0, \\ u(x,0) &= f(x), & 0 \leq x \leq L. \end{cases}$$

Here L and a are positive constants. Using separation of variables, we set

$$u(x,t) = X(x)T(t),$$

so that

$$\Rightarrow \qquad \begin{aligned} T'X &= aX''T \\ \frac{T'}{aT} &= \frac{X''}{X} = c, \end{aligned}$$

where c is our separation constant. Thus

$$X'' - cX = 0,$$

with boundary conditions

$$X(0) = 0 = X(L).$$

As before, we can use an energy integral argument to find that we only need to consider $c = -\lambda^2 < 0$. After imposing the boundary conditions, we find that

$$X_n = \sin \frac{n\pi x}{L} = \sin \lambda_n x.$$

Here we have defined

$$\lambda_n = \frac{n\pi}{L}, \quad n = 1, 2, \ldots.$$

Turning to the equation for T, we find that

$$\frac{T_n'}{aT_n} = -\lambda_n^2 \quad \Rightarrow \quad T_n = \alpha_n e^{-a\lambda_n^2 t},$$

again for $n = 1, 2, \ldots$. Further, α_n denotes an arbitrary constant, possibly different for every n.

A superposition of the fundamental solutions

$$u_n(x,t) = \alpha_n e^{-a\lambda_n^2 t} \sin \lambda_n x$$

gives

$$u(x,t) = \sum_{n=1}^{\infty} \alpha_n e^{-a\lambda_n^2 t} \sin \lambda_n x.$$

As before, using the initial condition gives us formulas for the α_n, $n = 1, 2, \ldots$:

$$f(x) = \sum_{n=1}^{\infty} \alpha_n \sin \lambda_n x \quad \Rightarrow \quad \alpha_n = \frac{2}{L} \int_0^L f(x) \sin \lambda_n x dx.$$

We have constructed a formal solution to the heat equation initial-boundary-value problem. This will be an actual solution, provided we can justify our Fourier series and the termwise differentiation. This proceeds along similar lines to what we did for the wave equation, so it will not be repeated here. We should just remark that for any $t > 0$, it suffices to have $\alpha_n < M$ (for some positive M), without needing any decay with increasing n. This is because the exponential decay in time provides the necessary decay of the Fourier coefficients to guarantee absolute and uniform convergence of the Fourier series. In summary, we have the following theorem.

Theorem 14 *If $f(x)$ is bounded and integrable, the formal Fourier series solution for $u(x,t)$ gives an actual solution of the heat equation initial-boundary-value problem. If $f(x)$ has a continuous second derivative with $f(0) = 0 = f(L)$, then $u(x,t)$ thus found is uniformly convergent and continuous for $t \geq 0$, $0 \leq x \leq L$, and satisfies the initial condition.*

6.2 A nonhomogeneous initial-boundary-value problem

Next, we consider the heat equation with an external forcing term:

$$\begin{cases} u_t &= au_{xx} + F(x,t), & t > 0, 0 < x < L, \\ u(0,t) &= 0, & t \geq 0, \\ u(L,t) &= 0, & t \geq 0, \\ u(x,0) &= f(x), & 0 \leq x \leq L. \end{cases}$$

The function $F(x,t)$ represents an external time- and space-dependent forcing. Suppose it is given by

$$F(x,t) = \sum_{n=1}^{\infty} F_n(t) \sin \lambda_n x.$$

This form of $F(x,t)$ implies that we've used an odd extension of $F(x,t)$ for its Fourier series representation. This is consistent with the boundary conditions: an external forcing at the boundaries will likely result in a function $u(x,t)$ which is not able to satisfy the boundary conditions.

Since we know how to solve the homogeneous solution, it suffices to find one particular solution. Once we have a particular solution, the general solution is given by the superposition of the homogeneous and particular solutions. Let's use variation of parameters to construct a particular solution. Thus, we'll look for a particular solution of the form

$$u_p(x,t) = \sum_{n=1}^{\infty} c_n(t) e^{-a\lambda_n^2 t} \sin \lambda_n x.$$

The temporal functions $c_n(t)$ replace the previously constant α_n. They are our "varying parameters". Since they are to be determined, we might as well save ourselves some writing and bundle them with the exponentials. This means we can look for

$$u_p(x,t) = \sum_{n=1}^{\infty} d_n(t) \sin \lambda_n x.$$

Substituting this in the equation, and formally calculating derivatives, gives

$$\sum_{n=1}^{\infty} d_n' \sin \lambda_n x = -a \sum_{n=1}^{\infty} \lambda_n^2 d_n \sin \lambda_n x + \sum_{n=1}^{\infty} F_n(t) \sin \lambda_n x.$$

Using orthogonality of the sine functions, we get

$$d_n' + a\lambda_n^2 d_n = F_n(t), \quad n = 1, 2, \ldots,$$

so that

$$\begin{aligned} \left(e^{a\lambda_n^2 t} d_n\right)' &= e^{a\lambda_n^2 t} F_n(t) \\ \Rightarrow \quad e^{a\lambda_n^2 t} d_n &= \int_0^t e^{a\lambda_n^2 \tau} F_n(\tau) d\tau \\ \Rightarrow \quad d_n &= e^{-a\lambda_n^2 t} \int_0^t e^{a\lambda_n^2 \tau} F_n(\tau) d\tau \\ &= \int_0^t e^{-a\lambda_n^2 (t-\tau)} F_n(\tau) d\tau. \end{aligned}$$

Note that we only care about finding *one* particular solution, so that we may ignore any constants during the integration process, as we've done here.

Thus the general solution is

$$u(x,t) = \sum_{n=1}^{\infty} \alpha_n e^{-a\lambda_n^2 t} \sin \lambda_n x + \sum_{n=1}^{\infty} \sin \lambda_n x \int_0^t e^{-a\lambda_n^2 (t-\tau)} F_n(\tau) d\tau.$$

At this point, the α_n's may be determined from the initial conditions. We've been smart enough to choose a particular solution that vanishes at $t = 0$, so that the formulas for the α_n in terms of the initial data are identical to those for the homogeneous case. As before, we can find conditions on $F(x,t)$ and $f(x)$ (our input functions) that ensure that the formal solution is indeed a solution.

6.3 Nonconstant boundary conditions

Suppose that the previous problem is modified so that we not have homogeneous boundary conditions, but instead some time-dependent behavior is fixed at the edges $x = 0$ and $x = L$. Thus, we have a problem as given below:

$$\begin{cases} u_t &= au_{xx} + F(x,t), & t > 0, 0 < x < L, \\ u(0,t) &= g(t), & t \geq 0, \\ u(L,t) &= h(t), & t \geq 0, \\ u(x,0) &= f(x), & 0 \leq x \leq L. \end{cases}$$

Physically, it would appear we are putting in time-dependent heat sources or sinks. This is indeed the case. Let

$$u(x,t) = v(x,t) + \frac{L-x}{L} g(t) + \frac{x}{L} h(t).$$

Then

$$\begin{aligned} u(0,t) = v(0,t) + g(t) \quad &\Rightarrow \quad v(0,t) = 0 \\ u(L,t) = v(L,t) + h(t) \quad &\Rightarrow \quad v(L,t) = 0. \end{aligned}$$

Thus the new function $v(x,t)$ satisfies homogeneous boundary conditions. Furthermore,

$$\begin{aligned} u_t &= v_t + \frac{L-x}{L}g'(t) + \frac{x}{L}h'(t), \\ u_{xx} &= v_{xx}, \end{aligned}$$

since the added terms are linear in x. Thus

$$\begin{cases} v_t &= av_{xx} + F(x,t) + \frac{x-L}{L}g'(t) - \frac{x}{L}h'(t), \quad t>0, 0<x<L, \\ v(0,t) &= 0, \qquad\qquad\qquad t \geq 0, \\ v(L,t) &= 0, \qquad\qquad\qquad t \geq 0, \\ v(x,0) &= f(x) - \frac{L-x}{L}g(0) - \frac{x}{L}h(0), \qquad 0 \leq x \leq L. \end{cases}$$

This is a similar problem to the one we solved in the previous section, but with a slightly different forcing term. We see that the effect of the time-dependent boundary conditions is indeed that of a forcing. In good ol' mathematical tradition, we've reduced the problem to one we've solved before. Therefore we're done.

6.4 The maximum principle in one dimension

So far, we've focused on providing exact solutions of partial differential equations. You should realize that in many situations, this is not possible, even for linear partial differential equations. Maybe the boundary conditions are too hard, maybe the domain is too complicated. Then for nonlinear equations all bets are off. We will most likely have to resort to numerical methods, or to other approximation methods such as asymptotics and perturbation theory. In situations like that it is crucial that we are able to determine some **a priori** bounds on the solution of the problem, which can act as a check on the approximation method. The maximum principle is a method that allows us to determine such bounds.

Let's get things going with a lemma. Why not indeed?

Lemma 3 *Let $a>0$, $T>0$ and $L>0$. Suppose that w satisfies $w_t - aw_{xx} < 0$ in $0<x<L$ and $0<t\leq T$. Then $w(x,t)$ cannot assume a local maximum at any point in $0<x<L$, $0<t\leq T$.*

Proof. We'll proceed by contradiction. Suppose there does exist a local maximum in our domain $\hat{x} \in (0,L)$ and $\hat{t} \in (0,T]$. At a local maximum

$$w_{xx}(\hat{x},\hat{t}) \leq 0.$$

Furthermore,

$$w_t(\hat{x},\hat{t}) \geq 0.$$

Note that usually we'd have $w_t(\hat{x},\hat{t}) = 0$ at a maximum, but in this case we have to allow for the possibility of $w_t(\hat{x},\hat{t}) > 0$, if $\hat{t} = T$. Thus we have

$$w_t(\hat{x},\hat{t}) - aw_{xx}(\hat{x},\hat{t}) \geq 0,$$

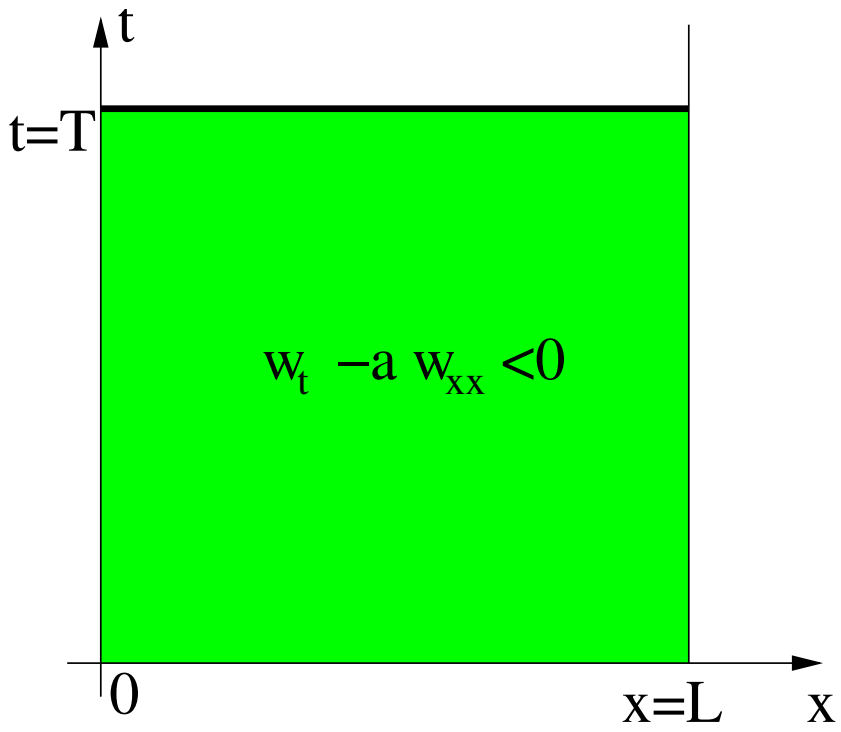

Figure 6.1: The domain under discussion in the lemma. The domain includes the top thick black line. (2.1)

which is a contradiction. This shows that such a local maximum cannot exist. ■.

The situation discussed in this lemma is illustrated in Fig. 6.1.

After every lemma, there's a theorem:

Theorem 15 *Suppose that $u(x,t)$ is continuous on $0 \le x \le L$, $0 \le t \le T$, and that*

$$\begin{aligned} u_t - a u_{xx} &\le 0 && \text{for } 0 < x < L, 0 < t \le T \\ u(0,t) &\le 0 && \text{on } 0 \le t \le T \\ u(L,t) &\le 0 && \text{on } 0 \le t \le T \\ u(x,0) &\le 0 && \text{on } 0 \le x \le L. \end{aligned}$$

Then $u(x,t) \le 0$ for $0 \le x \le L$, $0 \le t \le T$.

Proof. Again, we proceed by contradiction. It won't be the last time either. Suppose that

$$u(\hat{x}, \hat{t}) > 0,$$

for some $\hat{x} \in [0, L]$ and $\hat{t} \in [0, T]$. Consider

$$w(x,t) = u(x,t) - \epsilon t.$$

Here ϵ is a positive constant, but otherwise arbitrary. Then

$$\begin{aligned} w_t - aw_{xx} = u_t - au_{xx} - \epsilon < 0 \qquad & \text{for } \ 0 < x < L, 0 < t \leq T, \\ w(0,t) = u(0,t) - \epsilon t \leq 0 \qquad & \text{on } \ 0 \leq t \leq T \\ w(L,t) = u(L,t) - \epsilon t \leq 0 \qquad & \text{on } \ 0 \leq t \leq T \\ w(x,0) = u(x,0) \leq 0 \qquad & \text{on } \ 0 \leq x \leq L. \end{aligned}$$

Introducing the small positive parameter ϵ allows us to write the first expression with a strict inequality. That's what we need in order to invoke the previous lemma. Crafty, no? It follows from the lemma that $w(x,t)$ does not have a local maximum on the inside of our domain. This means that its largest value is obtained on the boundary. But all these boundary values are either zero or negative. However, for sufficiently small ϵ we have

$$w(\hat{x},\hat{t}) = u(\hat{x},\hat{t}) - \epsilon\hat{t} > 0,$$

since $u(\hat{x},\hat{t}) > 0$. This provides a contradiction, finishing the proof. ■

The following theorem is entirely similar to the previous one.

Theorem 16 *Suppose that $u(x,t)$ is continuous on $0 \leq x \leq L$, $0 \leq t \leq T$, and that*

$$\begin{aligned} u_t - au_{xx} \geq 0 \qquad & \textit{for } \ 0 < x < L, 0 < t \leq T \\ u(0,t) \geq 0 \qquad & \textit{on } \ 0 \leq t \leq T \\ u(L,t) \geq 0 \qquad & \textit{on } \ 0 \leq t \leq T \\ u(x,0) \geq 0 \qquad & \textit{on } \ 0 \leq x \leq L. \end{aligned}$$

Then $u(x,t) \geq 0$ for $0 \leq x \leq L$, $0 \leq t \leq T$.

Proof. The function $-u$ satisfies the conditions of the preceding theorem. ■

These two last theorems essentially tell us that heat equation problems have some kind of monotonicity to them: if everything on the boundary is negative, then so is the solution. Similarly, if everything on the boundary is positive, then so is the solution. So, solutions of the heat equation, while decaying to zero, don't overshoot their target. The following theorem finalizes these statements.

Theorem 17 (Minimum-Maximum principle) *Suppose that $u(x,t)$ is continuous on $0 \leq x \leq L$, $0 \leq t \leq T$, and that $u_t = au_{xx}$ for $0 < x < L$, $0 < t \leq T$. Define*

$$M = \max\{\max_{x\in[0,L]} u(x,0), \max_{t\in[0,T]} u(0,t), \max_{t\in[0,T]} u(L,t)\},$$

and

$$m = \min\{\min_{x\in[0,L]} u(x,0), \min_{t\in[0,T]} u(0,t), \min_{t\in[0,T]} u(L,t)\}.$$

Then

$$m \leq u(x,t) \leq M \quad \textit{for all} \quad (x,t) \in [0,L] \times [0,T].$$

Proof. The proof is straightforward: apply the previous two theorems to $M - u(x,t)$ and $u(x,t) - m$ respectively. ■

The Minimum-Maximum Principle states that a solution of the heat equation on a bounded domain is always in between the largest and smallest of its values on the boundary of its domain of definition. We'll see in the next setting that this conclusion holds in an arbitrary number of dimensions. Now we'll show that in addition to providing stringent a-priori bounds, the Minimum-Maximum Principle may also be used to prove some more fundamental results.

Corollary 2 *Let $f(x)$, $g(t)$ and $h(t)$ be continuous on their respective domains of definition. Then the solution of*

$$\begin{aligned} u_t = au_{xx} + F(x,t) \qquad & \textit{for } \ 0 < x < L, t > 0, \\ u(0,t) = g(t) \qquad & \textit{for } \ t \geq 0, \\ u(L,t) = h(t) \qquad & \textit{for } \ t \geq 0, \\ u(x,0) = f(x) \qquad & \textit{for } \ 0 \leq x \leq L \end{aligned}$$

is unique.

Proof. As before, we consider $w = u_1 - u_2$, where u_1 and u_2 both solve the initial-boundary-value problem posed above. Then w satisfies

$$\begin{aligned} w_t = aw_{xx} \qquad & \text{for } \ 0 < x < L, t > 0, \\ w(0,t) = 0 \qquad & \text{for } \ t \geq 0, \\ w(L,t) = 0 \qquad & \text{for } \ t \geq 0, \\ w(x,0) = 0 \qquad & \text{for } \ 0 \leq x \leq L. \end{aligned}$$

Using the definitions of the Minimum-Maximum Principle, we have that

$$M = 0, \quad m = 0,$$

from which

$$0 \leq w(x,t) \leq 0 \quad \Rightarrow \quad w(x,t) \equiv 0.$$

This results in $u_1 = u_2$, proving uniqueness. ■

Similarly, the Minimum-Maximum Principle can be used to show that solutions of this heat equation problem are not sensitively dependent on the input functions $f(x)$, $g(t)$, $h(t)$ and $F(x,t)$. However, this is not surprising: all solutions are exponentially damped, apart from the effect of the forcing term.

We'll now look at an example of how the Minimum-Maximum Principle is used to construct a-priori bounds on a solution, far improving on those mentioned in the theorem itself.

Example: Consider the initial-boundary-value problem

$$\begin{aligned} u_t &= u_{xx} && \text{for } 0 < x < 1,\ t > 0, \\ u(0,t) &= 0 && \text{for } t \geq 0, \\ u(1,t) &= 0 && \text{for } t \geq 0, \\ u(x,0) &= x(1-x)e^x && \text{for } x \in [0,1]. \end{aligned}$$

It follows immediately from the Minimum-Maximum Principle that $u \geq 0$. We could also compute the maximum of $u(x,0)$, and conclude that u is less than or equal to this maximum. But we can do better. The name of the game is comparison: let's find a good function to compare $u(x,t)$ with. Then we'll apply the Minimum-Maximum Principle to their difference.

The initial condition suggests we should compare $u(x,t)$ with

$$v(x,t) = \beta x(1-x)e^{-\alpha t}.$$

We've ignored the effect of the e^x factor in the initial condition. We can absorb its multiplicative effect in the coefficient β, to be determined. Similarly, α is an exponent to be determined. Guessing a decaying exponential is always a good idea for the solution to a heat equation. Our goal is to choose α and β so that

$$u(x,t) \leq v(x,t).$$

Let $w = u - v$, then w satisfies

$$\begin{aligned} w_t - w_{xx} &= u_t - v_t - u_{xx} + v_{xx} \\ &= -v_t + v_{xx} \\ &= \beta e^{-\alpha t}(\alpha x(1-x) - 2). \end{aligned}$$

We want this to be negative, thus we need to choose α so that

$$\begin{aligned} & && \alpha x(1-x) \leq 2 \\ \Rightarrow & && \alpha \leq \frac{2}{x(1-x)} \\ \Rightarrow & && \alpha \leq 8. \end{aligned}$$

We also have

$$w(0,t) = 0 = w(1,t),$$

for $t \geq 0$; and

$$\begin{aligned} w(x,0) &= x(1-x)e^x - \beta x(1-x) \\ &= x(1-x)\left(e^x - \beta\right). \end{aligned}$$

In order for this to be negative, we need

$$\begin{aligned} & \beta \geq e^x \\ \Rightarrow \qquad & \beta \geq e. \end{aligned}$$

Making these choices for α and β, we can apply the Minimum-Maximum Principle to the function $w(x,t)$, to find that

$$w(x,t) \leq 0 \quad \Rightarrow \quad u(x,t) \leq v(x,t),$$

so that we now know that

$$0 \leq u(x,t) \leq x(1-x)e^{1-8t}.$$

Since this upper and lower bound approach each other as $t \to \infty$, these inequalities provide quite a good test to any approximation method!

The real power of the Minimum-Maximum Principle becomes apparent in higher-dimensional settings, on irregular domains, as we'll see next.

6.5 The maximum principle in a multi-dimensional setting

Consider the heat equation in any number of dimensions d ($= 1, 2$ or 3, for applications). We have

$$\begin{aligned} d = 1 : \qquad & u_t = a u_{xx} \\ d = 2 : \qquad & u_t = a(u_{xx} + u_{yy}) \\ d = 3 : \qquad & u_t = a(u_{xx} + u_{yy} + u_{zz}). \end{aligned}$$

At this point we note that there was almost nothing in our proofs of the one-dimensional form of the Minimum-Maximum Principle that restricted it to one dimension. In this section, we'll extend it to the multidimensional setting.

We'll repeat some of those results here again, for completeness. This will also give us a preview on how to deal with the higher-dimensional setting, which will be useful later.

Let's introduce some notation:

$$S_T = \bar{D} \cup \{t \in [0,T]\}.$$

Here the notation $\bar{D}$ denotes a closed domain. This a domain that includes its own boundary: if D is an open domain then

$$\bar{D} = D \cup \partial D,$$

where ∂D denotes the boundary of the domain D. For our purposes, D is a domain in which the d-dimensional spatial variable lives. In what follows, $\mathring{S}_T$ is an open domain, not including its spatial boundary, but including its temporal boundary. As before, we have a lemma.

Lemma 4 *Let the function u be twice differentiable in the spatial variables, and once in the time variable. Let $a > 0$. If*

$$u_t - a\Delta u < 0$$

on $\mathring{S}_T$, then u has no local maximum in $\mathring{S}_T$.

Proof. Assume u has a local maximum, occurring at

$$(\hat{x}, \hat{t}) \in \mathring{S}_T.$$

Here $\hat{x}$ is a spatial coordinate of dimension d. Then, at this point,

$$\Delta u(\hat{x}, \hat{t}) \leq 0.$$

Usually we have a strict inequality here, but equality could occur in degenerate cases (e.g., the function $-x^4$). Also

$$u_t(\hat{x}, \hat{t}) \geq 0.$$

A similar comment is in order: usually we have equality here, but inequality could occur if the maximum is attained at $t = T$. Thus

$$u_t - a\Delta u \geq 0,$$

at $(\hat{x}, \hat{t})$, which provides a contradiction. This ends the proof. ■.

From this lemma, we easily obtain the following theorem.

Theorem 18 *Let the function u be twice differentiable in the spatial variables, and once in the time variable. Let $a > 0$. If*

$$u_t - a\Delta u \leq 0 \quad \textit{on} \quad \mathring{S}_T$$

and

$$u \leq 0 \quad \textit{on} \quad \partial S_T \textit{ (Spatial boundary only)},$$

then $u \leq 0$ on S_T.

Proof. Again, the proof is similar to the one-dimensional setting. Let $\epsilon > 0$ be an otherwise arbitrary constant. Consider

$$w = u - \epsilon t,$$

from which

$$w_t - a\Delta w = u_t - \epsilon - a\Delta u < 0,$$

for any ϵ. since we just bought ourselves a strict inequality, we may apply the previous lemma. Thus w assumes its maximum on the ∂S_T (spatial boundary). But $w \leq 0$ on the boundary, hence

$$w \leq 0, \quad \text{on} \quad S_T.$$

Since $t < T$ is finite, we can let the term ϵt become infinitely small, resulting in

$$u \leq 0, \quad \text{on} \quad S_T.$$

This is what we had to prove. ■

A similar theorem holds for positive values on the boundary, and a differential inequality of different sign. This is proven by letting $u \to -u$ in the preceding lemma and theorem.

We now get the Minimum-Maximum Principle for the heat equation in any number of dimensions. This result, much more than its one-dimensional counterpart, provides strong a-priori bounds on the solution u of the heat equation, in complicated domains where the problem cannot be solved analytically.

Theorem 19 *Suppose that u is continuous on $\mathring{S}_T$, twice differentiable in the spatial variables, and once in time. Lastly, let u be continuous on ∂S_T. If*

$$u_t = a\Delta u, \quad \text{on} \quad \mathring{S}_T,$$

then

$$\min_{\partial S_T} u \leq u \leq \max_{\partial S_T} u$$

Proof. $m - u$ satisfies the condition of our previous theorem, and so does $u - M$, from which the desired result follows. ■

As in the one-dimensional case, the Minimum-Maximum Principle allows us to prove uniqueness of the solution of the heat equation, even with forcing or nonhomogeneous boundary conditions. Try this as a good exercise.

6.6 The Green's function

Let's recall the notion of a fundamental matrix for a system of ordinary differential equations. For simplicity, we'll restrict ourselves to the case of constant coefficients. Thus we consider the system

$$x' = Ax,$$

where $x \in \mathbb{R}^d$, and $A \in \mathbb{R}^{d \times d}$, a constant matrix. We can write the solution of this system as

$$x = \phi(t)x_0,$$

where $\phi(t)$ is the fundamental matrix of the system, and x_0 is the initial condition for x:

$$x(0) = x_0.$$

For a constant coefficient system we can be more explicit: the fundamental matrix is given by

$$\phi(t) = e^{At},$$

defined using the standard matrix exponential (for instance, using a Taylor expansion). One of the good things about a fundamental matrix is that it allows us to solve the system many times for different initial conditions, merely by multiplying by a different vector. In other words, the fundamental matrix encodes everything about the system, except the initial condition.

We seek to generalize this concept to partial differential equations. Of course, for partial differential equations there are even more types of input: there's the equation itself, there are the initial conditions, and there are the boundary conditions. Depending on the context, a **Green's function** will encode all information contained in the partial differential equation (or ordinary differential equation, as we'll see later), and one of the initial data or boundary conditions. The idea is to get an operation that can easily be repeated many times, as one piece of the input data is changed. The context will dictate which input function we wish to be varied. We'll see that Green's functions have many nice properties, many of which are similar to those of the fundamental matrix.

Remark. This concept of a Green's function can be generalized even further to nonlinear partial differential equations, using the idea of a semi-group. We won't get further into this.

Let's make the concept of a Green's function a bit more concrete, before we construct a few for specific examples. Consider the linear partial differential equation

$$u_t = L(u),$$

where L is a linear differential operator in x acting on u. Suppose that we're given initial conditions

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

defined for $x \in \mathbb{R}$. We're looking for solutions such that $u \to 0$ as $x \to \pm\infty$. In this context, we're looking for an operator such that

$$u(x,t) = \theta(x,t)f(x).$$

We will see that in many examples this operator $\theta(x,t)$ is an integral operator. It's kernel will be the Green's function.

1. The heat equation on a finite domain, revisited

Let's consider the problem

$$\begin{cases} u_t &= au_{xx}, & t > 0, 0 < x < L, \\ u(0,t) &= 0, & t \geq 0, \\ u(L,t) &= 0, & t \geq 0, \\ u(x,0) &= f(x), & 0 \leq x \leq L. \end{cases}$$

We've considered this problem a few sections ago. It's solution is

$$u(x,t) = \sum_{n=1}^{\infty} \beta_n e^{-a\lambda_n^2 t} \sin \lambda_n x,$$

where $\lambda_n = n\pi/L$. In order to solve the initial-value problem, we need

$$\beta_n = \frac{2}{L}\int_0^L f(x)\sin\lambda_n x dx,$$

for $n = 1, 2, \ldots$.

Substituting this in the general solution formula, we find

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty}\left(\frac{2}{L}\int_0^L f(y)\sin\lambda_n y dy\right)e^{-a\lambda_n^2 t}\sin\lambda_n x \\ &= \int_0^L\left(\frac{2}{L}\sum_{n=1}^{\infty}e^{-a\lambda_n^2 t}\sin\lambda_n x\sin\lambda_n y\right)f(y)dy. \end{aligned}$$

We call

$$G(x,y,t) = \frac{2}{L}\sum_{n=1}^{\infty}e^{-a\lambda_n^2 t}\sin\lambda_n x\sin\lambda_n y$$

the Green's function for the heat equation on a finite domain with homogeneous boundary conditions. Then

$$u(x,t) = \int_0^L G(x,y,t)f(y)dy,$$

as desired.

Some of the steps taken above were formal (switching integrals and infinite sums, for instance), so some conditions on $f(x)$ are required for all steps to make sense. At this point, we can take the definition of the Green's function and the formula as a starting point for solving the given problem. It turns out that by doing this, we can dispense with many of our previous conditions. The reason for this is that integrals are smoothing operators.

Properties of the Green's function

1. The Green's function satisfies the partial differential equation, in both spatial variables for $t > 0$ and $0 < x < L$: thus

   $$G_t = aG_{xx}, \quad G_t = aG_{yy}.$$

   This follows immediately from the uniform convergence of the series representation of the Green's function and its derivatives, and from the functional form of the summands.

2. The Green's function satisfies the boundary conditions:

   $$G(0,y,t) = 0, \quad G(L,y,t) = 0,$$

   for $0 \le y \le L$, and for $t > 0$. This is immediately verified from the form of the Green's function.

3. The Green's function is symmetric in its spatial arguments:

$$G(x, y, t) = G(y, x, t),$$

for $0 \leq x, y \leq L$, and $t > 0$. Again, this follows immediately from the functional form of the Green's function.

4. If $f(x)$ is continuous and $f(0) = 0$, $f(L) = 0$ (*i.e.,* the initial condition is compatible with the boundary conditions), then

$$\lim_{t>0} \int_0^L G(x, y, t) f(y) dy = f(x).$$

(Thus the Green's function behaves like a delta function as $t \to 0$). This one requires a bit more work. Let's prove it.

Proof. Choose a sequence of twice differentiable functions $f_n(x)$ (n=1,2,...), uniformly convergent to $f(x)$, with $f_n(0) = 0 = f_n(L)$, for all n. Then, for all $\epsilon > 0$, there exists an $N(\epsilon)$ such that $n, m > N(\epsilon)$ implies that

$$|f_n(x) - f_m(x)| < \epsilon,$$

for $x \in [0, L]$. Let $u_n(x, t)$ be the solution of the heat equation with homogeneous boundary conditions corresponding to initial condition $f_n(x)$. We know from our theorem using the separation of variables solution that all the functions u_n are continuous.

By the Minimum-Maximum Principle we have

$$\min_{x\in[0,L]} (f_n(x) - f_m(x)) \leq u_n(x, t) - u_m(x, t) \leq \max_{x\in[0,L]} (f_n(x) - f_m(x)),$$

or

$$|u_n - u_m| \leq \epsilon.$$

Hence the sequence $\{u_n, n = 1, 2, \ldots\}$ is a uniform Cauchy sequence in the space of continuous functions. Since the space of continuous functions is complete, as long as consider uniform convergence (the uniform limit of a sequence of continuous functions is continuous), this Cauchy sequence is a convergent sequence, say $u_n \to u$, where the convergence is uniform, and u is continuous. Then

$$\begin{aligned}
u(x, t) &= \lim_{n\to\infty} u_n(x, t) \\
&= \lim_{n\to\infty} \int_0^L G(x, y, t) f_n(y) dy \\
&= \int_0^L \lim_{n\to\infty} G(x, y, t) f_n(y) dy \quad \text{(by uniform convergence)} \\
&= \int_0^L G(x, y, t) \lim_{n\to\infty} f_n(y) dy \\
&= \int_0^L G(x, y, t) f(y) dy.
\end{aligned}$$

On the other hand,

$$\begin{aligned} u(x,0) &= \lim_{t>0} u(x,t) \\ &= \lim_{t>0} \lim_{n\to\infty} u_n(x,t) \\ &= \lim_{n\to\infty} \lim_{t>0} u_n(x,t) \\ &= \lim_{n\to\infty} f_n(x) \\ &= f(x). \end{aligned}$$

We are allowed to switch the limits since the convergence in n is uniform. Taking the limit as $t \to 0^+$ of the previous result proves our delta function property. ■

Putting all these properties together, we have shown that the following theorem holds.

Theorem 20 *Let $f(x)$ be continuous on $x \in [0, L]$ with $f(0) = 0 = f(L)$. Then the problem*

$$\begin{cases} u_t &= au_{xx}, & t > 0, 0 < x < L, \\ u(0,t) &= 0, & t \geq 0, \\ u(L,t) &= 0, & t \geq 0, \\ u(x,0) &= f(x), & 0 \leq x \leq L. \end{cases}$$

has a unique, continuous solution for $t > 0$, given by

$$u(x,t) = \int_0^L G(x,y,t) f(y) dy.$$

This is quite a surprising result: we have found that using the Green's function formalism, we can obtain a solution of the heat equation requiring less regularity than we could before, using separation of variables. We'll take it!

2. The heat equation on an infinite domain

Consider the heat equation with initial data defined on the entire real line:

$$\begin{aligned} u_t &= au_{xx} && x \in \mathbb{R}, \\ u(x,0) &= f(x) && x \in \mathbb{R}, \end{aligned}$$

where $f(x)$ is sufficiently fast decaying, to allow for the formal analysis that follows.

As in our discussion of evolution equations, we look for solutions of the form

$$u = Ae^{ikx - i\omega t},$$

so that

$$\begin{aligned} u_x &= iku, \\ u_{xx} &= (ik)^2 u, \\ u_t &= -i\omega u. \end{aligned}$$

Thus $-i\omega = a(ik)^2 \;\Rightarrow\; \omega = -aik^2$, which is the dispersion relationship. Since its imaginary part is negative, we see that the solutions of the heat equation will decay as $t \to \infty$. We get

$$u_k(x,t) = A(k)e^{ikx-ak^2t}.$$

Superimposing these solutions gives

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx-ak^2t}dk.$$

Next, we impose the initial conditions to find

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)e^{ikx}dk \\ \Rightarrow \qquad A(k) &= \int_{-\infty}^{\infty} e^{-iky}f(y)dy, \end{aligned}$$

where we've introduced the dummy variable y. This allows us to rewrite the solution as

$$\begin{aligned} u(x,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx-ak^2t}\left(\int_{-\infty}^{\infty} e^{-iky}f(y)dy\right)dk \\ &= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx-ak^2t-iky}dk\right)f(y)dy \\ &= \int_{-\infty}^{\infty} G(x,y,t)f(y)dy. \end{aligned}$$

We have defined the Green's function as

$$\begin{aligned} G(x,y,t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx-ak^2t-iky}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ik(x-y)-ak^2t}dk \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos k(x-y)e^{-ak^2t}dk. \end{aligned}$$

The imaginary part of the integrand disappears, because it leads to the integral of an odd function. We can do some nifty manipulation on this integral in order to get an explicit (no

integral) expression for the Green's function. Let $x - y = z$, then

$$\begin{aligned}
\frac{dG}{dz} &= -\frac{1}{2\pi}\int_{-\infty}^{\infty} k \sin kz e^{-ak^2t} dk \\
&= \frac{1}{2\pi}\left(\frac{1}{2at} e^{-ak^2t} \sin kz \Big|_{-\infty}^{\infty} - \frac{z}{2at}\int_{-\infty}^{\infty} \cos kz e^{-ak^2t} dk \right) \\
&= -\frac{z}{2at} G(z) \\
\Rightarrow \qquad \frac{d}{dz}\ln G &= -\frac{z}{2at} \\
\Rightarrow \qquad \ln \frac{G}{G_0} &= -\frac{z^2}{4at} \\
\Rightarrow \qquad G &= G_0 e^{-\frac{(x-y)^2}{4at}}.
\end{aligned}$$

We can find G_0 directly:

$$\begin{aligned}
G_0 &= G(x,x,t) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ak^2t} dk \\
\Rightarrow \qquad G_0^2 &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} e^{-ak^2t} dk \int_{-\infty}^{\infty} e^{-al^2t} dl \\
&= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-at(k^2+l^2)} dk dl \\
&= \frac{1}{4\pi^2} 2\pi \int_0^{\infty} e^{-atr^2} r dr \\
&= \frac{1}{2\pi}\left. \frac{e^{-atr^2}}{-2at} \right|_{r=0}^{r=\infty} \\
&= \frac{1}{2\pi}\frac{1}{2at} \\
\Rightarrow \qquad G_0 &= \frac{1}{\sqrt{4\pi at}},
\end{aligned}$$

so that

$$G(x,y,t) = \frac{1}{\sqrt{4\pi at}} e^{-\frac{(x-y)^2}{4at}}.$$

This allows us to write the solution of the heat equation on an infinite domain as

$$u(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi at}} e^{-\frac{(x-y)^2}{4at}} f(y) dy.$$

Note that $G(x,y,t)$ has an essential singularity as $t \to 0$ ($t > 0$), in which limit $G(x,y,t) \to \delta(x-y)$. As before, $G(x,y,t)$ satisfies the heat equations

$$G_t = aG_{xx} = aG_{yy}.$$

With a little bit more effort, we can establish the following theorem.

Theorem 21 *Let $f(x)$ be bounded and piecewise continuous on $-\infty < x < \infty$. Then the heat equation problem on the whole line is solved by*

$$u(x,t) = \begin{cases} \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} f(y)dy & \text{for } t > 0, \\ f(x) & \text{for } t = 0. \end{cases}.$$

This solution is continuous for all $t > 0$. It is continuous at $t = 0$, except at the discontinuities of $f(x)$.

The proof is technical, but entirely similar to some of the previous ones. A surprising side note is that the above theorem does not require $f(x)$ to be integrable on the whole line! This is a testament to the smoothing and damping effect of the solution operator of the heat equation. Let's do an example.

Example: Solve the initial-value problem

$$\begin{cases} u_t &= a u_{xx} \\ u(x,0) &= f(x) = 1. \end{cases}$$

for $x \in \mathbb{R}$. Note that in this example, $f(x) = 1$ is not integrable. It is however piecewise continuous and it therefore satisfies the conditions of the theorem. Let's see what we get. For $t > 0$,

$$u(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} dy.$$

Let

$$\xi = \frac{y-x}{\sqrt{4at}} \quad \Rightarrow \quad d\xi = \frac{1}{\sqrt{4at}} dy,$$

and

$$u(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-\xi^2} d\xi = 1.$$

Direct substitution confirms that this is the correct solution. In this case there is no decay of energy: we've started with an infinite amount, and we keep an infinite amount throughout. As a consequence, there's no decay in the solution.

3. The heat equation on a semi-infinite domain

We can use the result of the previous section to solve the heat equation on a semi-infinite domain, such as the one shown in Fig. 6.2.

Thus we are interested in solving the problem

$$\begin{aligned} u_t = a u_{xx} \qquad & \text{for } x > 0, t > 0, \\ u(x,0) = f(x) \qquad & \text{for } x > 0, \\ u(0,t) = 0 \qquad & \text{for } t > 0. \end{aligned}$$

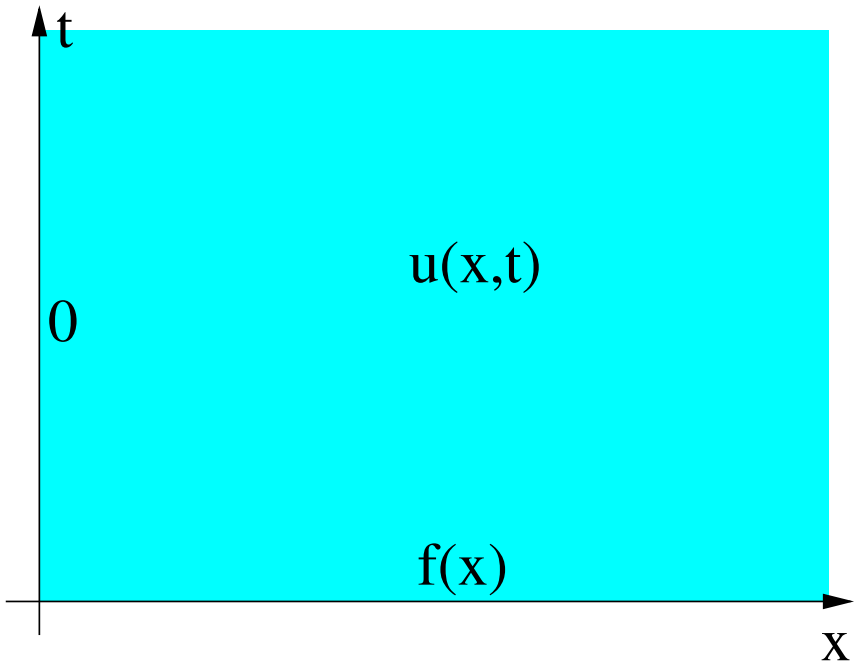

Figure 6.2: The domain under discussion for the heat equation with initial data on a half line, with homogeneous boundary conditions. (2.1)

We'd like to solve this problem using as much as possible from the previous section[1]. In order to do this, we need to extend the initial condition to the whole line in such a way that the $u(0,t) = 0$ boundary condition is automatically satisfied. It seems reasonable we'll accomplish this by using an odd extension of $f(x)$. In other words, we'll consider the problem

$$\begin{aligned} u_t &= a u_{xx} && \text{for } x \in \mathbb{R}, t > 0, \\ u(x,0) &= g(x) && \text{for } x \in \mathbb{R}, \end{aligned}$$

where $g(x)$ is the odd extension of $f(x)$:

$$g(x) = \begin{cases} f(x) & \text{for } x > 0, \\ 0 & \text{for } x = 0, \\ -f(-x) & \text{for } x < 0. \end{cases}$$

Using our previous result, we get

$$\begin{aligned} u(x,t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} g(y)\,dy \\ &= -\int_{-\infty}^{0} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} f(-y)\,dy + \int_{0}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} f(y)\,dy \end{aligned}$$

[1] call it pragmatic. It sounds better than lazy

$$= \int_{\infty}^{0} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x+y)^2}{4at}} f(y)dy + \int_{0}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} f(y)dy$$
$$= \int_{0}^{\infty} \frac{1}{\sqrt{4\pi a t}} \left(e^{-\frac{(x-y)^2}{4at}} - e^{-\frac{(x+y)^2}{4at}} \right) f(y)dy.$$

If this result is correct (wait for it... We'll check it in a minute) then the Green's function for the heat equation problem on the half-line is

$$G(x, y, t) = \frac{1}{\sqrt{4\pi a t}} \left(e^{-\frac{(x-y)^2}{4at}} - e^{-\frac{(x+y)^2}{4at}} \right).$$

Let's check that the solution we've constructed is indeed a solution of the original initial-value problem with homogeneous boundary conditions.

- **The solution satisfies the partial differential equation.** We could calculate some fancy derivatives and do some tedious calculus. Let's not. Instead, we can notice that the solution constructed is a superposition of two solutions to the heat equation. Since the heat equation is linear, this superposition still satisfies the equation.

- **The solution satisfies the initial condition.** Let's use the alternate form

  $$u(x, t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi a t}} e^{-\frac{(x-y)^2}{4at}} g(y)dy.$$

  Evaluating this at $t = 0$ gives that

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

  which is equal to $f(x)$ on the positive half line, as desired.

- **The solution satisfies the boundary condition.** Indeed:

  $$u(0, t) = \int_{0}^{\infty} \frac{1}{\sqrt{4\pi a t}} \left(e^{-\frac{y^2}{4at}} - e^{-\frac{y^2}{4at}} \right) f(y)dy = 0.$$

We have verified directly that we have constructed a solution of the heat equation on the positive half line.

6.7 Exercises

1. Consider the linear scalar evolution equation

   $$u_t = L(u),$$

   where L is a linear differential operator in x of order N, acting on u. Assume that this equation is defined on the entire real line, $x \in \mathbb{R}$, with initial condition

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

Here $f(x)$ is a "nice" function so that all formal calculations can be justified. Using formal calculations, and the dispersion relation, find the Green's function for this problem. (Note: your final answer may have an integral in it.)

2. Find a formal solution to the following initial-value problem. Here $x \in [0, L]$.

$$\begin{cases} u_t = \sigma u_{xx} + F(x,t), \\ u(x,0) = 0, \\ u(0,t) = 0, \\ u_x(L,t) = 0. \end{cases}$$

Chapter 7

Integral eqations, Sturm-Liouville problems, and eigenfunction expansions

So far, we've looked at expansions of functions in terms of a simpler set of basis functions in the simplest case: rectangular domains, and problems with constant coefficients.

In general, more complicated scenarios arise if either of these conditions is relaxed. In many cases, separation of variables is still successful, but the ordinary differential equations that need to be solved are more complex. One often has to solve a Sturm-Liouville problem of the form

$$-\left(p(x)y'\right)' + q(x)y = r(x)\lambda y,$$

with boundary conditions

$$\begin{cases} a_1 y(0) + a_2 y'(0) &= 0, \\ b_1 y(1) + b_2 y'(1) &= 0. \end{cases}$$

Note that we can always scale x so that its domain is $x \in [0,1]$. Let's define the differential operator L so that

$$Ly = -\left(k(x)y'\right)' + l(x)y.$$

This allows us to rewrite the Sturm-Liouville problem as

$$Ly = \lambda r(x)y.$$

7.1 Green's functions for Sturm-Liouville problems

We'll review many properties of these systems as we proceed in this chapter. We'll even prove them. We start by discussing the nonhomogeneous problem

$$\begin{cases} -\left(p(x)y'\right)' + q(x)y = f(x), & \text{for } 0 < x < 1, \\ y(0) = h_0 y'(0),\ y(1) = -h_1 y'(1), \end{cases}$$

where $p(x) \neq 0$, $p'(x)$, $q(x)$ and $f(x)$ are continuous on $x \in [0,1]$.

1. Definition and derivation

It is natural to look for a solution of this problem using a Green's function: since we regard $f(x)$ as our input function, we'd like to write

$$y(x) = \int_0^1 g(x,s)f(s)ds,$$

where $g(x,s)$ is the Green's function we're looking for. You might want to think of this formula as kind of a variation of parameters idea. In that case the solution is also given as an integral over the forcing function.

A function $g(x,s)$ is called a Green's function for our Sturm-Liouville problem if it is continuous on $x \in [0,1]$ and

$$y(x) = \int_0^1 g(x,s)f(s)ds$$

is the unique solution of the Sturm-Liouville boundary-value problem.

To show the validity of our solution formula, we first need Lagrange's identity.

Lemma 5 (Lagrange's identity) *If u, v are twice differentiable, then*

$$vLu - uLv = \frac{d}{dx}\left(p(x)W(u,v)\right),$$

where $W(u,v) = uv' - u'v$ is the Wronskian of the functions u and v.

Proof. With $Ly = -(py')' + qy$, we have

$$\begin{aligned} vLu - uLv &= v\left(-(pu')' + qu\right) - u\left(-(pv')' + qv\right) \\ &= -v(pu')' + u(pv')' \\ &= -vp'u' - vpu'' + up'v' + upv'' \\ &= p'(uv' - u'v) + p(uv'' - vu'') \\ &= p'W(u,v) + pW'(u,v) \\ &= \frac{d}{dx}\left(p(x)W(u,v)\right), \end{aligned}$$

which finishes the proof. ■

In what follows we will assume Dirichlet boundary conditions:

$$y(0) = 0, \quad y(1) = 0,$$

to simplify the calculations. It would be a good exercise to repeat what follows for the more general original boundary conditions.

Consider the two initial-value problems

$$\begin{cases} Lu &= 0, \\ u(0) &= 0, \end{cases}$$

and

$$\begin{cases} Lv &= 0, \\ v(1) &= 0. \end{cases}$$

Let y denote the solution to the nonhomogeneous problem we wish to solve:

$$Ly = f(x), \quad y(0) = 0, \ y(1) = 0.$$

We'll use Lagrange's identity twice.

1. First, we use it with u and y:

$$\begin{aligned} yLu - uLy &= \frac{d}{dx} pW(u, y) \\ \Rightarrow \qquad -uf &= \frac{d}{dx} pW(u, y). \end{aligned}$$

2. Next, we use it with v and y:

$$\begin{aligned} yLv - vLy &= \frac{d}{dx} pW(v, y) \\ \Rightarrow \qquad -vf &= \frac{d}{dx} pW(v, y). \end{aligned}$$

Writing these out, we have

$$-\int_0^x u(s)f(s)ds = p(uy' - u'y)$$

and

$$\int_1^x v(s)f(s)ds = p(vy' - v'y)$$

$$\begin{aligned} \Rightarrow \qquad v(x)\int_0^x u(s)f(s)ds + u(x)\int_x^1 v(s)f(s)ds &= pvu'y - pv'uy \\ &= -pW(u, v)y. \end{aligned}$$

We can get the Wronskian using Abel's formula:

$$\begin{aligned} LY = 0 \Leftrightarrow \qquad -pY'' - p'Y' + qY &= 0 \\ \Rightarrow \qquad Y'' + \frac{p'}{p}Y' - \frac{q}{p}Y &= 0 \\ \Rightarrow \qquad W &= Ce^{-\int \frac{p'}{p} dx} \\ &= Ce^{-\ln p} \\ &= \frac{C}{p}, \end{aligned}$$

where C is a constant. It can be chosen arbitrarily, as long as it's not chosen to be zero. We'll choose $C = -1$ here. Thus

$$y = v(x)\int_0^x u(s)f(s)ds + u(x)\int_x^1 v(s)f(s)ds$$
$$y = \int_0^1 g(x,s)f(s)ds,$$

with

$$g(x,s) = \begin{cases} v(x)u(s) & 0 \le s \le x \le 1, \\ u(x)v(s) & 0 \le x \le s \le 1. \end{cases}$$

So, if there is a Green's function, it provides a unique solution to the nonhomogeneous Sturm-Liouville problem. Next we establish when a Green's function exists.

Theorem 22 *The Sturm-Liouville problem has a Green's function if and only if the homogeneous problem (corresponding to $f(x) = 0$) only has the trivial solution $y = 0$.*

Before we prove this, we should realize that a statement like this makes sense: since the Green's function provides a solution to the problem, it inverts the Sturm-Liouville operator L in some sense. This is only possible if the homogeneous problem has no nontrivial solutions, just like in linear algebra. This is a special case of the very general **Fredholm alternative**, which we'll prove near the end of this chapter.

Proof. First we prove $\Rightarrow$. If there is a Green's function then the solution of the homogeneous problem is (put $f(x) \equiv 0$)

$$y = \int_0^1 g(x,s)0ds = 0.$$

No other solutions are possible.

Next we prove $\Leftarrow$. Assume that the problem

$$Ly = 0, \quad y(0) = 0, \ y(1) = 0$$

only has the trivial $y = 0$ solution. We can always find nontrivial solutions to the initial-value problems

$$Ly = 0 \ \text{ and } \ y(0) = 0,$$

and

$$Ly = 0 \ \text{ and } y(1) = 0.$$

Call these solutions u and v respectively. These solutions are nontrivial, since the general solutions will depend on two arbitrary constants since we're solving a second-order equation. We're only fixing one of these constants using the one initial condition given. The question is whether it is still possible to satisfy the condition

$$C = -1,$$

or

$$W(u,v) = -1/p.$$

If this is possible we've constructed our Green's function. Suppose that it is not possible, which happens if $W(u,v) = 0$ and $C = 0$. In that case u and v have to be proportional, implying that

$$u(1) = 0,\ v(0) = 0.$$

Then u (or v) is a nontrivial solution of the homogeneous problem, violating our assumption. Thus the functions u and v constructed are not proportional, and therefore not linearly dependent. Hence their Wronskian does not vanish and C can be chosen to be any nonzero constant, especially -1. Thus if the homogeneous problem only has the trivial solution, we can contruct the Green's function.

This completes the proof. ■

2. Properties of the Green's function

We group the properties of the Green's function in the following theorem.

Theorem 23 *The Green's function for the nonhomogeneous Sturm-Liouville problem satisfies the following five properties:*
(1) Symmetry: $g(x,s) = g(s,x)$,
(2) $g(x,s)$ *is continuous on* $0 \le x, s \le 1$,
(3) The second derivatives of $g(x,s)$ *are continuous on each of the triangles* $0 \le s < x \le 1$ *and* $0 \le x < s \le 1$. *Further, on each triangle* $Lg = 0$, *where* L *is a differential-in-*x *operator,*
(4) $g(0,s) = 0$, $g(1,s) = 0$, *and*
(5) For $0 < s < 1$, $g_x(s^+,s) - g_x(s^-,s) = -1/p(s)$.

Proof. The domain on which the Green's function is defined is illustrated in Fig. 7.1.

Proving (1) is done by direct inspection. Next we turn to proving (2). The only issue of continuity might be on the diagonal on the square domain of definition. Let's check. We have

$$g(x,s) = \begin{cases} v(x)u(s), & 0 \le s \le x \le 1 \text{ gold triangle,} \\ u(x)v(s), & 0 \le x \le s \le 1 \text{ purple rectangle.} \end{cases}$$

On the diagonal $x = s$ we have that

$$g(s^-,s) = v(s)u(s) = g(s^+,s),$$

which proves (2). Also,

$$g(0,s) = u(0)v(s) = 0,$$

and

$$g(1,s) = v(1)u(s) = 0,$$

which proves (4). Because g is a multiple of either $u(x)$ or $v(x)$ on either the gold or purple triangle, it is clear that the second derivatives are continuous on these triangles and that

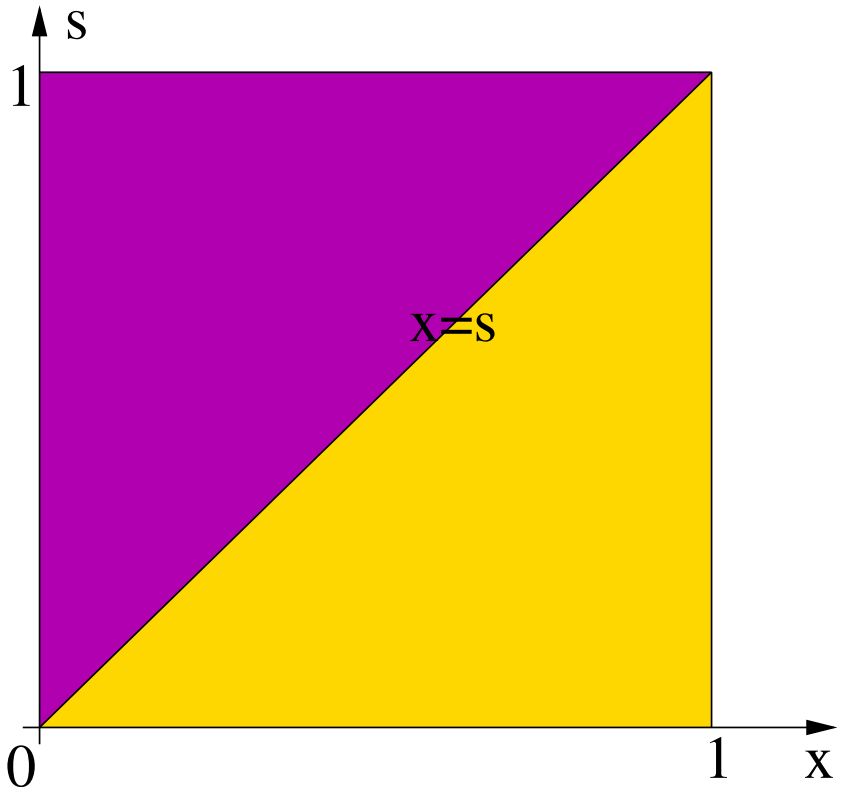

Figure 7.1: The domain on which the Green's function is defined.

$Lg = 0$ there. This proves (3). It remains to prove (5). This is also a straightforward calculation:

$$g_x(x,s) = \begin{cases} v'(x)u(s), & 0 \le s \le x \le 1 \text{ gold triangle,} \\ u'(x)v(s), & 0 \le x \le s \le 1 \text{ purple rectangle.} \end{cases}$$

Thus

$$g_x(s^+,s) - g_x(s^-,s) = -u'(s)v(s) + v'(s)u(s) = W(u,v) = -\frac{1}{p},$$

which finishes the proof. ■

Example: Consider the constant coefficient problem

$$\begin{cases} -ky'' &= f(x) \\ y(0) = 0, & y(1) = 0. \end{cases}$$

Here k is a positive constant. Let's follow the steps outlined above and construct the Green's function.

1. find u, which satisfies the initial-value problem

$$\begin{cases} -ku'' &= 0 \\ u(0) &= 0 \end{cases} \quad \Rightarrow \quad u = Ax,$$

where A is a constant.

2. find v, which satisfies the initial-value problem

$$\left\{ \begin{array}{rcl} -kv'' & = & 0 \\ v(1) & = & 0 \end{array} \right. \quad \Rightarrow \quad v = B(1-x),$$

where B is a constant.

3. Next we need to ensure that the Wronskian of these two solutions is $-1/p = -1/k$. We have

$$\begin{aligned} W &= uv' - u'v \\ &= Ax(-B) - AB(1-x) \\ &= -AB. \end{aligned}$$

Thus we need to impose

$$-AB = -\frac{1}{k} \;\Rightarrow\; AB = \frac{1}{k}.$$

There's some freedom to solve this. We can choose $A = 1$, $B = 1/k$. Then $u = x$, $v = (1-x)/k$. Note that this choice of constants does not influence the Green's function which only depends on the product of u and v.

Finally, we find that the Green's function is given by

$$g(x,s) = \left\{ \begin{array}{ll} \frac{1}{k}(1-x)s & 0 \le s \le x \le 1 \\ \frac{1}{k}(1-s)x & 0 \le x \le s \le 1 \end{array} \right. .$$

A plot of this Green's function is shown in Fig. 7.2.

The following theorem provides practical conditions on when a Green's function exists for our problem.

Theorem 24 *The Green's function exists if $p(x) > 0$, $q(x) \ge 0$, and $h_0 \ge 0$, $h_1 \ge 0$.*

Proof. We need to show that under these conditions the homogeneous problem

$$Ly = 0, \quad y(0) = h_0 y'(0), \quad y(1) = -h_1 y'(1)$$

only has the zero solution. We have

$$\begin{aligned} & -(py')' + qy = 0 \\ \Rightarrow \quad & -(py')'y + qy^2 = 0 \\ \Rightarrow \quad & -\int_0^1 (py')'y dx + \int_0^1 qy^2 dx = 0 \\ \Rightarrow \quad & -py'y|_0^1 + \int_0^1 \left(py'^2 + qy^2 \right) dx = 0 \\ \Rightarrow \quad & p(1)y'^2(1)h(1) + p(0)y'^2(0)h(0) + \int_0^1 \left(py'^2 + qy^2 \right) dx = 0. \end{aligned}$$

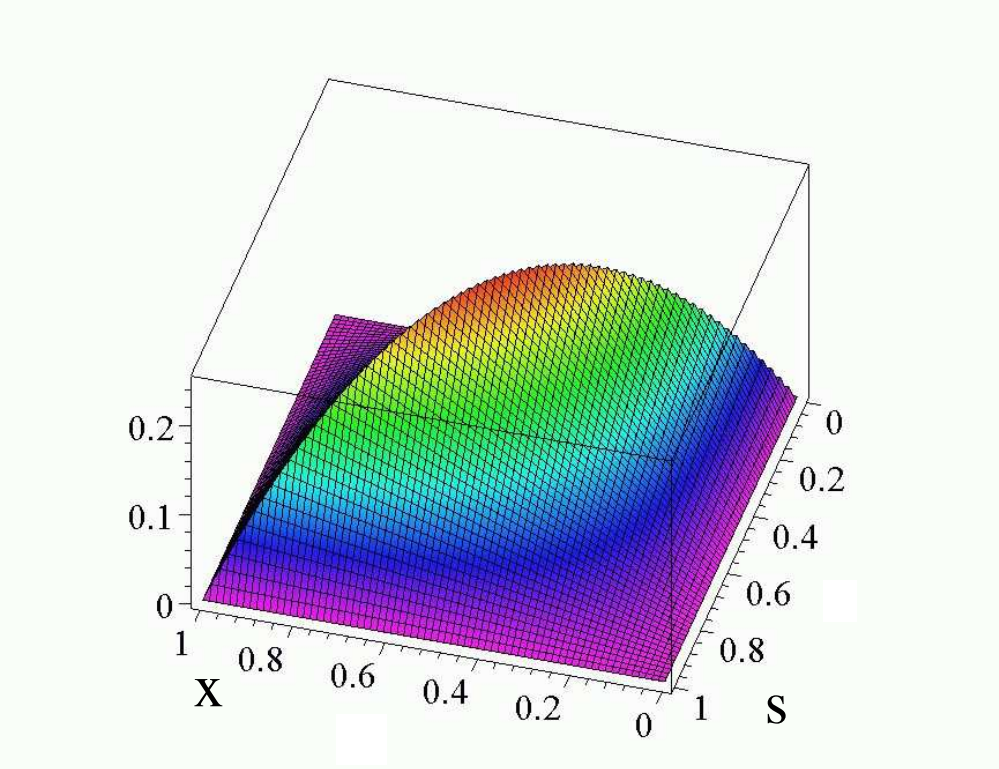

Figure 7.2: The Green's function for the constant coefficient example.

All terms in this last expression are positive, so all terms need to vanish individually. In particular we need ${y'}^2(0) = 0$, which implies that $y(x)$ is constant. This constant is forced to be zero by the boundary conditions. This concludes the proof. ■

3. The Sturm-Liouville problem as an integral equation

Next we recast the general Sturm-Liouville eigenvalue problem in integral form. It is this form that we will use later on.

As a starting point we have

$$\begin{cases} Ly - \lambda r(x)y = f(x) & 0 < x < 1 \\ y(0) = h_0 y'(0), y(1) = -h_1 y'(1) \end{cases}.$$

Any solution of

$$Ly = f(x) + \lambda r(x)y$$

satisfies

$$y = \int_0^1 g(x,s)\left[f(s) + \lambda r(s)y(s)\right] ds,$$

or

$$y(x) = \lambda \int_0^1 g(x,s)r(s)y(s)ds + F(x),$$

where

$$F(x) = \int_0^1 g(x,s)f(s)ds.$$

This problem is completely equivalent with the original one. The integral equation we have obtained is known as a **Fredholm integral equation of the second kind** with kernel $g(x,s)r(s)$.

Note that the kernel of this integral equation can be symmetrized. Assuming that $r(x)$ (which plays the role of a weight function) is positive, we have that

$$\sqrt{r(x)}y = \lambda \int_0^1 g(x,s)\sqrt{r(x)}\sqrt{r(s)}\,\sqrt{r(s)}y(s)ds + F(x)\sqrt{r(x)}.$$

Let's introduce

$$z(x) = \sqrt{r(x)}y(x).$$

Then

$$z(x) = \lambda \int_0^1 k(x,s)z(s)ds + G(x),$$

with

$$k(x,s) = g(x,s)\sqrt{r(x)}\sqrt{r(s)},$$
$$G(x) = F(x)\sqrt{r(x)}.$$

In the immediate future, we'll see more about Fredholm integral equations, setting us up for using all of it in the context of partial differential equations.

7.2 Inner product spaces

Let's introduce a somewhat more abstract setting. This will allow us to do things in more generality.

Definition (Inner product space). A linear space V is an **inner product space** if for each $u, v \in V$, there is a scalar $\langle u, v\rangle$ (their **inner product**) such that for all $u, v, w \in V$
(1) $\langle u, v\rangle = \langle v, u\rangle^*$ ($*$ denotes complex conjugate),
(2) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$,
(3) $\langle \alpha u, v\rangle = \alpha \langle u, v\rangle$, for all $\alpha \in \mathbb{C}$,
(4) $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0$ implies $v = 0$.

With the inner product, we define the norm of $v \in V$ to be

$$||v|| = \sqrt{\langle v, v\rangle}.$$

If V is complete with respect to this norm, then V is a Hilbert space.

Let's give a few examples.

Example: $\mathbb{R}^n$ is a Hilbert space with respect to the inner product defined by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The norm induced by this inner product is the regular Euclidean norm:

$$||x|| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Example: The space of all continuous functions on an interval $[a, b]$ is denoted by $C([a, b])$. It is an inner product space with

$$\langle f, g\rangle = \int_a^b f(x)g^*(x)dx,$$

where f and g are continuous functions on $[a, b]$. The induced norm is

$$||f|| = \sqrt{\langle f, f\rangle} = \sqrt{\int_a^b |f|^2(x)dx}.$$

Using this norm, $C([a, b])$ is not a Hilbert space.

Example: The space of all square-integrable functions on the interval $[a, b]$ is denoted as $L^2([a, b])$. It is a Hilbert space[1] with as inner product

$$\langle f, g\rangle = \int_a^b f(x)g^*(x)dx,$$

where f and g are square integrable. The induced norm is

$$||f|| = \sqrt{\langle f, f\rangle} = \sqrt{\int_a^b |f|^2(x)dx}.$$

Theorem 25 (The Schwarz inequality) *For any $u, v \in V$ we have*

$$|\langle u, v\rangle| \leq ||u||\,||v||.$$

[1]Strictly speaking we need to use the Lebesgue integral here, instead of the Riemann integral. We won't go further into this.

Proof. Let t be a real scalar. Consider

$$\begin{aligned} ||u+tv||^2 &= \langle u+tv, u+tv\rangle \\ &= ||u||^2 + t^2||v||^2 + t\langle u,v\rangle + t\langle v,u\rangle . \end{aligned}$$

This is a quadratic expression in t. Since this is always positive or zero we need its discriminant to be negative or zero:

$$\begin{aligned} & & \Delta &= 4(\Re\langle u,v\rangle)^2 - 4||u||^2||v||^2 \\ \Rightarrow & & (\Re\langle u,v\rangle)^2 &\leq ||u||^2||v||^2 \\ \Rightarrow & & \Re\langle u,v\rangle &\leq ||u||\,||v||. \end{aligned}$$

This inequality is valid for all u and v. Let

$$v = \langle u,w\rangle\, w,$$

where $w \in V$. The above inequality becomes

$$\begin{aligned} & & \Re\langle u, \langle u,w\rangle w\rangle &\leq ||u||\,||\langle u,w\rangle w|| \\ \Rightarrow & & \Re\langle u,w\rangle^*\langle u,w\rangle &\leq |\langle u,w\rangle|\,||u||\,||w|| \\ \Rightarrow & & \Re|\langle u,w\rangle|^2 &\leq |\langle u,w\rangle|\,||u||\,||w|| \\ \Rightarrow & & |\langle u,w\rangle|^2 &\leq |\langle u,w\rangle|\,||u||\,||w||. \end{aligned}$$

If $\langle u,w\rangle = 0$, then the Schwarz inequality is clearly satisfied. If $\langle u,w\rangle \neq 0$ then we get

$$|\langle u,w\rangle| \leq ||u||\,||w||,$$

which is what we had to prove. ■

A few remarks are in order.

- For $V = \mathbb{R}^n$, the Schwarz inequality states that $\cos\theta \leq 1$. As usual, you should check this.

- If you're interested in quantum mechanics, you spend a lot of time living in a Hilbert space. The Schwarz inequality has a physical interpretation as the Heisenberg uncertainty principle:

  $$\Delta q\,\Delta p \geq \frac{\hbar}{2}.$$

  This states that the product of the uncertainties in determining position and momentum is bounded from below: if you know the position of a particle accurately, your uncertainty on its momentum is huge. If you know momentum accurately, you have no idea where the particle is.

Theorem 26 (Bessel's inequality) *Let $\{v_1, v_2, \ldots\}$ be an orthonormal sequence in V. This means that $\langle v_i, v_j \rangle = \delta_{ij}$. Then*

$$\sum_{n=1}^{\infty} |\langle v, v_n \rangle|^2 \leq ||v||^2,$$

for any $v \in V$.

Proof. The proof is straightforward calculation.

$$\begin{aligned}
||v - \sum_{k=1}^{N} \langle v, v_k \rangle v_k||^2 &= \left\langle v - \sum_{k=1}^{N} \langle v, v_k \rangle v_k, v - \sum_{n=1}^{N} \langle v, v_n \rangle v_n \right\rangle \\
&= ||v||^2 - \sum_{k=1}^{N} \langle v, v_k \rangle \langle v_k, v \rangle - \sum_{n=1}^{N} \langle v, v_n \rangle^* \langle v, v_n \rangle + \\
&\quad \sum_{k=1}^{N} \sum_{n=1}^{N} \langle v, v_k \rangle \langle v, v_n \rangle^* \langle v_k, v_n \rangle \\
&= ||v||^2 - 2 \sum_{k=1}^{N} |\langle v, v_k \rangle|^2 + \sum_{k=1}^{N} |\langle v, v_k \rangle|^2 \\
&= ||v||^2 - \sum_{k=1}^{N} |\langle v, v_k \rangle|^2.
\end{aligned}$$

Letting $N \to \infty$ we obtain

$$\begin{aligned}
& \lim_{N \to \infty} ||v - \sum_{k=1}^{N} \langle v, v_k \rangle v_k||^2 = ||v||^2 - \sum_{k=1}^{\infty} |\langle v, v_k \rangle|^2 \\
\Rightarrow \qquad & \sum_{k=1}^{\infty} |\langle v, v_k \rangle|^2 \leq ||v||^2,
\end{aligned}$$

which is what we had to prove. ■

An immediate consequence of this proof is the following:

Corollary 3 (Parseval's identity) *If $\sum_{k=1}^{\infty} \langle v, v_k \rangle v_k \to v$ in our norm, then*

$$\sum_{k=1}^{\infty} |\langle v, v_k \rangle|^2 = ||v||^2.$$

The following result does not require that we have an inner product space. However, the generality of the result mandates that we include in this section, rather than include it with more specific sections later.

Let R denote an operator on V: the operator R maps elements of V to different[2] elements of V:

$$R : V \to V : v \in V \to Rv \in V.$$

We define the norm of an operator as

$$||R|| = \sup_{||u||=1} ||Ru||.$$

We have the following result.

Theorem 27 *The operator $I - R$ has an inverse if $||R|| < 1$ and V is complete. Here I denotes the identity operator.*

Proof. Consider the operator

$$\hat{R} = I + R + R^2 + \ldots.$$

Formally

$$\hat{R}(I - R) = I + R + R^2 + \ldots - R - R^2 - \ldots = I.$$

Thus, if we can show that the series defining $\hat{R}$ converges, then $\hat{R}$ is the inverse of $I - R$. We'll do this by showing that $\{\hat{R}_n = I + R + R^2 + \ldots + R^n\}$ is a Cauchy sequence. If we can do this, then convergence follows since the space if complete.

Without loss of generality, assume that $n > m$. Then

$$\begin{aligned}
\hat{R}_n - \hat{R}_m &= R^{m+1} + \ldots + R^n \\
&= R^{m+1}\left(I + R + \ldots + R^{n-m-1}\right) \\
\Rightarrow \qquad ||\hat{R}_n - \hat{R}_m|| &= ||R^{m+1}\left(I + R + \ldots + R^{n-m-1}\right)|| \\
&\leq ||R^{m+1}||\,||I + R + \ldots + R^{n-m-1}|| \\
&\leq ||R||^{m+1}\left(1 + ||R|| + ||R||^2 + \ldots + ||R||^{n-m-1}\right) \\
&\to 0, \quad \text{as } m, n \to \infty
\end{aligned}$$

provided $||R|| < 1$. Note that we have repeatedly used that

$$||AB|| \leq ||A||\,||B||,$$

for two operators $A, B : V \to V$. There's a good chance this will be a homework problem. You could proactively check it. This concludes our proof. ■

[2]Unless it's the identity operator.

7.3 The Fredholm alternative

Let's return to our study of integral equations. A Fredholm integral equation has the form

$$y(x) = \int_0^1 k(x,s)y(s)ds + f(x).$$

We can write this in operator form as

$$\begin{aligned} & y(x) = Ky(x) + f(x) \\ \Rightarrow \quad & y = Ky + f \\ \Rightarrow \quad & y - Ky = f \\ \Rightarrow \quad & (I-K)y = f, \end{aligned}$$

which is the generic form of a linear nonhomogeneous problem. The operator K is given by

$$Ky(x) = \int_0^1 k(x,s)y(s)ds.$$

In this section we will establish the **Fredholm alternative** using various kernels $k(x,s)$. Let's preview the result we'll get (we'll be more precise later):

The Fredholm alternative: either the equation

$$(I-K)y = f$$

has a unique solution for each f, or else[3]

$$(I-K)y = 0$$

has at least one nontrivial solution.

You have already encountered the Fredholm alternative in many settings:

- **Linear algebra:** For the nonhomogeneous equation $Ax = b$ to have a unique solution, it is necessary that the homogeneous problem $Ax = 0$ only has the trivial solution $x = 0$.
- **Linear ordinary differential equations:** The solution of an inhomogenous linear ordinary differential equation are given by the solution of the homogeneous problem, plus a particular solution.

We'll put $\lambda = 1$ from now on, for convenience. We can always absorb it in the definition of K.

[3]Here's your alternative

1. Proof for degenerate kernels

First we prove the Fredholm alternative for **degenerate kernels**. A degenerate kernel is a kernel of the form

$$k(x,s) = \sum_{j=1}^{n} a_j(x) b_j(s).$$

The integral equation becomes

$$\begin{aligned} y(x) &= f(x) + \int_0^1 \sum_{j=1}^{n} a_j(x) b_j(s) y(s) ds \\ &= f(x) + \sum_{i=1}^{n} a_j(x) \langle b_j, y \rangle . \end{aligned}$$

Here's a trivial observation from the above equation: if we can find $\langle b_j, y\rangle$ for all functions b_j in the degenerate kernel $k(x,s)$, then we can find y from the above. This observation is actually useful: let's multiply the above equation by $b_i(x)$ and integrate from 0 to 1. Then

$$\langle b_i, y\rangle = \langle b_i, f\rangle + \sum_{j=1}^{n} \langle b_i, a_j\rangle \langle b_j, y\rangle .$$

Define the vectors Y and F as

$$\begin{aligned} Y &= \left(\langle b_1, y\rangle, \ldots, \langle b_n, y\rangle\right)^T, \quad \text{and} \\ F &= \left(\langle b_1, f\rangle, \ldots, \langle b_n, f\rangle\right)^T . \end{aligned}$$

Lastly, the matrix C is given by

$$C = (C_{ij})_{i,j=1}^{n}, \quad \text{with} \quad C_{ij} = \langle b_i, a_j\rangle .$$

The above equation is rewritten as

$$\begin{aligned} & Y = F + CY \\ \Rightarrow \qquad & (I - C)Y = F. \end{aligned}$$

In other words, we have reduced the problem to one of linear algebra.

1. First we show that the statement "$(I-C)Y = 0$ only has the trivial solution $Y = 0$" implies the statement "$(I-K)y = 0$ only has the trivial solution $y = 0$". Indeed, if Y is the only solution of $(I-C)Y = 0$, then it follows from the above that $y = 0$.

2. Having established the above, it follows that it suffices for us to prove that "$(I-C)Y = 0$ only has the trivial solution $Y = 0$" implies "$(I-K)y = f$ has a unique solution".

   This is straightforward: If $(I-C)Y = 0$ only has the trivial solution $Y = 0$, then $I-C$ is an invertible matrix and the only solution of $(I-C)Y = F$ is $Y = (I-C)^{-1}F$. Having determined Y uniquely, we may use our previous equation to determine y uniquely.

This finishes the proof for degenerate kernels. ■

Remark: the above is the only thing we have to prove: if a nontrivial solution of $(I-K)y=0$ exists, then clearly, any solution of $(I-K)y=f$ is not unique if one exists. We could always add the nontrivial solutions of the homogeneous problem to it.

2. Continuous kernels

In this section we prove the Fredholm alternative for kernels that are continuous functions of x and s.

Let $k(x,s)$ be continuous on $[0,1]\times[0,1]$. Let K be the operator defined by

$$Ky=\int_0^1 k(x,s)y(s)ds.$$

We see that K maps functions that are continuous on $[0,1]$ to functions that remain continuous on $[0,1]$:

$$K:\ y\in C([0,1])\to Ky\in C([0,1]).$$

As a norm we use

$$||y||=\max_{x\in[0,1]}|y(x)|,$$

implying that we are dealing with functions that are *uniformly* continuous.

The first step of our proof consists in approximating $k(x,s)$ by a polynomial $p(x,s)$ such that

$$|k(x,s)-p(x,s)|<\frac{1}{\rho},$$

with $\rho>1$. This is easily done, using the polynomial form of the Weierstrass approximation theorem. Let P denote the integral operator corresponding to $p(x,s)$.

Let

$$r(x,s)=k(x,s)-p(x,s),$$

with R being the integral operator with $r(x,s)$ as its kernel:

$$Ry=\int_0^1 r(x,s)y(s)ds.$$

Then

$$\begin{aligned} |Ry| &= \left| \int_0^1 r(x,s)y(s) \right| \\ &\leq \int_0^1 |r(x,s)||y(s)|ds \\ &\leq ||y|| \int_0^1 |k(x,s) - p(x,s)|ds \\ &\leq \frac{||y||}{\rho} \\ \Rightarrow \qquad ||R|| &\leq \frac{1}{\rho} < 1. \end{aligned}$$

Now consider the equation

$$\begin{aligned} & (I-K)y = f \\ \Leftrightarrow \quad & (I-R-P)y = f \\ \Leftrightarrow \quad & (I-R)y - Py = f \\ \Leftrightarrow \quad & y - (I-R)^{-1}Py = (I-R)^{-1}f \\ \Leftrightarrow \quad & (1-L)y = \hat{f}, \end{aligned}$$

where $\hat{f} = (I-R)^{-1}f$ and $L = (I-R)^{-1}P$. We've used that $I-R$ has an inverse since our space is complete and $\rho > 1$ implies that $||R|| < 1$.

If we can establish the Fredholm alternative for this last equation, it holds for the original problem: these two problems are completely equivalent as the above shows. We now show that our last equation has a degenerate kernel. We have

$$Ly = (I-R)^{-1} \int_0^1 p(x,s)y(s)ds.$$

It is clear that $p(x,s)$ is a degenerate kernel:

$$\begin{aligned} p(x,s) &= A_n(s)x^n + \cdots + A_1(s)x + A_0(s) \\ &= \sum_{i=0}^{n} A_i(s)B_i(x), \end{aligned}$$

with

$$B_i(x) = x^i.$$

Now to show that L has a degenerate kernel:

$$\begin{aligned} Ly &= (I-R)^{-1}\int_0^1 \sum_{i=0}^{n} A_i(s)B_i(x)y(s)ds \\ &= (I-R)^{-1}\int_0^1 \sum_{i=0}^{n} B_i(x)A_i(s)y(s)ds \\ &= \int_0^1 \sum_{i=0}^{n} \left[(I-R)^{-1}B_i(x)\right] A_i(s)y(s)ds, \end{aligned}$$

which has a degenerate kernel! This concludes the proof of the Fredholm alternative for continuous kernels. ■

Remark: completely analogous to the previous two proofs, we can prove the Fredholm alternative for kernels that are square integrable (this requires Lebesgue integration, instead of Riemann integration), or that are weakly singular (*i.e.*, have singularities that are integrable).

3. The strong form of the Fredholm alternative

The Fredholm alternative may be rephrased in a somewhat more comprehensive form than the one we've been using to far. Let's illustrate this for the linear algebra setting first, where you've probably seen it all before.

Strong form of the Fredholm alternative: matrix formulation

Consider the matrix problem

$$Ax = b$$

and its adjoint problem

$$A^\dagger y = c,$$

where $A^\dagger$ is the adjoint matrix of A, determined by

$$\left\langle x, A^\dagger y\right\rangle = \left\langle Ax, y\right\rangle.$$

If A is real, then $A^\dagger$ is the transposed matrix: $A^\dagger = A^T$. The following theorem holds:

Theorem 28 (Fredholm alternative (strong form; matrix setting)) *If the homogeneous equations $A^\dagger y = 0$ and $Ax = 0$ only have the trivial solution, then there is a unique solution to the nonhomogeneous equations.*

If there is a nontrivial solution to the homogeneous problem, then

a) $Ax = 0$ and $A^\dagger y = 0$ have the same number of linearly independent solutions (dim ker A= dim ker $A^\dagger$),

b) $Ax = b$ has a solution if and only if b is orthogonal to all solutions of $A^\dagger y = 0$ ($b \perp \ker A^\dagger$),

c) The general solution of $Ax = b$ (if it exists) is given by a particular solution of $Ax = b$ added to the most general solution of the homogeneous problem $Ax = 0$.

Strong form of the Fredholm alternative: integral operator setting

Completely analogous to the matrix setting, we now formulate the strong form of the Fredholm alternative for integral equations of the second kind.

Consider the Fredholm integral equation of the second kind

$$(I - K)y = f,$$

together with its adjoint equation

$$(I - K^\dagger)y = g.$$

Here the operator $K^\dagger$ is the adjoint of K. It is determined by

$$\langle Ky_1, y_2\rangle = \left\langle y_1, K^\dagger y_2\right\rangle.$$

In terms of integral kernels, this implies that the kernel of the integral operator $K^\dagger$ is $k^\dagger(x, s) = k(s, x)^*$. We have the following theorem.

Theorem 29 (Fredholm alternative (strong form; integral operator setting)) *If the homogeneous equations $(I - K)y = 0$ and $(I - K^\dagger)y = 0$ only have the trivial solution, then there is a unique solution to the nonhomogeneous equations.*

If there is a nontrivial solution to the homogeneous problem, then

a) $(I-K)y = 0$ and $(I-K^\dagger)y = 0$ have the same number of linearly independent solutions (dim ker $(I - K)$= dim ker $(I - K^\dagger)$),

b) $(I-K)y = f$ has a solution if and only if f is orthogonal to all solutions of $(I-K^\dagger)y = 0$ ($f \perp$ ker $(I - K^\dagger)$),

c) The general solution of $(I - K)y = f$ (if it exists) is given by a particular solution of $(I - K)y = f$ added to the most general solution of the homogeneous problem $(I - K)y = 0$.

These additions to the form of the Fredholm alternative follow almost immediately from our previous considerations.

7.4 Eigenvalues and eigenfunctions

According to the Fredholm alternative, the equation

$$(I - \lambda K)y = f$$

has a solution for all f, or else the homogeneous equation

$$(I - \lambda K)y = 0$$

has a nontrivial solution. The values of λ for which this latter case occurs are called **eigenvalues** of the kernel $k(x, s)$ or of the operator K. The nontrivial solutions of the homogeneous equation corresponding to an eigenvalue are called its **eigenfunctions**. The maximal number of linearly independent eigenfunctions corresponding to an eigenvalue λ is called its **multiplicity**. The following theorem is almost too easy to prove.

Theorem 30 *For an eigenvalue λ of multiplicity n there are n orthonormal eigenfunctions, which form a basis for the eigenspace of λ.*

Proof. This is immediately achieved using the Gram-Schmidt procedure. ■

More familiar results are the topic of the next theorem.

Theorem 31 *For a self-adjoint operator (*i.e.*, $K = K^\dagger$), all eigenvalues are real, and eigenfunctions corresponding to distinct eigenvalues are orthogonal.*

Proof. Assume λ is an eigenvalue of K. Then there exists a $y \neq 0$ such that

$$y = \lambda K y.$$

Assume that $\lambda \neq 0$. Then

$$\begin{aligned}
\langle y, y\rangle &= \langle \lambda K y, y\rangle \\
&= \lambda \langle y, Ky\rangle \\
&= \lambda \left\langle y, \frac{y}{\lambda}\right\rangle \\
&= \frac{\lambda}{\lambda^*} \langle y, y\rangle \\
\Rightarrow \quad (\lambda^* - \lambda)\langle y, y\rangle &= 0 \\
\Rightarrow \quad (\lambda^* - \lambda)||y||^2 &= 0 \\
\Rightarrow \quad \lambda &= \lambda^* \\
\Rightarrow \quad \lambda &\in \mathbb{R}.
\end{aligned}$$

This proves that all eigenvalues of self-adjoint integral operators are real.

Next, let

$$y = \lambda K y,$$

and

$$z = \mu K z,$$

with $\lambda \neq \mu$. Then

$$\begin{aligned} \langle y, z \rangle &= \langle \lambda K y, z \rangle \\ &= \lambda \langle y, K z \rangle . \end{aligned}$$

On the other hand, we also have

$$\begin{aligned} \langle y, z \rangle &= \langle y, \mu K z \rangle \\ &= \mu^* \langle y, K z \rangle \\ &= \mu \langle y, K z \rangle . \end{aligned}$$

We conclude that

$$\lambda \langle y, K z \rangle = \mu \langle y, K z \rangle \quad \Rightarrow \quad \langle y, K z \rangle = 0 \quad \Rightarrow \quad \langle y, z \rangle = 0,$$

since $\lambda \neq \mu$. This concludes our proof. ■

If $k(x, s)$ is real and $k(x, s) = k(s, x)$ then k is called symmetric. Symmetric kernels are also self-adjoint.

Theorem 32 *Let K be self-adjoint. Let $\lambda_k, k = 1, 2, \ldots$ be a sequence of ordered eigenvalues of K, with eigenvalues of multiplicity n repeated n times. Then the corresponding list of eigenfunctions is an orthogonal set, and can be made orthonormal.*

Proof. This is an immediate consequence of our two previous theorems. ■

Next we note that it follows from the proof of the Fredholm alternative for continuous kernels that the matrix of the degenerate kernel $C(\lambda) = (I - \lambda R)^{-1} \lambda P$ (we've reinserted the appropriate λ dependence) is analytic as a function of λ as long as $\lambda < \rho$ (so that $||\lambda R|| < 1$). The same is true of $\det C(\lambda)$, which is a polynomial combination of analytic functions.

We use this knowledge in the following theorem.

Theorem 33 *The set of eigenvalues of $k(x, s)$ is at most countable. If the set is infinite, $|\lambda_n| \to \infty$ as $n \to \infty$.*

Proof. Fix $\rho > 1$. This determines a circular area for λ in which our expansion for the inverse of $I - \lambda R$ is valid, and analytic in λ. As for the degenerate kernels, we can solve the problem by constructing a linear system for the expansion coefficients. The matrix of this linear system is $C(\lambda)$, introduced above. It follows that nontrivial solutions to the homogeneous problem exist for those values of λ for which

$$\det C(\lambda) = 0.$$

We may refer to this equation as the eigenvalue equation. As stated above, the left-hand side is an analytic function of λ as long as $\lambda < \rho$. Since an analytic function has at most a finite number of zeroes in a bounded region of the complex plane, it follows that the eigenvalue equation has a finite number of solutions in $\lambda < \rho$. Equivalently, it follows that the operator K only has a finite number of eigenvalues in $\lambda < \rho$. Dividing the complex plane in a countable number of concentric annuli, and letting $\rho \to \infty$, we reach our conclusion. This concludes the proof of the theorem. ■

In principle, we still have to prove that eigenvalues actually exist. We will not prove that in this course. It is a rather obvious thing to believe, given what we know about linear equations in various settings, especially matrix problems. As we're not proving theorems right now, let's also not prove the following theorem.

Theorem 34 (Rayleigh-Ritz quotient) *Assume that K is self-adjoint, and not identically equal to zero. Then K has an eigenvalue of smallest modulus λ_1, and*

$$\frac{1}{\lambda_1^2} = \max_{y \neq 0} \frac{\langle Ky, Ky \rangle}{\langle y, y \rangle}.$$

7.5 Eigenfunction expansions

In this section we prove the Hilbert-Schmidt theorem, which is a statement about the completeness of the set of eigenfunctions of a self-adjoint kernel.

1. Preliminaries

We begin with a lemma.

Lemma 6 *Let k be self-adjoint. Any eigenfunction of*

$$k^{(n)}(x,s) = k(x,s) - \sum_{j=1}^{n} \frac{\phi_j(x)\phi_j^*(s)}{\lambda_j}$$

is an eigenfunction of $k(x,s)$ with the same eigenvalue. Here $\phi_j(x)$ denotes an eigenfunction of $k(x,s)$ corresponding to eigenvalue λ_j. These eigenvalues are considered to be ordered by modulus, with multiplicity taken into account.

Proof. The operator $K^{(n)}$ corresponding to the kernel $k^{(n)}(x,s)$ is defined by

$$\begin{aligned} K^{(n)}z &= Kz - \sum_{j=1}^{n} \int_0^1 \frac{\phi_j(x)\phi_j^*(s)}{\lambda_j} z(s)ds \\ &= Kz - \sum_{j=1}^{n} \frac{\langle z, \phi_j \rangle}{\lambda_j} \phi_j. \end{aligned}$$

If ψ is an eigenfunction of $K^{(n)}$ with eigenvalue μ, then $\psi = \mu K^{(n)}\psi$, and for all $r = 1, 2, \ldots, n$ we have

$$\begin{aligned}
\langle \psi, \phi_r \rangle &= \mu \left\langle K^{(n)}\psi, \phi_r \right\rangle \\
&= \mu \langle K\psi, \phi_r \rangle - \mu \sum_{j=1}^{n} \frac{\langle \psi, \phi_j \rangle \langle \phi_j, \phi_r \rangle}{\lambda_j} \\
&= \mu \langle K\psi, \phi_r \rangle - \mu \frac{\langle \psi, \phi_r \rangle}{\lambda_r} \\
&= \mu \langle \psi, K\phi_r \rangle - \frac{\mu}{\lambda_r} \langle \psi, \phi_r \rangle \\
&= \frac{\mu}{\lambda_r} \left(\langle \psi, \phi_r \rangle - \langle \psi, \phi_r \rangle \right) \\
&= 0.
\end{aligned}$$

We conclude that

$$K^{(n)}\psi = K\psi - \sum_{j=1}^{n} \frac{\langle \psi, \phi_j \rangle}{\lambda_j} \phi_j = K\psi,$$

from which it follows that

$$\psi = \mu K\psi,$$

which is what we had to prove. ■

This leads us to the next theorem.

Theorem 35 *Let k be self-adjoint. Then $k(x, s)$ is degenerate if and only if it has a finite number of eigenvalues.*

Proof. First, we prove that if k is degenerate it only has a finite number of eigenvalues. We already know this, since in this case the problem reduces to linear algebra.

Next, we show that if k gives rise to only a finite number of eigenvalues, then it is a degenerate kernel. Suppose k has a finite number n of eigenvalues, with corresponding eigenfunctions. Construct $k^{(n)}$ as in the lemma above. We claim that $k^{(n)} \equiv 0$, from which it follows immediately that

$$k(x, s) = \sum_{j=1}^{n} \frac{\phi_j(x)\phi_j^*(s)}{\lambda_j}.$$

This kernel is clearly degenerate. Thus, if we can prove that $k^{(n)} \equiv 0$, then we are done.

Assume that $k^{(n)} \neq 0$. Then it is self-adjoint and has an eigenvalue μ with eigenvector ψ. It follows from the lemma that ψ is an eigenfunction of k with the same eigenvalue. This implies that ψ must be a linear combination of the finite number of eigenfunctions of k:

$$\psi = \sum_{j=1}^{n} a_j \phi_j.$$

The calculation in the proof of the previous lemma shows that for all $j = 1, 2, \ldots, n$ we have

$$\langle \psi, \phi_j \rangle = 0.$$

This implies that $a_j = 0$ for all j, from which $\psi \equiv 0$. This is a contradiction, from which the theorem follows. ■

The next theorem is the last one we need before we can prove our completeness result.

Theorem 36 *Let k be a self-adjoint, non-degenerate kernel with eigenvalues $\{\lambda_1, \lambda_2, \ldots\}$ (including multiplicities) and orthonormal eigenfunctions $\{\phi_1, \phi_2, \ldots\}$. Then $k^{(n)}(x, s)$ has eigenvalues $\{\lambda_{n+1}, \lambda_{n+2}, \ldots\}$ and for any g*

$$\lim_{n\to\infty} \left\langle K^{(n)} g, K^{(n)} g \right\rangle = 0.$$

Proof. Let $r > n$. Then

$$\begin{aligned} K^{(n)}\phi_r &= K\phi_r - \sum_{j=1}^{n} \frac{\langle \phi_r, \phi_j \rangle}{\lambda_j} \phi_j \\ &= K\phi_r, \end{aligned}$$

since $\langle \phi_r, \phi_j \rangle = 0$, for $j = 1, 2, \ldots, n$. This implies that

$$\phi_r = \lambda_r K \phi_r = \lambda_r K^{(n)} \phi_r.$$

Furthermore, all eigenvalues and eigenfunctions of $K^{(n)}$ are obtained this way. Indeed, suppose there is an eigenfunction ψ of $K^{(n)}$ that is not spanned by $\{\phi_{n+1}, \phi_{n+2}, \ldots\}$. We know that ψ is also an eigenfunction of K:

$$\psi = \mu K^{(n)} \psi = \mu K \psi.$$

Since ψ is not spanned by $\{\phi_{n+1}, \phi_{n+2}, \ldots\}$, it must be spanned by $\{\phi_1, \phi_2, \ldots, \phi_n\}$:

$$\psi = \sum_{j=1}^{n} a_j \phi_j.$$

But for the eigenfunctions of $K^{(n)}$ we know that $a_j = 0$, which provides the required contradiction.

Lastly, we show that

$$\lim_{n\to\infty} \left\langle K^{(n)} g, K^{(n)} g \right\rangle = 0,$$

for any $g \in V$. From the Rayleigh-Ritz quotient we have that

$$\left\langle K^{(n)} g, K^{(n)} g \right\rangle \leq \frac{\langle g, g \rangle}{\lambda_{n+1}^2},$$

since

$$\frac{1}{\lambda_{n+1}^2} = \max_{g \neq 0} \frac{\left\langle K^{(n)}g, K^{(n)}g \right\rangle}{\langle g, g \rangle}.$$

It follows from this that

$$\lim_{n\to\infty} \left\langle K^{(n)}g, K^{(n)}g \right\rangle = 0,$$

since $|\lambda_{n+1}| \to \infty$ as $n \to \infty$. This concludes the proof. ■

2. The Hilbert-Schmidt theorem

Theorem 37 (Hilbert-Schmidt) *Let $k(x,s) \neq 0$ be the kernel of a self-adjoint kernel K, so that $k(x,s)$ is continuous on $0 \leq x, s \leq 1$. Then each function in the range of K has an eigenfunction expansion which converges in the least-squares sense. Thus, if*

$$f = Kg,$$

then

$$\sum_{n=1}^{\infty} \langle f, \phi_n \rangle \phi_n \to f,$$

in a least-squares sense. Here $\{\phi_n, n = 1, 2, \ldots\}$ is the sequence of eigenfunctions of K.

Proof. Let

$$\begin{aligned} S_n &= \sum_{j=1}^{n} \langle f, \phi_j \rangle \phi_j \\ &= \sum_{j=1}^{n} \langle Kg, \phi_j \rangle \phi_j \\ &= \sum_{j=1}^{n} \langle g, K\phi_j \rangle \phi_j \\ &= \sum_{j=1}^{n} \frac{\langle g, \phi_j \rangle}{\lambda_j} \phi_j. \end{aligned}$$

Then

$$\begin{aligned} f - S_n &= Kg - \sum_{j=1}^{n} \frac{\langle g, \phi_j \rangle}{\lambda_j} \phi_j \\ &= K^{(n)}g. \end{aligned}$$

We know that

$$\langle f - S_n, f - S_n \rangle = \left\langle K^{(n)}g, K^{(n)}g \right\rangle \to 0,$$

by the previous theorem. This is what we had to show. ■

Written more explicitly, we have shown that

$$\lim_{n\to\infty} \int_0^1 \left| f(x) - \sum_{j=1}^{n} \langle f, \phi_j \rangle \, \phi_j \right|^2 dx = 0.$$

We should note that this proof can be strengthened to give absolute and uniform convergence of the eigenfunction expansion series of f in the case when f is continuous on $0 \leq x \leq 1$. An immediate consequence of the above is that Parseval's identity holds.

7.6 Sturm-Liouville problems in applications

Let's return to our original Sturm-Liouville set-up.

$$\begin{cases} -(py')' + qy = \lambda r y, & \text{for } x \in (0,1), \\ y(0) = h_0 y'(0), & y(1) = -h_1 y'(1). \end{cases}$$

Here $p(x)$, $p'(x)$, $q(x)$ and $r(x)$ are continuous and $p(x), r(x) > 0$, $q(x) \geq 0$, and $h_0, h_1 \geq 0$.

Under these conditions, we've shown that this problem is equivalent to that of studying the integral equation

$$z(x) = \lambda \int_0^1 k(x,s) z(s) ds,$$

with $z(x) = \sqrt{r(x)} y(x)$, and

$$k(x,s) = g(x,s)\sqrt{r(x)}\sqrt{r(s)},$$

where $g(x,s)$ is the Green's function for the Sturm-Liouville problem:

$$g(x,s) = \begin{cases} u(s)v(x) & 0 \leq s \leq x \leq 1 \\ u(x)v(s) & 0 \leq x \leq s \leq 1 \end{cases}.$$

The two fundamental solutions $u(x)$ and $v(x)$ satisfy the boundary conditions at the left and right sides respectively. They are connected by

$$pW(u,v) = -1.$$

1. Some more lemmas and theorems

Having set up all our integral equation machinery, we can now state the following theorem.

Theorem 38 *If p, $r > 0$, $q \geq 0$ and $h_0, h_1 \geq 0$, and if p, p', r and q are continuous on $[0,1]$, then*
(a) The Sturm-Liouville problem has an infinite number of eigenvalues λ_1, λ_2, ..., with corresponding eigenfunctions y_1, y_2, ...
(b) These eigenvalues are real, and their corresponding eigenfunctions can be chosen real and orthonormal, with weight function $r(x)$:

$$\int_0^1 y_m y_n r(x) dx = \delta_{nm}.$$

(c) If f, f', f'' are continuous on $[0,1]$, and satisfy the boundary conditions, then $f(x)$ has an absolutely and uniformly convergent eigenfunction expansion

$$f(x) = \sum_{n=1}^{\infty} c_n y_n(x),$$

with

$$c_n = \int_0^1 f(x) y_n(x) r(x) dx.$$

Proof.

(a-b) The kernel $k(x,s)$ is real and symmetric, thus self-adjoint. Thus $k(x,s)$ has a (possibly finite) set of eigenvalues (real) with corresponding eigenfunctions, such that

$$\int_0^1 z_m(x) z_n(x) dx = \delta_{nm},$$

from which

$$\int_0^1 y_m(x) y_n(x) r(x) dx = \delta_{nm}.$$

This proves (a-b), apart from the statement that there are an infinite number of eigenvalues and eigenvectors. We'll come back to this.

(c) Since $k(x,s)$ is symmetric and continuous, the Hilbert-Schmidt theorem states that any function in its range has an absolutely and uniformly convergent expansion in terms of $z_1(x)$, $z_2(x)$, ... Define

$$h(x) = -(pf')' + qf.$$

Then $y(x) = f(x)$ solves the boundary-value problem

$$\begin{cases} Ly = h, & 0 < x < 1 \\ y(0) = h_0 y'(0), \quad y(1) = -h_1 y'(1). \end{cases}$$

Thus

$$\begin{aligned} y &= \int_0^1 g(x,s)h(s)ds \\ \Rightarrow \qquad f(x) &= \int_0^1 \frac{k(x,s)}{\sqrt{r(x)}\sqrt{r(s)}}h(s)ds \\ \Rightarrow \qquad \sqrt{r(x)}f(x) &= \int_0^1 k(x,s)\frac{h(s)}{\sqrt{r(s)}}ds. \end{aligned}$$

Thus $\sqrt{r(x)}f(x)$ is in the range of the $k(x,s)$. Using the Hilbert-Schmidt theorem we get

$$\sqrt{r(x)}f(x) = \sum_j \left\langle \sqrt{r(x)}f(x), z_j(x) \right\rangle z_j(x),$$

which is absolutely and uniformly convergent on $[0,1]$. Then

$$f(x) = \sum_j \left\langle \sqrt{r}f, z_j \right\rangle y_j(x) = \sum_j c_j y_j(x).$$

The expansion coefficients are

$$\begin{aligned} c_j &= \int_0^1 \sqrt{r(x)}f(x)z_j(x)dx \\ &= \int_0^1 \sqrt{r(x)}f(x)y_j(x)\sqrt{r(x)}dx \\ &= \int_0^1 f(x)y_j(x)r(x)dx, \end{aligned}$$

which proves (c).

(a) **(cont'd)** At this point we still need to settle that there are an infinite number of eigenvalues, not a finite number. If there would be only a finite number, it follows from our proof of (c) that there would only exist a finite number of linearly independent functions $f(x)$ satisfying the boundary conditions which have two continuous derivatives. But the infinite sequence of functions

$$p_n(x) = (x(1-x))^{2n}$$

satisfies these conditions for $h_0 = 0$ and $h_1 = 0$. You can construct a suitably modified sequence for $h_0 \neq 0$ and $h_1 \neq 0$. This provides a contradiction, establishing the existence of an infinite number of eigenvalues.

This concludes the proof of the theorem. ■

Theorem 39 *Under the assumptions of the previous theorem, all eigenvalues are positive, and each eigenvalue is simple,* i.e., *has multiplicity one.*

Proof. We have

$$\begin{aligned}
\lambda_n r(x) y_n^2(x) &= -(py_n')' y_n + q y_n^2 \\
\Rightarrow \qquad \lambda_n \int_0^1 r(x) y_n^2 dx &= -\int_0^1 (py_n')' y_n dx + \int_0^1 q y_n^2 dx \\
\Rightarrow \qquad \lambda_n &= -\, py_n' y_n|_0^1 + \int_0^1 p y_n'^2 dx + \int_0^1 q y_n^2 dx \\
&= h_1 p(1) y_n'^2(1) + h_0 p(0) y_n'^2(0) + \int_0^1 (p y_n'^2 + q y_n^2) dx,
\end{aligned}$$

from which it follows immediately that $\lambda_n \geq 0$. The only possibility to achieve $\lambda_n = 0$ is to have $y_n' = 0$. Then $y_n(0) = y_n(1) = -h_1 y_n'(1) = 0$, thus this results only in the trivial solution. Thus $\lambda_n > 0$ for all n.

If y_n and $\tilde{y}_n$ are eigenfunctions belonging to the same λ_n, then

$$\begin{aligned}
W(y_n, \tilde{y}_n)(0) &= \det \begin{pmatrix} y_n(0) & \tilde{y}_n(0) \\ y_n'(0) & \tilde{y}_n'(0) \end{pmatrix} \\
&= \det \begin{pmatrix} h_0 y_n'(0) & h_0 \tilde{y}_n'(0) \\ y_n'(0) & \tilde{y}_n'(0) \end{pmatrix} \\
&= 0.
\end{aligned}$$

It follows that the two eigenfunctions are linearly dependent. In other words, the multiplicity of λ_n is one, as we had to establish. ■

2. Example: a nonhomogeneous heat equation

Let us consider an example to illustrate the use of all of the previous material on Sturm-Liouville problems, as applied to a nonhomogeneous heat equation.

Consider the following problem:

$$\rho(x) c(x) \frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(k(x) \frac{\partial u}{\partial x} \right) - l(x) u,$$

with $\rho(x)$ denoting the density, $c(x)$ the specific heat, $u(x)$ the temperature, $k(x)$ the thermal conductivity and $l(x)$ a radiation coefficient.

We have boundary conditions

$$u(0,t) = h_0 u_x(0,t), \quad u(1,t) = -h_1 u_x(1,t).$$

This form for the boundary conditions is an immediate consequence of Newton's law of cooling. Further we have an initial temperature distribution given by

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

In the above the functions $\rho(x)$, $c(x)$ are strictly positive, and $k(x) \geq 0$. Lastly, the constants h_0 and h_1 are positive or zero as well.

Let's try using separation of variables on this problem. Set

$$u(x,t) = X(x)T(t),$$

then

$$\begin{aligned} & \rho(x)c(x)XT' = (k(x)X')'T - l(x)XT \\ \Rightarrow \qquad & \frac{T'}{T} = \frac{(k(x)X' - l(x)X)'}{\rho(x)c(x)X} = -\lambda, \end{aligned}$$

where $-\lambda$ is our separation constant.

It follows that

$$T(t) = e^{-\lambda t},$$

upto a multiplicative constant.

Turning to the spatial problem, we need that

$$X(0) = h_0 X'(0), \quad X(1) = -h_1 X'(1).$$

These are the boundary conditions for the Sturm-Liouville problem

$$-(k(x)X')' + l(x)X = \lambda r(x)X,$$

where $r(x) = \rho(x)c(x)$.

We know from our various theorems that under the conditions satisfied by the coefficient functions there are an infinite number of eigenvalues for this problem, and they are all positive. Further, they give rise to a complete set of mutually orthonormal eigenfunctions. Thus

$$u(x,t) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n t} X_n(x).$$

It remains to determine the expansion coefficients a_n using the initial condition:

$$f(x) = \sum_{n=1}^{\infty} a_n X_n(x),$$

so that

$$a_n = \int_0^1 f(x) X_n(x) r(x) dx.$$

This concludes the solution of the problem of heat conduction in a nonhomogeneous rod, at least at a formal level. Going beyond this formal level, we know that the series for $u(x,t)$ is absolutely and uniformly convergent (this follows from the Hilbert-Schmidt theorem), but some more work is required to establish this for its derivatives. The positivity of the eigenvalues makes this an easy exercise.

7.7 Exercises

1. Consider the nonhomogeneous boundary value problem ($a > 0$)

$$\begin{cases} -u'' + a^2 u = f(x), \\ u(0) = 0, u(1) = 0. \end{cases}$$

   a) For which values of a does this problem have a unique solution?

   b) For these values of a, construct the Green's function $g(x, s)$ of this problem.

   c) Write down the solution of the nonhomogeneous problem, using your Green's function from part (b).

   d) Use variation of parameters to obtain a different form of your solution.

   e) Verify that your two solutions are identical. Which formulation would you prefer?

2. Define the norm of a linear operator L acting on a function space $\mathcal{L} = \{u\}$by

$$||L|| = \sup_{u \neq 0} \frac{||Lu||}{||u||} = \sup_{||u||=1} ||Lu||.$$

   Prove that for two linear operators A and B, $||AB|| \leq ||A||||B||$.

Chapter 8

Laplace's equation and potential theory

Potential theory deals with the solutions of the two partial differential equations

$$\Delta u = 0,$$

and

$$\Delta u = f(x).$$

The first equation is called Laplace's equation, the second one Poisson's equation. Solution to Laplace's equation are called **harmonic functions**. We will consider these two equations in two or three dimensions, in an open domain D, whose boundary is denoted by ∂D.

The **Dirichlet problem** for either of these equations imposes given values of u on ∂D:

$$\begin{cases} \Delta u = f(x) & \text{for } x \in D, \\ u = h(x) & \text{for } x \in \partial D. \end{cases}$$

On the other hand, the **Neumann problem** imposes conditions on the normal derivative of u on the boundary:

$$\begin{cases} \Delta u = f(x) & \text{for } x \in D, \\ \dfrac{\partial u}{\partial n} = g(x) & \text{for } x \in \partial D. \end{cases}$$

Here n denotes the notmal vector to ∂D. Lastly, the **Robin problem** mixes these two types of boundary conditions:

$$\begin{cases} \Delta u = f(x) & \text{for } x \in D, \\ \alpha(x)u + \beta(x)\dfrac{\partial u}{\partial n} = k(x) & \text{for } x \in \partial D. \end{cases}$$

8.1 Laplace's equation on separable domains

The domain D for our equations can be arbitrarily complicated since we are no longer in one dimension. The separation of variables method works essentially in rectangular domains, or

domains that become rectangular after a transformation to a different coordinate system. Such domains include spherical domains, cylindrical domains, or any other domains where one of the families of orthogonal coordinates transforms our boundaries to line segments parallel to the new coordinate axes. We refer to such domains as separable.

As an example, we consider Laplace's equation in two dimensions, defined on the inside of a circle of radius a:

$$\begin{cases} \Delta u = 0 & \text{for } x \in \{|x| < a\}, \\ u(a,\theta) = f(\theta) = f(\theta + 2\pi) & \text{for } |x| = a. \end{cases}$$

Here θ is the angular variable along the circle. The requirement that

$$f(\theta) = f(\theta + 2\pi)$$

is imposed to ensure continuity of the boundary condition.

We start by expressing Laplaces equation in polar coordinates:

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0,$$

valid for $0 \le r < a$.

1. Separation of variables

Let's look for solutions of the form

$$u(r,\theta) = R(r)\Theta(\theta).$$

Substitution in Laplace's equation gives

$$\begin{aligned} & R''\Theta + \frac{1}{r}R'\Theta + \frac{1}{r^2}R\Theta'' = 0 \\ \Rightarrow \quad & \frac{R''}{R} + \frac{R'}{rR} + \frac{1}{r^2}\frac{\Theta''}{\Theta} = 0 \\ \Rightarrow \quad & r^2\frac{R''}{R} + r\frac{R'}{R} + \frac{\Theta''}{\Theta} = 0 \\ \Rightarrow \quad & r^2\frac{R''}{R} + r\frac{R'}{R} = -\frac{\Theta''}{\Theta} = c, \end{aligned}$$

where c is a separation constant.

Let's consider the equation for $\Theta(\theta)$ first. We have

$$\Theta'' + c\Theta = 0.$$

Since we want $\Theta(\theta)$ to be a periodic function of the angle variable θ, we need to impose

$$c \ge 0.$$

This results in

$$\Theta = c_1 \cos(\sqrt{c}\theta) + c_2 \sin(\sqrt{c}\theta).$$

Imposing not only that this function is periodic, but that its period is 2π requires

$$c = n^2, \quad \text{for} \quad n = 0, 1, 2, \ldots.$$

Thus

$$\Theta_n = c_1 \cos n\theta + c_2 \sin n\theta.$$

Next we consider the problem for $R(r)$. We have

$$r^2 R'' + rR' - n^2 R = 0.$$

This is an Euler equation with indicial equation

$$\begin{aligned} & s(s-1) + s - n^2 = 0 \\ \Rightarrow \quad & s^2 = n^2 \\ \Rightarrow \quad & s = \pm n. \end{aligned}$$

We need to split this up into two cases:

1. $n = 0$**:** in this case $s = 0$ and
$$R_0 = \alpha_0 + \beta_0 \ln r.$$
This solution has a singularity at $r = 0$, thus we need to choose $\beta_0 = 0$. Thus
$$R_0 = \alpha_0,$$
a constant.

2. $n > 0$**:** now $s = \pm n$, so that
$$R_n = \alpha_n r^n + \beta_n r^{-n}.$$
The second term is singular as $r \to 0$, so we impose $\beta_n = 0$. Thus
$$R_n = \alpha_n r^n.$$

Superimposing all the solutions we've just found results in

$$u(r, \theta) = \alpha_0 + \sum_{n=1}^{\infty} r^n \left(\alpha_n \cos n\theta + \beta_n \sin n\theta\right).$$

Notice that we've redefined the constants for convenience.

Lastly, we need to determine the constants α_n $(n = 0, 1, \ldots)$ and β_n (n=1,2,...) by imposing the boundary condition at $r = a$. We obtain

$$f(\theta) = \alpha_0 + \sum_{n=1}^{\infty} a^n \left(\alpha_n \cos n\theta + \beta_n \sin n\theta\right).$$

From these, it follows that

$$\alpha_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)d\theta,$$
$$\alpha_n = \frac{1}{\pi a^n}\int_0^{2\pi} f(\theta)\cos n\theta\, d\theta,$$
$$\beta_n = \frac{1}{\pi a^n}\int_0^{2\pi} f(\theta)\sin n\theta\, d\theta.$$

This completely determines the solution to our problem defined by Laplace's equation on the disk, with Dirichlet boundary conditions.

2. The Green's function

Using the above solution, we can proceed to find the Green's function for our problem. We have

$$\begin{aligned} u &= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)d\theta + \frac{1}{\pi}\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n \int_0^{2\pi} f(s)(\cos ns\cos n\theta + \sin ns\sin n\theta)ds \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\theta)d\theta + \frac{1}{\pi}\int_0^{2\pi} f(s)ds\sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n \cos n(s-\theta) \\ &= \frac{1}{\pi}\int_0^{2\pi} f(s)ds\left(\frac{1}{2} + \sum_{n=1}^{\infty}\left(\frac{r}{a}\right)^n \cos n(s-\theta)\right). \end{aligned}$$

We'll come back to this result. As an aside, we now compute for $|v| < 1$ and $v, w \in \mathbb{R}$

$$\begin{aligned} \frac{1}{2} + \sum_{n=1}^{\infty} v^n \cos nw &= \Re\left(\frac{1}{2} + \sum_{n=1}^{\infty}\left(ve^{iw}\right)^n\right) \\ &= \Re\left(\frac{1}{2} + ve^{iw}\frac{1}{1-ve^{iw}}\right) \\ &= \Re\left(\frac{1}{2} + ve^{iw}\frac{1}{1-ve^{iw}}\frac{1-ve^{-iw}}{1-ve^{-iw}}\right) \\ &= \Re\left(\frac{1}{2} + \frac{ve^{iw}\left(1-ve^{-iw}\right)}{1+v^2-2v\cos w}\right) \\ &= \Re\left(\frac{1+v^2-2v\cos w + 2ve^{iw} - 2v^2}{2(1+v^2-2v\cos w)}\right) \\ &= \frac{1-v^2}{2(1+v^2-2v\cos w)} \end{aligned}$$

Using this result we obtain for the solution of our problem

$$u(r,\theta) = \frac{1}{2\pi}\int_0^{2\pi}\frac{a^2-r^2}{a^2+r^2-2ra\cos(s-\theta)}f(s)ds.$$

From this formula, it is easy to read off the Green's function for this problem. This form of the solution is known as Poisson's formula. It's an extremely important result, as it gives very accurate answers even if one uses the crudest numerical integration techniques to compute the integral.

8.2 The Hankel transform

In this section we introduce an integral transform other than the Fourier transform. We'll see that in essence it is a Fourier transform suitable for domains with cylindrical symmetry.

The **Bessel function of the first kind of order 0** is defined as

$$J_0(z) = \frac{1}{2\pi}\int_0^{2\pi} \cos(z\cos\theta)d\theta.$$

We can use this definition to show that the Bessel function satisfies the following second-order ordinary differential equation:

$$\begin{aligned}
&zJ_0''(z) + J_0'(z) + zJ_0(z) \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left(-z\cos(z\cos\theta)\cos^2\theta - \sin(z\cos\theta)\cos\theta + z\cos(z\cos\theta)\right) d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left(z\cos(z\cos\theta)(1-\cos^2\theta) - \sin(z\cos\theta)\cos\theta\right) d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi} \left(z\cos(z\cos\theta)\sin^2\theta - \sin(z\cos\theta)\cos\theta\right) d\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi} \frac{\partial}{\partial\theta}\left(-\sin(z\cos\theta)\sin\theta\right) d\theta \\
&= 0.
\end{aligned}$$

Thus

$$zJ_0''(z) + J_0'(z) + zJ_0(z) = 0.$$

The **Hankel transform** of $f(x)$ is

$$\hat{f}(\rho) = \int_0^\infty f(r)J_0(\rho r)rdr.$$

Theorem 40 (Inverse Hankel transform) *The inverse of the Hankel transform is given by*

$$f(r) = \int_0^\infty \hat{f}(\rho)J_0(\rho r)\rho d\rho.$$

Thus, the functional recipes for the Hankel transform and its inverse are identical.

Proof. Let $F(x,y) = f(r)$ be a radially symmetric function. Its two-dimensional Fourier transform is

$$\hat{F}(k_1,k_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)e^{-i(xk_1+yk_2)}dxdy.$$

We introduce polar coordinates in both (x,y)-space and in (k_1,k_2)-space:

$$\begin{cases} x &= r\cos\theta \\ y &= r\sin\theta \end{cases}, \qquad \begin{cases} k_1 &= \rho\cos\alpha \\ k_2 &= \rho\sin\alpha \end{cases}.$$

Equating $\hat{F}(k_1,k_2) = \hat{f}(\rho,\alpha)$, we get

$$\begin{aligned}
\hat{f}(\rho,\alpha) &= \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi} f(r)e^{-ir\rho(\cos\theta\cos\alpha+\sin\theta\sin\alpha)}rdrd\theta \\
&= \frac{1}{2\pi}\int_0^{\infty} f(r)rdr\int_0^{2\pi} e^{-ir\rho\cos(\theta-\alpha)}d\theta \\
&= \frac{1}{2\pi}\int_0^{\infty} f(r)rdr\int_0^{2\pi} e^{-ir\rho\cos\theta}d\theta \\
&= \frac{1}{2\pi}\int_0^{\infty} f(r)rdr\int_0^{2\pi} \cos(r\rho\cos\theta)d\theta \\
&= \int_0^{\infty} f(r)J_0(r\rho)rdr.
\end{aligned}$$

We see that $\hat{f}(\rho,\alpha)$ depends only on ρ. Note that $\int_0^{2\pi}\sin(r\rho\cos\theta)rdr = 0$. This is easily seen by replacing θ by $\theta+\pi$. Since we're integrating over an entire period of a periodic function, this shift of the independent variable has no effect on the overall result.

Given that we're using two-dimensional Fourier transforms, we know what the inverse transform is:

$$\begin{aligned}
f(r) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{f}(\rho)e^{i(xk_1+yk_2)}dk_1dk_2 \\
&= \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi}\hat{f}(\rho)e^{i\rho r\cos(\theta-\alpha)}\rho d\rho d\alpha \\
&= \int_0^{\infty}\hat{f}(\rho)J_0(\rho r)\rho d\rho.
\end{aligned}$$

This is what we had to prove. ■

Next we use the Hankel transform to solve the following Dirichlet problem for Laplace's equation.

$$\begin{cases} \Delta u = 0 & \text{for } z>0, (x,y)\in\mathbb{R}^2, \\ u(x,y,0) = f(r) = f(\sqrt{x^2+y^2}), & \text{for } (x,y)\in\mathbb{R}^2. \end{cases}$$

Because of the boundary condition, we are prompted to look for a radially symmetric solution: a function that depends on r and z, but not on x or y individually:

$$u = u(r,z).$$

In cylindrical coordinates without angular dependence, our Dirichlet problem becomes

$$\begin{cases} \dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\dfrac{\partial u}{\partial r}\right)+\dfrac{\partial^2 u}{\partial z^2}=0 & \text{for } z>0, r\geq 0, \\ \\ u(r,0)=f(r), & \text{for } (x,y)\in\mathbb{R}^2. \end{cases}$$

We multiply this partial differential equation by $rJ_0(\rho r)$ and integrate over all r values. In what follows, Hankel transforms are denoted by a hat.

$$\begin{aligned}
& \int_0^\infty \left(J_0(\rho r)\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+rJ_0(\rho r)\frac{\partial^2 u}{\partial z^2}\right)dr=0 \\
\Rightarrow \quad & \frac{\partial^2 \hat{u}}{\partial z^2}+\int_0^\infty J_0(\rho r)\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)dr=0 \\
\Rightarrow \quad & \frac{\partial^2 \hat{u}}{\partial z^2}+\int_0^\infty \left(\frac{\partial}{\partial r}\left(J_0(\rho r)r\frac{\partial u}{\partial r}\right)-\rho r\frac{\partial u}{\partial r}J_0'(\rho r)\right)dr=0 \\
\Rightarrow \quad & \frac{\partial^2 \hat{u}}{\partial z^2}+J_0(\rho r)r\frac{\partial u}{\partial r}\bigg|_{r=0}^{r\to\infty}-\int_0^\infty \rho r\frac{\partial u}{\partial r}J_0'(\rho r)dr=0 \\
\Rightarrow \quad & \frac{\partial^2 \hat{u}}{\partial z^2}-\int_0^\infty \rho r\frac{\partial u}{\partial r}J_0'(\rho r)dr=0.
\end{aligned}$$

The integrated term vanishes because we want our solution and its derivatives to decay as $r\to\infty$. Thus

$$\begin{aligned}
\frac{\partial^2 \hat{u}}{\partial z^2} &= \int_0^\infty \rho r\frac{\partial u}{\partial r}J_0'(\rho r)dr \\
&= \int_0^\infty \frac{\partial u}{\partial r}\rho r J_0'(\rho r)dr \\
&= -\int_0^\infty u\frac{\partial}{\partial r}\left(\rho r J_0'(\rho r)\right)dr \\
&= -\int_0^\infty u\left(\rho J_0'(\rho r)+\rho^2 r J_0''(\rho r)\right)dr \\
&= -\int_0^\infty \frac{u}{r}\left(\rho r J_0'(\rho r)+\rho^2 r^2 J_0''(\rho r)\right)dr \\
&= \int_0^\infty \frac{u}{r}\rho^2 r^2 J_0(\rho r)dr \\
&= \rho^2\int_0^\infty uJ_0(\rho r)rdr \\
&= \rho^2\hat{u}.
\end{aligned}$$

This differential equation is easily solved:

$$\hat{u}=A(\rho)e^{\rho z}+B(\rho)e^{-\rho z}.$$

It follows immediately that we need to eliminate the first term, as it $\to \infty$ as $z \to \infty$. Thus

$$\hat{u} = B(\rho)e^{-\rho z}.$$

It follows that

$$\begin{aligned} u(r,z) &= \int_0^\infty \hat{u} J_0(r\rho)\rho d\rho \\ &= \int_0^\infty B(\rho)e^{-\rho z} J_0(r\rho)\rho d\rho. \end{aligned}$$

At this point we impose the boundary condition at $z = 0$.

$$\begin{aligned} & f(r) = \int_0^\infty B(\rho) J_0(r\rho)\rho d\rho \\ \Rightarrow \qquad & B(\rho) = \hat{f} = \int_0^\infty f(r) J_0(r\rho) r dr. \end{aligned}$$

This concludes the solution of our Dirichlet problem on the half-space. We could put these results together to obtain a Green's function, for instance. Provided with a radially symmetric boundary condition at $z = 0$, the above formula determines $B(\rho)$, which in turn determines the final solution by

$$u(r,z) = \int_0^\infty B(\rho)e^{-\rho z} J_0(r\rho)\rho d\rho.$$

8.3 The maximum principle

We have studied the maximum principle for the heat equation in reasonable detail. Such principles can be used for the equations of potential theory with equal success.

Lemma 7 *Let $u \in C^2(D) \cap C^1(\bar{D})$, where D is connected and bounded*[1]*. If $-\Delta u \geq 0$ in D and $u \geq 0$ on ∂D, then $u \geq 0$ on D. Similarly, if $-\Delta u \leq 0$ in D and $u \leq 0$ on ∂D, then $u \leq 0$ on D.*

Proof. We'll prove the first statement. The second statement follows from it by replacing u by $-u$. We proceed by contradiction. Suppose that there exists an $x_0 \in D$ such that

$$u(x_0) < 0.$$

Then u achieves a negative minimum value somewhere in D, say at $x = c$:

$$u(c) < 0.$$

[1]This means that u is twice continuously differentiable on D, and once on $\bar{D} = D \cup \partial D$.

For $\epsilon > 0$ construct

$$v(x) = u(x) - \epsilon|x - c|^2.$$

Note that $v(c) = u(c)$. We choose ϵ so that for all $x \in \partial D$

$$v(x) = u(x) - \epsilon|x - c|^2 \geq -\epsilon|x - c|^2 \geq -\epsilon d^2 > u(c) = v(c).$$

Here d is the diameter of D:

$$d = \max_{x_1, x_2 \in D} |x_1 - x_2|.$$

It follows that $v(x)$ achieves its minimum at some point interior to D, say at $p \in D$. Then

$$\Delta v(p) \geq 0.$$

Let $\mathcal{D}$ denote the dimension of D. Then

$$\Delta u(p) = \Delta v(p) + 2\epsilon\mathcal{D} > 0,$$

which is a contradiction. Thus $u(x) \geq 0$, for all $x \in D$. This is what we had to show. ■

With this lemma at hand, it is easy to prove the Maximum-Minimum Principle.

Theorem 41 (Maximum-Minimum Principle) *Let D be connected and bounded, and let $u \in C^2(D) \cap C^1(\bar{D})$. Then*
(i) if $-\Delta u \leq 0$ in D, then $u(x) \leq \max_{y \in \partial D} u(y)$, for all $x \in D$,
(ii) if $-\Delta u \geq 0$ in D, then $u(x) \geq \min_{y \in \partial D} u(y)$, for all $x \in D$,
(iii) if $-\Delta u = 0$ in D, then $\min_{y \in \partial D} u(y) \leq u(x) \leq \max_{y \in \partial D} u(y)$, for all $x \in D$.

Proof (i) Let $M = \max_{y \in \partial D} u(y)$. Then

$$-\Delta(u - M) \leq 0$$

for $x \in D$, and

$$u - M \leq 0$$

for $x \in \partial D$. By the previous lemma

$$u - M \leq 0$$

on D. This proves (i). Items (ii) and (iii) follow similarly. ■

We can put this result to good use to prove uniqueness for the Dirichlet problem.

Theorem 42 *Let D be connected and bounded. Then there is at most one solution to*

$$\begin{cases} \Delta u = h(x) & \text{for } x \in D \\ u = f(x) & \text{for } x \in \partial D \end{cases},$$

with $u \in C^2(D) \cap C^1(\bar{D})$.

Proof. Suppose u_1 and u_2 both satisfy the Dirichlet problem. Then $u = u_1 - u_2$ satisfies

$$\begin{cases} \Delta u = 0 & \text{for } x \in D \\ u = 0 & \text{for } x \in \partial D \end{cases}.$$

By the Maximum-Minimum Principle we find that

$$0 \leq u \leq 0 \quad \Rightarrow \quad u = 0 \text{ for } x \in D.$$

This proves the assertion. ■

Just like for the heat equation, we can use the Maximum-Minimum Principle to find a-priori estimates for solutions of Poisson's equation.

Theorem 43 *Let D be connected and bounded, and $u \in C^2(D) \cap C^1(\bar{D})$. Suppose u satisfies*

$$\begin{cases} \Delta u = h(x) & \text{for } x \in D \\ u = f(x) & \text{for } x \in \partial D \end{cases}.$$

Finally, let $|h(x)| \leq H$ in D and $|f(x)| < F$ on ∂D. Then

$$|u(x)| < F + CH,$$

for $x \in D$. Here C is a constant that depends only on D.

Proof. Let $x_1 = c_1$ be a plane entirely to the left of D, as indicated in Fig. 8.1. Then

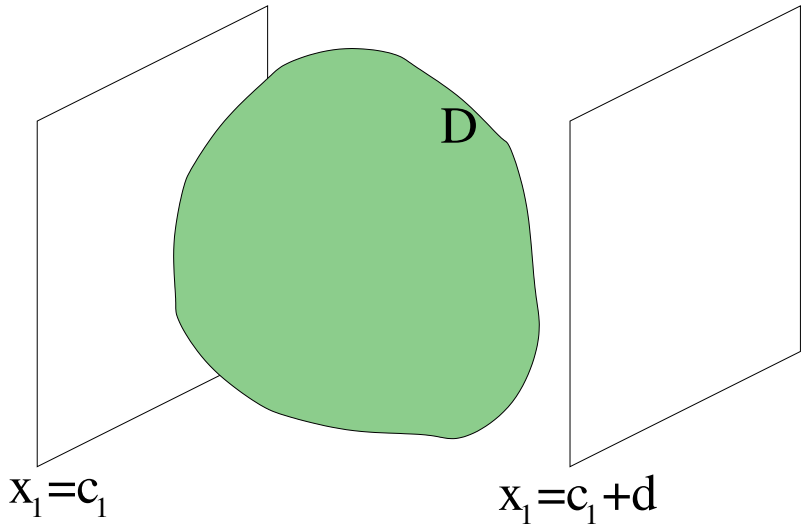

Figure 8.1: The choice of the coordinate planes $x_1 = c_1$ and $x_1 = c_1 + d$.

$c_1 < x_1$ for $x \in \bar{D}$. Fix d such that $x_1 = c_1 + d$ is a coordinate plane entirely to the right of D. Let

$$w = F + H(e^d - e^{x_1 - c_1}), \quad \text{for } x \in \bar{D}.$$

Then

$$\Delta w = -He^{x_1 - c_1} \leq -H, \quad \text{for } x \in D,$$

which implies

$$-\Delta w \geq 0, \quad \text{for } x \in D,$$

and

$$w \geq F, \quad \text{for } x \in \partial D.$$

Next we consider

$$W_{\pm} = w \pm u.$$

Then

$$\begin{aligned} -\Delta W_{\pm} &= -\Delta w \mp \Delta u \\ &\geq H \mp h \\ &\geq 0, \quad \text{for } x \in D. \end{aligned}$$

On ∂D we have

$$W_{\pm} = w \pm u \geq F \pm u \geq 0.$$

Now we can use the Maximum-Minimum Principle to conclude that

$$W_{\pm} \geq 0 \ \Rightarrow \ w \pm u \geq 0, \quad \text{for } x \in \bar{D}.$$

It follows from this that

$$|u| \leq w(x) \leq F + CH.$$

This last inequality follows from the definition of w. For the constant C, we could use $C = e^d$. This concludes the proof. ■

8.4 Solving Poisson's equation

So far, we've mainly talked about solving Laplace's equation. We'll continue this for a bit longer. It turns out that the current discussion of singular solutions of Laplace's equation will lead to solutions of Poisson's equation.

1. The fundamental solution of Laplace's equation

Let's consider Laplace's equation in n dimensions:

$$\sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2} = 0.$$

We will look for special solutions to this equation that have spherical symmetry: solutions that depend only on the radial variable, and that have no angular dependence. Thus

$$u(x) = v(r),$$

where

$$r^2 = \sum_{j=1}^{n} x_j^2.$$

We immediately get

$$2r\frac{\partial r}{\partial x_j} = 2x_j \quad \Rightarrow \quad \frac{\partial r}{\partial x_j} = \frac{x_j}{r},$$

for $j = 1, \ldots, n$. Using this and the chain rule we obtain

$$\begin{aligned} & \frac{\partial u}{\partial x_j} = v'(r)\frac{x_j}{r} \\ \Rightarrow \qquad & \frac{\partial^2 u}{\partial x_j^2} = v''(r)\frac{x_j^2}{r^2} + \frac{v'(r)}{r} - v'(r)\frac{x_j^2}{r^3}. \end{aligned}$$

This allows us to rewrite Laplace's equation for spherically symmetric functions as

$$\begin{aligned} & v'' + n\frac{v'}{r} - \frac{v'}{r} = 0 \\ \Rightarrow \qquad & v'' + \frac{n-1}{r}v' = 0 \\ \Rightarrow \qquad & \frac{v''}{v'} = \frac{1-n}{r} \\ \Rightarrow \qquad & (\ln v')' = \frac{1-n}{r} \\ \Rightarrow \qquad & \ln\frac{v'}{c} = (1-n)\ln r \\ \Rightarrow \qquad & v' = cr^{1-n} = \frac{c}{r^{n-1}}. \end{aligned}$$

Here c is an arbitrary real constant.

1. If $n = 2$ then we get

$$v' = \frac{c}{r} \quad \Rightarrow \quad v = c\ln r + d,$$

   where d is a second arbitrary constant.

2. If $n \geq 3$ then

$$v = \frac{c}{(2-n)r^{n-2}} + d.$$

With an appropriate choice of the constants we call these solutions the fundamental solution $\phi(x)$ of Laplace's equation:

$$\phi(x) = \begin{cases} -\dfrac{1}{2\pi}\ln|x| & \text{for } n = 2 \\ \dfrac{1}{4\pi|x|} & \text{for } n = 3, \end{cases}$$

and so on for other dimensions. Note that these solutions are singular at $x = 0$. This should make you ponder. Are these solutions useful? Or are they mere mathematical artifacts only to be appreciated by the true aesthete? The answer as always is all of the above. Let's demonstrate this.

Clearly, $\phi(x - y)$ is a solution of Laplace's equation as well, for any fixed y. So is

$$\phi(x - y)f(y),$$

for any function $f(y)$. Could we argue that the superposition

$$u(x) = \int_{\mathbb{R}^n} \phi(x - y)f(y)dy$$

is the general solution? This turns out to be wrong: due to the singularity at the origin, we cannot bring the Laplacian inside the integral sign. The infinite extent of the integration domain does not help either. So, what can be said about this superposition?

2. Poisson's equation on $\mathbb{R}^n$

Theorem 44 *Consider Poisson's equation on $\mathbb{R}^n$: let $u \in C^2(\mathbb{R}^n)$, and*

$$-\Delta u = f(x),$$

for $x \in \mathbb{R}^n$. If $f(x) \in C^2(\mathbb{R}^n)$ this problem is solved by

$$u(x) = \int_{\mathbb{R}^n} \phi(x - y)f(y)dy.$$

Proof. Let $z = x - y$. Then

$$u(x) = \int_{\mathbb{R}^n} \phi(z)f(x - z)dz.$$

It follows that

$$\frac{\partial u}{\partial x_j} = \int_{\mathbb{R}^n} \phi(z)\frac{\partial f}{\partial x_j}(x - z)dz,$$

and

$$\begin{aligned}\Delta u &= \int_{\mathbb{R}^n} \phi(z)\Delta_x f(x - z)dz \\ &= \int_{B(0,\epsilon)} \phi(z)\Delta_x f(x - z)dz + \int_{\mathbb{R}^n - B(0,\epsilon)} \phi(z)\Delta_x f(x - z)dz \\ &= I_\epsilon + J_\epsilon.\end{aligned}$$

Here Δ_x denotes the Laplacian with partial derivatives with respect to the components of x. Also, $B(0, \epsilon)$ denotes the ball of radius ϵ centered at 0.

Let's look at I_ϵ first. This is the part containing the singularity of the fundamental solution, but it's a small part. In contrast J_ϵ does not contain any singularities, but it's huge. We have

$$|I_\epsilon| \leq \max_{z\in\mathbb{R}^n} |\Delta_x f(x-z)| \int_{B(0,\epsilon)} |\phi(z)|dz.$$

We can say more about this last integral:

1. If $n = 2$ then we get

$$\begin{aligned}\int_{B(0,\epsilon)} |\phi(z)|dz &= 2\pi \int_0^\epsilon \frac{1}{2\pi} |\ln r|\, r dr \\ &= \int_0^\epsilon r|\ln r| dr \\ &\leq \epsilon^2 |\ln \epsilon|,\end{aligned}$$

since $r|\ln r|$ is monotone increasing for small ϵ. Thus

$$|I_\epsilon| \leq \max_{z\in\mathbb{R}^2} |\Delta_x f(x-z)| \epsilon^2 |\ln \epsilon| \to 0,$$

as $\epsilon \to 0$.

2. If $n = 3$ then

$$\begin{aligned}\int_{B(0,\epsilon)} |\phi(z)|dz &= 4\pi \int_0^\epsilon \frac{1}{4\pi}\frac{1}{r} r^2 dr \\ &= \int_0^\epsilon r dr \\ &= \frac{1}{2}\epsilon^2.\end{aligned}$$

Thus

$$|I_\epsilon| \leq \max_{z\in\mathbb{R}^3} |\Delta_x f(x-z)| \frac{\epsilon^2}{2} \to 0,$$

as $\epsilon \to 0$.

This the contribution from $I_\epsilon \to 0$ as $\epsilon \to 0$, and the I_ϵ part is innocent.

How about J_ϵ? We have

$$\begin{aligned}J_\epsilon &= \int_{\mathbb{R}^n - B(0,\epsilon)} \phi(z) \Delta_x f(x-z) dz \\ &= \int_{\mathbb{R}^n - B(0,\epsilon)} \phi(z) \Delta_z f(x-z) dz.\end{aligned}$$

We use one of Green's identities:

$$\int_D (v\Delta w + \nabla v \cdot \nabla w) dx = \int_{\partial D} v \frac{\partial w}{\partial n} d\sigma,$$

where σ is a "surface" element on ∂D. Here surface denotes anything that's on the boundary ∂D. It does not necessarily entail a two-dimensional space. Using this identity we get

$$\begin{aligned} J_\epsilon &= \int_{\partial B(0,\epsilon)} \phi(z) \frac{\partial f}{\partial n}(x-z) d\sigma - \int_{\mathbb{R}^n - B(0,\epsilon)} \nabla_z \phi(z) \cdot \nabla_z f(x-z) dz \\ &= K_\epsilon + L_\epsilon. \end{aligned}$$

Again, we've split things in two parts. for the first part, we get

$$\begin{aligned} K_\epsilon &= \int_{\partial B(0,\epsilon)} \phi(z) \frac{\partial f}{\partial n}(x-z) d\sigma \\ \Rightarrow \qquad |K_\epsilon| &\leq \max_{z \in \mathbb{R}^n} |\nabla f| \int_{\partial B(0,\epsilon)} |\phi(z)| d\sigma. \end{aligned}$$

We may proceed to bound this last integral as before:

1. If $n = 2$ then we get

$$\begin{aligned} \int_{\partial B(0,\epsilon)} |\phi(z)| dz &\leq \int_{\partial B(0,\epsilon)} \frac{1}{2\pi} |\ln r| \, d\sigma \\ &= \int_0^{2\pi} \frac{1}{2\pi} |\ln \epsilon| \epsilon d\theta \\ &= \epsilon |\ln \epsilon|. \end{aligned}$$

   Thus

$$|K_\epsilon| \leq \max_{z \in \mathbb{R}^n} |\nabla f| \epsilon |\ln \epsilon| \to 0,$$

   as $\epsilon \to 0$.

2. If $n = 3$ then

$$\begin{aligned} \int_{\partial B(0,\epsilon)} |\phi(z)| dz &= \int_{\partial B(0,\epsilon)} \frac{1}{4\pi r} d\sigma \\ &= \int_0^{2\pi} \int_0^{\pi} \frac{1}{4\pi\epsilon} \epsilon^2 \sin\varphi d\theta d\varphi \\ &= \epsilon. \end{aligned}$$

   Thus

$$|K_\epsilon| \leq \max_{z \in \mathbb{R}^n} |\nabla f| \epsilon \to 0,$$

   as $\epsilon \to 0$.

We see that also K_ϵ does not contribute to the integral, as $\epsilon \to 0$. All our hope lies with L_ϵ. We have

$$L_\epsilon = - \int_{\mathbb{R}^n - B(0,\epsilon)} \nabla_z \phi(z) \cdot \nabla_z f(x-z) dz.$$

We use the same Green's identity again, but with the roles of v and w switched. This gives us

$$\begin{aligned} L_\epsilon &= \int_{\mathbb{R}^n - B(0,\epsilon)} f(x-z)\Delta_z\phi(z)dz - \int_{\partial B(0,\epsilon)} f(x-z)\frac{\partial\phi}{\partial n}(z)d\sigma \\ &= -\int_{\partial B(0,\epsilon)} f(x-z)\frac{\partial\phi}{\partial n}(z)d\sigma, \end{aligned}$$

since $\Delta_z\phi(z) = 0$. Let's compute the normal derivative of ϕ: let r denotes $\sqrt{z_1^2 + \ldots + z_n^2}$. Then

$$\begin{aligned} & & \frac{\partial\phi}{\partial z_j} &= \phi'(r)\frac{z_j}{r} \\ \text{and} & & n_i &= -\frac{z_j}{r} \\ \Rightarrow & & \frac{\partial\phi}{\partial n} &= \sum_{j=1}^{n}\frac{\partial\phi}{\partial z_j}n_j \\ & & &= -\phi'(r) \\ & & &= \begin{cases} \dfrac{1}{2\pi r} & \text{for } \ n=2 \\ \dfrac{1}{4\pi r^2} & \text{for } \ n=3 \end{cases}. \end{aligned}$$

We see that

$$\begin{aligned} L_\epsilon &= \phi'(\epsilon)\int_{\partial B(0,\epsilon)} f(x-z)d\sigma \\ &= -\frac{1}{|\partial B(0,\sigma)|}\int_{\partial B(0,\epsilon)} f(x-z)d\sigma \\ &= -\langle f(x-z)\rangle_{\partial B(0,\epsilon)} \\ &= -\langle f(x)\rangle_{\partial B(0,\epsilon)}. \end{aligned}$$

Here $\langle f(x-z)\rangle$ denotes the average of $f(x-z)$ over the indicated region. Clearly, as $\epsilon \to 0$, we have

$$\lim_{\epsilon\to\epsilon} L_\epsilon = -\lim_{\epsilon\to\epsilon}\langle f(x-z)\rangle_{\partial B(0,\epsilon)} = -f(x).$$

This allows us to finally conclude that $u(x)$ solves

$$-\Delta u = f(x),$$

so that u indeed solves Poisson's equation. This is what we had to prove. ■

If we do a more formal calculation, we get

$$\begin{aligned} -\Delta u &= \int_{\mathbb{R}^n} -\Delta_x \phi(x-y) f(y) dy \\ &= \int_{\mathbb{R}^n} \delta(x-y) f(y) dy \\ &= f(x). \end{aligned}$$

We see that the fundamental solutions of Laplace's equation may be used as Green's functions for Poisson's equation. Thus singular solutions can be of great interest indeed!

8.5 Exercises

1. We have derived Poisson's formula for the solution of Laplace's equation inside of a circle of radius a with Dirichlet boundary conditions given on $r = a$. Show that this formula, when properly interpreted, is also valid for the solution of Laplace's equation outside of a circle of radius a with Dirichlet boundary conditions given on $r = a$. Note: you can't merely use a transformation $r \to 1/r$ for this, as this transformation is singular. You're supposed to repeat what we did in this chapter, but taking into account different boundary conditions, because of the different type of domain.

2. Find a formal solution to the following initial-value problem for the heat equation with two spatial dimensions with diffusion coefficient σ:

$$\begin{cases} u_t - \sigma(u_{xx} + u_{yy}) = f(x,y), \\ u(x,y,0) = 0, \end{cases}$$

   where $f(x,y)$ only depends on the radial variable $\sqrt{x^2+y^2}$ with $(x,y) \in \mathbb{R}^2$ and $t \geq 0$.

3. Find a formal solution to the following initial-value problem for the wave equation with speed c:

$$\begin{cases} u_{tt} = c^2(u_{xx} + u_{yy}) + f(x,y), \\ u(x,y,0) = 0, \quad u_t(x,y,0) = 0, \end{cases}$$

   where $f(x,y)$ only depends on the radial variable $\sqrt{x^2+y^2}$ with $(x,y) \in \mathbb{R}^2$ and $t \geq 0$.

4. Let D be a bounded, connected region in $\mathbb{R}^n$. Suppose that u satisfies the Neumann problem

$$\begin{aligned} \Delta u &= F(x), \qquad x \in D, \\ \frac{\partial u}{\partial n} &= h(x), \qquad x \in \partial D. \end{aligned}$$

   Discuss the existence and uniqueness of any solutions. The Green's identity

$$\int_D \Delta u dx = \int_{\partial D} \frac{\partial u}{\partial n} d\sigma$$

may be useful.