Partial Differential Equations

R.Wait

October 16, 2008

UPPSALA UNIVERSITET
Inst. för informationsteknologi
Avd. för teknisk databehandling

UPPSALA UNIVERSITY
Information Technology
Dept. of Scientific Computing

Contents

Chapter 1

Preliminaries

1.1 Notation

In $\mathbb{R}^n$ assume that $u = u(x_1, \dots, x_n)$ then the partial derivatives w.r.t. x_i (assuming x_j, $j \neq i$ are constant) can be denoted by $\frac{\partial u}{\partial x_i}$, u_{x_i} or $\partial_{x_i} u$ and the differential operator can be denoted by $\frac{\partial}{\partial x_i}$ or ∂_{x_i}. A second derivative can be $\frac{\partial^2 u}{\partial x^2}$, u_{xx} or $\partial_{xx} u$. The operator notation is often extended to the *multi-index* notation, where for $x = (x_1, \dots, x_n)$ and $\alpha = (\alpha_1, \dots, \alpha_n)$ with $x_i \in \mathbb{R}$ and α_i non-negative integers,

$$\partial^\alpha = \frac{\partial^{\alpha_1}}{\partial {x_1}^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial {x_n}^{\alpha_n}}$$

The gradient operator is denoted by the vector $\nabla \equiv \left(\begin{array}{ccc} \dfrac{\partial}{\partial x_1} & \cdots & \frac{\partial}{\partial x_n} \end{array} \right)^T$.

These notes include material taken from on-line lecture notes [9], [3] and from Wikipedia [11]

1.2 Fourier Series

Any *reasonable* function that is periodic with period $2l$, can be written as a infinite linear combination of sin and cos functions

$$f(x) = \frac{1}{2} B_0 + \sum_{n=1}^{\infty} \left(A_n \sin\left(\frac{n\pi x}{l}\right) + B_n \cos\left(\frac{n\pi x}{l}\right)\right) \tag{1.1}$$

Given the identities

$$\frac{1}{l} \int_{-l}^{l} \sin\left(\frac{n\pi x}{l}\right) \sin\left(\frac{m\pi x}{l}\right) = \delta_{mn}$$

$$\frac{1}{l} \int_{-l}^{l} \cos\left(\frac{n\pi x}{l}\right) \cos\left(\frac{m\pi x}{l}\right) = \delta_{mn}$$

$$\frac{1}{l} \int_{-l}^{l} \sin\left(\frac{n\pi x}{l}\right) \cos\left(\frac{m\pi x}{l}\right) = 0$$

It follows that

$$A_n = \frac{1}{l}\int_{-l}^{l} f(x)\cos\left(\frac{n\pi x}{l}\right)\,dx.$$

and

$$B_n = \frac{1}{l}\int_{-l}^{l} f(x)\sin\left(\frac{n\pi x}{l}\right)\,dx.$$

Assume the $f(-x) = -f(x)$ (f is an odd function) then $B_n = 0$ and (1.1) becomes

$$f(x) = \sum_{n=1}^{\infty} A_n \sin\left(\frac{n\pi x}{l}\right) \tag{1.2}$$

with

$$A_n = \frac{2}{l}\int_{0}^{l} f(x)\cos\frac{n\pi x}{l}\,dx.$$

if f is even then $A_n = 0$ and

$$f(x) = \frac{1}{2}B_0 + \sum_{n=1}^{\infty} B_n \cos\left(\frac{n\pi x}{l}\right) \tag{1.3}$$

and

$$B_n = \frac{2}{l}\int_{0}^{l} f(x)\sin\frac{n\pi x}{l}\,dx.$$

Theorem 1 *Parseval's*[1] *Theorem: If*

$$A(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx} \text{ and } B(x) = \sum_{n=-\infty}^{\infty} b_n e^{inx}$$

respectively. Then

$$\sum_{n=-\infty}^{\infty} a_n\overline{b_n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(x)\overline{B(x)}dx$$

1.3 Fourier Transforms

The Fourier Transform of function ϕ is

$$\mathcal{F}\phi(\xi) = \hat{\phi}(\xi) = \int_{\mathbb{R}^n} e^{-ix\cdot\xi}\phi(x)dx$$

and the inverse transformation is

$$\mathcal{F}^{-1}\psi(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{ix\cdot\xi}\psi(\xi)d\xi$$

These are also known as *continuous Fourier transforms* to distinguish them from the *discrete Fourier transforms* mentioned later.

[1] Marc-Antoine Parseval des Chênes (April 27, 1755 – August 16, 1836) was a French mathematician

1.3.1 Table of important Fourier transforms

$a \cdot g(t) + b \cdot h(t)$	$a \cdot G(\xi) + b \cdot H(\xi)$
$g(t-a)$	$e^{-ia\xi}G(\xi)$
$e^{iat}g(t)$	$G(\xi - a)$
$g(at)$	$\frac{1}{\lvert a\rvert}G\left(\frac{\xi}{a}\right)$
$G(t)$	$g(-\xi)$
$\frac{d^n g(t)}{dt^n}$	$(i\xi)^n G(\xi)$
$t^n g(t)$	$i^n \frac{d^n G(\xi)}{d\xi^n}$
$(g * h)(t)$	$\sqrt{2\pi}G(\xi)H(\xi)$
$g(t)h(t)$	$\frac{(G*H)(\xi)}{\sqrt{2\pi}}$
$e^{-\alpha t^2}$	$\frac{1}{\sqrt{2\alpha}} \cdot e^{-\frac{\xi^2}{4\alpha}}$
e^{iat^2}	$\frac{1}{\sqrt{2a}} \cdot e^{-i\left(\frac{\xi^2}{4a} - \frac{\pi}{4}\right)}$
$\cos(at^2)$	$\frac{1}{\sqrt{2a}}\cos\left(\frac{\xi^2}{4a} - \frac{\pi}{4}\right)$
$\sin(at^2)$	$\frac{-1}{\sqrt{2a}}\sin\left(\frac{\xi^2}{4a} - \frac{\pi}{4}\right)$
$\mathrm{e}^{-a\lvert t\rvert}$	$\sqrt{\frac{2}{\pi}} \cdot \frac{a}{a^2+\xi^2}$
$\frac{1}{\sqrt{\lvert t\rvert}}$	$\frac{1}{\sqrt{\lvert\xi\rvert}}$

Theorem 2 *Plancherel's*[2] *theorem:*

$$\int_{\mathbb{R}^n} |f(t)|^2 \, dt = \int_{\mathbb{R}^n} |F(\xi)|^2 \, d\xi.$$

Fourier transforms are often used to prove stability, *i.e.* boundedness, of the solution by invoking Plancherel's theorem once the boundedness of the Fourier transform has been established (see sections 3.4 and 9.4.1). Chapters 7 and 8 also require Fourier transforms. These sections and chapters can be omitted if Fourier transforms have not been studied previously.

1.4 The Divergence theorem

The fundamental theorem of calculus states that:

$$\int_a^b f' \; dx \; = \; [f]_a^b$$

with $f = vw$, leads to integration parts

$$\int_a^b (vw)' \; dx \; = \; [vw]_a^b$$

hence

$$\int_a^b vw' \; dx \; = \; -\int_a^b v'w \; dx \; + \; [vw]_a^b$$

[2] Michel Plancherel (16 January 1885 – 4 March 1967) was a Swiss mathematician

if $w = u'$ then

$$\int_a^b vu''\ dx \ = \ -\int_a^b v'u'\ dx \ + \ [vu']_a^b$$

In higher dimensions there is the divergence theorem, which in its simplest form is:

Theorem 3

$$\int_\Omega \nabla \cdot \mathbf{F}\ d\Omega = \int_\Gamma \mathbf{F} \cdot \mathbf{n}\ ds$$

with $\mathbf{F} = v\mathbf{w}$ and $\mathbf{w} = \nabla u$ this leads to

$$\int_\Omega \nabla \cdot (v\nabla u)\ d\Omega = \int_\Gamma v\nabla u \cdot \mathbf{n}\ ds$$

as $\nabla u \cdot \mathbf{n} = \partial_n u$, the *outward normal derivative* this can be written as

$$\int_\Omega v\nabla \cdot \nabla u\ d\Omega = -\int_\Omega \nabla v \cdot \nabla u\ d\Omega + \int_\Gamma v\partial_n u\ ds$$

Chapter 2

Introduction to PDEs

2.1 1st order Equations

Definition 4 *A partial differential equation[7] is a relation of the form*

$$F(x, y, \ldots, u, u_x, u_y, \ldots, u_{xx}, u_{xy}, \ldots) = 0 \tag{2.1}$$

where F is a function of the variables $x, y, \ldots, u, u_x, u_y, \ldots, u_{xx}, u_{xy}, \ldots$; a function $u(x, y, \ldots)$ of the independent variables $x, y, \ldots$ is sought such that (2.1) is identically satisfied in these independent variables if $u(x, y, \ldots)$ and its derivatives are substituted in F.

Ordinary differential equations are typically satisfied by a family of functions that depend on a number of arbitrary parameters, for example the general solution of $u'' - 4u - \sin(x) = 0$ is $u(x) = Ae^{2x} + Be^{-2x} - \frac{1}{5}\sin(x)$ where A and B are arbitrary. The solution of partial differential equations is not so straightforward. Consider first the solution of a 1st order linear homogeneous equation with constant coefficients in $\mathbb{R}^2$.

$$au_x + bu_y = 0 \qquad \equiv \qquad (a, b) \cdot \nabla u = 0 \tag{2.2}$$

That is the directional derivative of u along the direction $\boldsymbol{v} = (a, b)$ is specified. This means that the function $u(x, y)$ on any line of direction (a, b) remains constant (different constant for different lines). The direction $\boldsymbol{v} = (a, b)$ is along a line

$$bx - ay = c \qquad c \in \mathbb{R}$$

these are known as the *characteristic lines* for (2.2). Alternatively the lines define $y(x)$ such that $\frac{d}{dx}y = \frac{b}{a}$ or $\frac{dx}{a} = \frac{dy}{b}$. If u does not change along these lines then

$$u(x, y)|_{bx-ay=c} = f(c)$$

so

$$u(x, y) = f(bx - ay)$$

to be more precise about the form of the solution, it is necessary to specify additional conditions (see below).

Example 5 *Consider the case of variable coefficients*

$$yu_x - xu_y = 0 \tag{2.3}$$

this means that u is constant along curves *that have tangent vectors (y, x) so if as for the line this curve is represented by $y = y(x)$ then*

$$\frac{d}{dx}y = -\frac{x}{y} \qquad \Leftrightarrow \qquad \frac{d}{dx}\left(\frac{1}{2}y^2\right) = -x \qquad \Leftrightarrow \qquad \frac{1}{2}y^2 = -\frac{1}{2}x^2 + c$$

hence the characteristic curves are $y^2 + x^2 = c$ and so

$$u(x, y) = f(y^2 + x^2)$$

In order to define a well posed problem*, that is a problem for which there exists a unique solution, it is necessary to impose supplementary conditions. For example, if the solution of (2.3) also satisfies $u(0, y) = e^{-y^2}$ then*

$$f(y^2) = e^{-y^2} \qquad \Rightarrow \qquad f(t) = e^{-t} \quad \forall t$$

so

$$u(x, y) = e^{-y^2 - x^2}.$$

In general, the equation

$$a(x, y)u_x + b(x, y)u_y = 0$$

can be solved as long as

$$\frac{dy}{dx} = \frac{b(x, y)}{a(x, y)}$$

can be solved as an ODE to give the characteristic curves.

To solve $au_x + bu_y = c$ first find a particular solution. For simplicity, look for a solution of the form $u_0(x, y) = \alpha x$

$$a\alpha + b.0 = c \qquad \Rightarrow \qquad \alpha = \frac{c}{a} \qquad \Rightarrow \qquad u_0(x, y) = \frac{c}{a}x$$

so the general solution is

$$u(x, y) = f(bx - ay) + \frac{c}{a}x$$

To solve $au_x + bu_y + cu = 0$ find the set of solutions, assuming $u \neq 0$,

$$a\frac{u_x}{u} + b\frac{u_y}{u} + c = 0$$

Now define $v = \log u$ then $v_x = \frac{u_x}{u}$ and $v_y = \frac{u_y}{u}$ so

$$av_x + bv_y + c = 0$$

from the previous example

$$v(x, y) = f(bx - ay) - \frac{c}{a}x$$

now $u = e^v$ so

$$\begin{aligned} u(x,y) &= e^{f(bx-ay)-\frac{c}{a}x} \\ &= e^{f(bx-ay)}\frac{1}{e^{\frac{c}{a}x}} \end{aligned}$$

The solution of the general linear equation

$$a(x,y)u_x + b(x,y)u_y = c(x,y)$$

is not constant along the characteristics which satisfy $\frac{dx}{a(x,y)} = \frac{dy}{b(x,y)}$. If s is the tangential variable then $ds = \dfrac{dx}{a(x,y)}$, $ds = \dfrac{dy}{b(x,y)}$ and so as $a(x,y)u_x + b(x,y)u_y$ is the tangential derivative $ds = \dfrac{du}{c(x,y)}$ and hence

$$\frac{dx}{a(x,y)} = \frac{dy}{b(x,y)} = \frac{du}{c(x,y)}$$

assuming $a, b, c \neq 0$, in this formulation a, b and c could be also functions of u.

Example 6

$$u_t + xu_x = 0$$

with $u(x,0) = f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \textit{otherwise} \end{cases}$

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

and the characteristics are curves $xe^{-t} =$ *constant.*

$$\tilde{u}(x,0) = f(x) \quad \Rightarrow \quad u(x,t) = f(xe^{-t}) = \begin{cases} 1 & 0 \leq x \leq e^t \\ 0 & \textit{otherwise} \end{cases}$$

Exercises

Exercise 2.1

- *What are the characteristic curves for the equation $xyu_x + u_y = 0$, assuming $x, y > 0$?*
- *What is the general solution?*
- *What is the general solution of the equation $xyu_x + u_y = 1$*
- *Is*
$$\begin{aligned} xyu_x + u_y = 0, \quad y > 0, -\infty > x > \infty & \\ u(x,0) = f(x), \qquad & x > 0 \\ u(0,y) = g(y), \qquad & y > 0 \end{aligned}$$
a well posed problem for any functions f and g?
- *Is*
$$\left.\begin{aligned} xyu_x + u_y = 0, \quad y > 0 \\ u(x,0) = f(x) \end{aligned}\right\} - \infty > x > \infty$$
a well posed problem for any function f?

Exercise 2.2 *A first order system of p.d.e.s such as*

$$\begin{aligned} u_t + cv_x &= 0 \\ v_t + cu_x &= 0 \end{aligned} \tag{2.4}$$

Writing $\boldsymbol{w} = \begin{pmatrix} u \\ v \end{pmatrix}$, the system (2.4) can be written in terms of matrices as

$$\boldsymbol{w}_t + A\boldsymbol{w}_x = 0$$

Assume that λ is an eigenvalue of A and $\mathbf{u}$ is the corresponding eigenvector. Show that for any differentiable function f,

$$\boldsymbol{w} = f(x - \lambda t)\boldsymbol{u}$$

is a solution. What are the characteristics for (2.4)?

Exercise 2.3 *Given first order system of p.d.e.s*

$$\begin{aligned} u_t + 2u_x + v_x &= 0 \\ v_t + u_x + 2v_x &= 0 \end{aligned} \tag{2.5}$$

- *What are the characteristics?*
- *What is the general solution?*
- *What is the solution such that*
$$\left.\begin{aligned} u(x,0) = \phi(x) \\ v(x,0) = \psi(x) \end{aligned}\right\} \quad - \infty < x < \infty$$

Chapter 3

Well posed problems

3.1 Applications of PDEs

Physical problems have in general 3 characteristics which should be reflected in the mathematical equations:

1. Existence - The phenomenon exists
2. Uniqueness - Given the same conditions, the physical process is repeatable at a later time
3. Stability - A small change in the initial conditions should lead to a small change in the output

Example 7 The transport equation*:*

Assume a drop of ink in a pipe full of water moving a velocity c the concentration is denoted by $u(x,t)$. The amount of ink in a fixed length x of pipe at time t is $\int_0^x u(z,t)dz$ at a later time $t+h$ the ink has moved, with velocity c so the same ink is now $\int_{ch}^{x+ch} u(z,t+h)dz$ thus

$$\int_0^x u(z,t)dz = \int_{ch}^{x+ch} u(z,t+h)dz$$

differentiating w.r.t. x

$$u(x,t) = u(x+ch,t+h)$$

differentiating w.r.t. h at $h=0$

$$u_t(x,t) + cu_x(x,t) = 0$$

In general equations written in the form

$$u_t(x,t) + \frac{\partial}{\partial x}F(u(x,t)) = 0$$

are know as conservation laws

Example 8 The wave equation*:*

Consider the motion of a string of length l, fixed at the endpoints. From Newton's law of motion, the force of the tension is balanced by the acceleration, so considering two neighbouring points x_0 and x_1

$$Tu_x|_{x_0}^{x_1} = \int_{x_0}^{x_1} \rho u_{tt} dx$$

differentiate w.r.t x

$$(Tu_x)_x = \rho u_{tt} \quad \rightarrow \quad u_{tt} = c^2 u_{xx}$$

Diffusion (Heat Conduction):

Given a region $\Omega \in \mathbb{R}^3$, the amount of heat at time t is denoted by $H(t)$ then

$$H(t) = \int_{\Omega} c\rho u \, d\Omega$$

where

- *c is the specific heat*
- *ρ is the density*
- *u is the temperature*
- *k is the conductivity*

then

$$\begin{array}{rcl} \frac{dH}{dt} & = & \int_{\Omega} c\rho\partial_t u \, d\Omega \\ & = & \textit{heat flux out of } \Omega \\ & = & \int_{\Gamma} k\partial_n u \, ds \quad (\partial_n u = \boldsymbol{n} \cdot \nabla u \textit{ the outward normal derivative}) \\ & = & \int_{\Omega} \nabla \cdot (k\nabla u) d\Omega \quad \textit{The divergence theorem} \end{array}$$

so

$$c\rho\partial_t u = \nabla \cdot (k\nabla u)$$

Example 9 Laplace's Equation*: Consider small displacements of a membrane, whose energy difference from no displacement is approximated by*

$$I(\varphi) = \frac{1}{2} \int_{\Omega} |\nabla\varphi|^2 \, d\Omega,$$

The functional I is to be minimised among all trial functions φ that assume prescribed values on the boundary Γ of Ω. If u is the minimising function and v is an arbitrary smooth function that vanishes on Γ, then the first variation of $I[u + v]$ must vanish:

$$\frac{d}{d\epsilon} I(u + \epsilon v)\bigg|_{\epsilon=0} = \int_{\Omega} \nabla u \cdot \nabla v \, d\Omega = 0.$$

Provided that u is sufficiently differentiable, it is possible to apply the divergence theorem to obtain

$$\int_\Omega \nabla \cdot (v\nabla u)\, d\Omega = \int_\Omega (\nabla u \cdot \nabla v + v\nabla \cdot \nabla u)\ d\Omega + \int_\Gamma v\partial_n u ds,$$

where s is arc-length along Γ. Since v vanishes on Γ and the first variation vanishes, the result is

$$\int_\Omega v\nabla u \cdot \nabla u\, d\Omega = 0$$

for all smooth functions v that vanish on Γ. It then follows that

$$\nabla^2 u = 0 \ \ \textit{in}\ \Omega,$$

assumption that the minimising function u has two derivatives.

Frequently first order (nonlinear) systems are written in *conservation form*

Example 10

- $$\begin{aligned} \tfrac{\partial i}{\partial x} + C\tfrac{\partial v}{\partial t} &= -Gv \\ \tfrac{\partial v}{\partial x} + L\tfrac{\partial i}{\partial t} &= -Ri \end{aligned}$$

  where $i(x,t)$ is current in amperes, $v(x,t)$ is voltage

R	-	*resistance*	
L	-	*inductance*	*is a hyperbolic system*
C	-	*capacitance*	
G	-	*leakage*	

- *Flow of an ideal gas in one dimension (*Euler's Equations*)*

  $$\frac{\partial}{\partial t}U + \frac{\partial}{\partial x}F(U) = 0$$

  where

  $$U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}, \quad F(U) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E+p)u \end{pmatrix}$$

  and where ρ is the density, u is the velocity, p is the pressure and E is the total energy per unit mass. Given that

  $$E = \frac{1}{2}\rho u^2 + \rho e, \quad p = (\gamma - 1)\rho e$$

  where e is the internal energy per unit mass and γ is the specific heat ($\gamma = 1.4$ for air). This is aslo a hyperbolic system.

Example 11 The linearised shallow water equations

$$\boldsymbol{u}_t = A_1\boldsymbol{u}_x + A_2\boldsymbol{u}_y + \boldsymbol{f}$$

are hyperbolic, with

$$\boldsymbol{u} = \begin{pmatrix} u \\ v \\ \phi \end{pmatrix} \textit{ and } \boldsymbol{f} = \begin{pmatrix} 0 & -f & 0 \\ f & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{u},$$

$$A_1 = \begin{pmatrix} U & 0 & 1 \\ 0 & V & 0 \\ \Phi & 0 & U \end{pmatrix} \textit{ and } A_2 = \begin{pmatrix} 0 & V & 1 \\ U & 0 & 0 \\ 0 & \Phi & V \end{pmatrix}$$

3.2 Initial and Boundary Conditions

In order to obtain *exactly one solution*, it it is necessary to impose certain assumptions on the problem. Assumptions that can be justified by the physical meaning of the PDE are of two types:

- Initial Conditions: In the case of the wave, it is reasonable to specify the shape of the wave at $t = 0$ thus

  $$u(x, 0) = g(x) \quad \text{is given}$$

  for the wave equation it is also necessary to specify the velocity at $t = 0$ (does the wave start from rest?) so

  $$\partial_t u(x, 0) = f(x) \quad \text{is also known}$$

- Boundary Conditions: For the heat equation (stationary or not) it is necessary to specify what happens on the boundary Γ, there are three well known boundary conditions
  - $u|_\Gamma$ *Dirichlet boundary condition*, where the surface temperature is specified
  - $\partial_n u|_\Gamma$ *Neumann boundary condition*, where the heat flux out of the body is specified
  - $(\partial_n u + au)|_\Gamma$ where a is given, *Robin boundary condition*, the heat flux depends on the surface temperature (the body loses heat by conduction)

Definition 12 *: If the boundary condition is constantly equal to zero, then the condition is said to be* homogeneous*. If not then it is* inhomogeneous*.*

The heat conduction problem for which the heat distribution is known for $t = 0$ (an initial condition) and for which the surface temperature is specified (a boundary condition) is an example of an *initial-boundary value problem* or I.-B.V.P.

3.3 Well Posed Problems

3.3.1 Heat Equation

Assuming that all the data is continuous, consider the problem where $\Omega \subset \mathbb{R}^2$ is a simple connected open region with boundary Γ

$$\begin{cases} \partial_t u - \Delta u = q \quad x \in \Omega, \ 0 < t < T \\ u(x, 0) = f, \ x \in \Omega; \quad u(x, t) = h, \ x \in \Gamma, 0 < t < T \end{cases}$$

If there are two solutions u_1 and u_2 then $w = u_1 - u_2$ satisfies

$$\begin{cases} \partial_t w - \Delta w = 0 \quad x \in \Omega,\ 0 < t < T \\ w(x,0) = 0\ x \in \Omega; \quad w(x,t) = 0\ x \in \Gamma, 0 < t < T \end{cases}$$

multiply by w and integrate over the space domain

$$0 = \int_\Omega w\left(\partial_t w - \Delta w\right) d\Omega = \frac{1}{2}\frac{d}{dt}\int_\Omega |w|^2 d\Omega + \int_\Omega |\nabla w|^2 d\Omega - \int_\Gamma w\partial_n w d\Gamma$$

since $w = 0$ on Ω

$$0 = \frac{1}{2}\frac{d}{dt}\int_\Omega |w|^2 d\Omega + \int_\Omega |\nabla w|^2 d\Omega \geq \frac{1}{2}\frac{d}{dt}\int_\Omega |w|^2 d\Omega \quad 0 < t < T \tag{3.1}$$

The initial condition implies $0 = \int_\Omega |w|^2 d\Omega$ at $t = 0$ so the only way that (3.1) can be satisfied is if $0 = \int_\Omega |w|^2 d\Omega$ for all $t < T$ so from the continuity this implies that $w \equiv 0$ for all $x \in \Omega$ and all $0 \leq t < T$. So assuming a solution exits, it is unique.

3.4 Well Posed Cauchy Problems

Consider the Cauchy Problem for a linear *system*

$$\begin{cases} \boldsymbol{u}_t = A\boldsymbol{u}_x \quad \boldsymbol{u} \in \mathbb{R}^n \quad -\infty < x < \infty,\ t > 0 \\ \boldsymbol{u}(x,0) = \boldsymbol{f}, \quad -\infty < x < \infty. \end{cases} \tag{3.2}$$

Taking Fourier transforms, this leads to

$$\hat{\boldsymbol{u}}_t = -i\xi A\hat{\boldsymbol{u}} \tag{3.3}$$

assume that the $n \times n$ matrix A has eigenvalues and eigenvectors $\lambda_j,\ \boldsymbol{v}_j$ respectively for $j = 1, \ldots, n$ then a solution of (3.3) can be written as

$$\hat{\boldsymbol{u}} = \sum_{j=1}^n a_j e^{-i\lambda_j \xi t}\boldsymbol{v}_j \tag{3.4}$$

where the coefficients a_j are determined from the initial condition

$$\sum_{j=1}^n a_j \boldsymbol{v}_j = \hat{\boldsymbol{f}}$$

Consider $\boldsymbol{u} \in \mathbb{R}^2$ with $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, the eigenvalues are $\lambda = \pm i$ hence (3.4) becomes

$$\hat{\boldsymbol{u}} = a_1 \begin{pmatrix} 1 \\ -i \end{pmatrix} e^{\xi t} + a_2 \begin{pmatrix} 1 \\ i \end{pmatrix} e^{-\xi t}$$

which shows that in this case (3.2) has solutions that grow exponentially, as e^{Rt} for arbitrarily large R. This does not correspond to the behaviour of any physical system. Alternatively consider $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, the eigenvalues are $\lambda = \pm 1$, then

$$\hat{\boldsymbol{u}} = a_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{i\xi t} + a_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-i\xi t}$$

then

$$a_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \hat{\boldsymbol{f}} = \begin{pmatrix} \hat{\boldsymbol{f}}_1 \\ \hat{\boldsymbol{f}}_2 \end{pmatrix}$$

and

$$\hat{\boldsymbol{u}} = \begin{pmatrix} \hat{\boldsymbol{u}}_1 \\ \hat{\boldsymbol{u}}_2 \end{pmatrix} = \frac{1}{2}(\hat{\boldsymbol{f}}_1 + \hat{\boldsymbol{f}}_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}(\hat{\boldsymbol{f}}_1 - \hat{\boldsymbol{f}}_2) \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

hence

$$|\hat{\boldsymbol{u}}|^2 = |\hat{\boldsymbol{u}}_1|^2 + |\hat{\boldsymbol{u}}_2|^2 = |\hat{\boldsymbol{f}}_1|^2 + |\hat{\boldsymbol{f}}_2|^2 = |\hat{\boldsymbol{f}}|^2$$

which from Parseval's identity gives

$$\|\boldsymbol{u}(,t)\|^2 = \|\boldsymbol{u}(,0)\|^2$$

Definition 13 *: Consider the problem (3.2) for all initial values $\boldsymbol{f}$ for which $\|\boldsymbol{f}\| < \infty$. The problem is well posed if there are constants K and α such that*

$$\|\boldsymbol{u}(,t)\| \leq Ke^{\alpha t}\|\boldsymbol{u}(,0)\| \quad \forall t$$

The equation in (3.2) is said to be hyperbolic *if the eigenvalues of A are all real.*

There is a straightforward extension of this definition to systems in higher dimensions, given

$$\boldsymbol{u}_t = \sum_{j=1}^{d} A_j \boldsymbol{u}_x$$

applying Fourier transforms in $\mathbb{R}^d$ leads to

$$\hat{\boldsymbol{u}}_t = -i \sum_{j=1}^{d} \xi_j A_j \hat{\boldsymbol{u}}_x$$

Then the system is hyperbolic if all the eigenvalues of

$$P = \sum_{j=1}^{d} \xi_j A_j$$

are real for all $\xi_j, \ j = 1, \ldots, d$.

Consider the system $\boldsymbol{u}_t + A\boldsymbol{u}_x = 0$. If A can be diagonalised and all the eigenvalues are real then there exists P such that

$$PAP^{-1} = \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

then

$$P\boldsymbol{u}_t + PAPP^{-1}\boldsymbol{u}_x = 0$$
$$\boldsymbol{w}_t + \Lambda \boldsymbol{w}_x = 0$$

a set of *uncoupled* equations. A system such as $\boldsymbol{u}_t + A\boldsymbol{u}_x + B\boldsymbol{u} = 0$ would still be coupled after the transformation as B would not be diagonalised.

3.5 Discontinuities

Consider the equation

$$u_t + uu_x = 0, \qquad \begin{cases} -\infty < x < \infty \\ t > 0 \end{cases}$$

which has a general solution that can be written as

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

with the initial condition

$$u(x,0) = f(x) \equiv \begin{cases} A & x < 0 \\ B & x > 0 \end{cases}$$

i.e. a jump discontinuity in the initial condition. If $A < B$ then from the characteristics, which are straight lines,

$$u(x,t) = \begin{cases} A & x < At \\ B & x > Bt \end{cases}$$

and in the gap (see figure 3.1a), by interpolation,

$$x - ut = 0, \qquad At \le x \le Bt$$

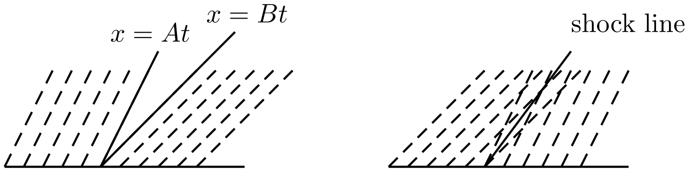


Figure 3.1: (a) (b)

However, if $A > B$, then the characteristics overlap in the region $Bt \le x \le At$. From the characteristics,

$$u(x,t) = \begin{cases} A & x < Bt \\ B & x > At \end{cases}$$

Within the overlap region the solution there are two possibilities, so the solution is not uniquely defined without some additional information. If the equations are modelling gas dynamics then it is known that a shock wave exists and there is a jump discontinuity along the shock (in this example the shock would be $x = \frac{A+B}{2}t$, see figure 3.1b). Further study of this topic is beyond the scope of this course, see for example Wikipedia [12].

Chapter 4

Classification of PDEs

4.1 Equations of Second Order

Consider

$$au_{xx} + bu_{xy} + cu_{yy} = 0 \tag{4.1}$$

write as a 1st order system with $w = u_x$ and $v = u_y$ then equations are

$$\begin{array}{lcl} w_x + \frac{b}{a} w_y + \frac{c}{a} v_y & = & 0 \\ v_x - w_y & = & 0 \end{array}$$

or $\boldsymbol{w}_x + A\boldsymbol{w}_y = 0$,

$$A = \begin{pmatrix} \frac{b}{a} & \frac{c}{a} \\ -1 & \end{pmatrix}$$

so eigenvalues λ satisfy $a\lambda^2 - b\lambda + c = 0$ roots are real if $b^2 - 4ac \geq 0$, in which case the equation is *hyperbolic*.

In general in $\mathbb{R}^2$ consider

$$0 = a_{11}u_{xx} + 2a_{12}u_{xy} + a_{22}u_{yy} + a_1 u_x + a_2 u_y + a_0 u$$

note $u_{xy} = u_{yx}$ hence $a_{12} + a_{21} = 2a_{12}$. Then the second order part of the differential operator is

$$a_{11}\partial_{xx} + 2a_{12}\partial_{xy} + a_{22}\partial_{yy}$$

w.l.o.g. assume $a_{11} = 1$ then

$$\partial_{xx} + 2a_{12}\partial_{xy} + a_{22}\partial_{yy} = (\partial_x + a_{12}\partial_y)^2 + (-a_{12}^2 + a_{22})\partial_{yy}$$

consider 3 cases depending on the sign of the second coefficient!

4.1.1 $(a_{22} - a_{12}^2) > 0$

Change variables

$$\begin{aligned} x &= \alpha\xi + \beta\zeta \\ y &= \gamma\xi + \delta\zeta \\ \partial_\xi u &= u_x \tfrac{\partial x}{\partial \xi} + u_y \tfrac{\partial y}{\partial \xi} \\ \partial_\zeta u &= u_x \tfrac{\partial x}{\partial \zeta} + u_y \tfrac{\partial y}{\partial \zeta} \end{aligned}$$

so that the operator becomes

$$\partial_\xi^2 + \partial_\zeta^2$$

such equations are known as *Elliptic Equations.* But

$$\begin{array}{ll} \tfrac{\partial x}{\partial \xi} = \alpha & \tfrac{\partial y}{\partial \xi} = \gamma \\ \tfrac{\partial x}{\partial \zeta} = \beta & \tfrac{\partial y}{\partial \zeta} = \delta \end{array}$$

so

$$\left.\begin{aligned} \partial_\xi &= \alpha\partial_x + \gamma\partial_y \\ \partial_\zeta &= \beta\partial_x + \delta\partial_y \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{array}{ll} \alpha = 1 & \gamma = a_{12} \\ \beta = 0 & \delta = \sqrt{a_{22} - a_{12}^2} \end{array}\right.$$

hence the condition $(a_{22} - a_{12}^2) > 0$ is essential and then

$$x = \xi \qquad y = a_{12}\xi + \left(\sqrt{(a_{22} - a_{12}^2)}\right)\zeta$$

4.1.2 $(a_{22} - a_{12}^2) < 0$

Using the transformation with $\delta = \sqrt{a_{12}^2 - a_{22}}$ leads to the operator

$$\partial_{\xi\xi} - \partial_{\zeta\zeta} \equiv (\partial_\xi - \partial_\zeta)(\partial_\xi + \partial_\zeta)$$

such equations are known as *Hyperbolic Equations.*

4.1.3 $(a_{22} - a_{12}^2) = 0$

The transformation leads to

$$\partial_{\xi\xi} + \text{first order terms} \tag{4.2}$$

such equations are known as *Parabolic Equations.*

4.1.4 Higher Dimensions

In $\mathbb{R}^d$ a second order equation can be written as

$$\sum_{i,j=1}^{d} a_{ij} u_{x_i x_j} + \sum_{i=1}^{d} a_i u_{x_i} + a_0 u = c$$

assume $a_{ij} = a_{ji}$, *i.e.* coefficient matrix A is *symmetric.*

Definition 14

- *Equation (4.2) is Elliptic iff A or $-A$ is* positive definite
- *Equation (4.2) is Hyperbolic if all eigenvalues are non zero and all but one have the same sign.*
- *Equation (4.2) is Parabolic iff there is exactly one zero eigenvalue and all others have the same sign*

In $\mathbb{R}^2$,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$$

eigenvalues λ satisfy

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}^2 = 0$$
$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}^2)$$

so $\lambda = \frac{1}{2}\left((a_{11} + a_{22}) \pm \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}^2)}\right)$ both roots are positive iff $a_{11}a_{22} - a_{12}^2 < 0$, with $a_{11} = 1$ this leads to the same condition as before. In $\mathbb{R}^3$, Laplace's Equation

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

defines $A = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$ and hence is elliptic, whereas the wave equation

$$u_{xx} - u_{yy} - u_{zz} = 0$$

defines $A = \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix}$ and hence is hyperbolic. The heat equation

$$u_t - u_{xx} - u_{yy} = 0$$

defines $A = \begin{pmatrix} 0 & & \\ & -1 & \\ & & -1 \end{pmatrix}$ and hence is parabolic.

4.2 Wave Equation (in $\mathbb{R}^2$)

$$u_{tt} - c^2 u_{xx} = 0$$
$$(\partial_t - c\partial_x)(\partial_t + c\partial_x)u = 0$$

Change variables

$$\xi = x + ct, \qquad \zeta = x - ct$$

then

$$\begin{cases} \partial_x u &= \partial_\xi u \frac{\partial \xi}{\partial x} + \partial_\zeta u \frac{\partial \zeta}{\partial x} \\ \partial_t u &= \partial_\xi u \frac{\partial \xi}{\partial t} + \partial_\zeta u \frac{\partial \zeta}{\partial t} \end{cases}$$

so

$$\partial_x = \partial_\xi + \partial_\zeta \qquad \partial_t = c\partial_\xi - c\partial_\zeta$$

and

$$\partial_t - c\partial_x = -2c\partial_\zeta \qquad \partial_t + c\partial_x = 2c\partial_\xi$$

and

$$4c\partial_{\xi\zeta} u = 0 \qquad \Rightarrow \qquad u(\xi, \zeta) = g(\xi) + f(\zeta)$$

so

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

Both $g(x+ct)$ and $f(x-ct)$ are waves that travel along the lines $x+ct = \alpha$ and $x-ct = \beta$. Assume $c > 0$, then $f(x-ct)$ travels to the right with speed c.

4.2.1 Initial Value Problem

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0 \\ u(x,0) = \phi(x) \quad \partial_t u(x,0) = \psi(x) \end{cases}$$

then

$$\begin{aligned} u(x,t) &= g(x+ct) + f(x-ct) \\ \partial_t u(x,t) &= cg'(x+ct) - cf'(x-ct) \end{aligned}$$

so at time $t = 0$

$$\begin{aligned} g(x) + f(x) &= \phi(x) \\ cg'(x) - cf'(x) &= \psi(x) \\ g'(x) + f'(x) &= \phi'(x) \\ g'(x) - f'(x) &= \tfrac{1}{c}\psi(x) \end{aligned}$$

and

$$\begin{aligned} g'(x) &= \tfrac{1}{2}\left(\phi'(x) + \tfrac{1}{c}\psi(x)\right) \\ f'(x) &= \tfrac{1}{2}\left(\phi'(x) - \tfrac{1}{c}\psi(x)\right) \end{aligned}$$

so

$$\begin{aligned} g(z) &= \int_0^z \tfrac{1}{2}\left(\phi'(x) + \tfrac{1}{c}\psi(x)\right) dx + g(0) \\ f(z) &= \int_0^z \tfrac{1}{2}\left(\phi'(x) - \tfrac{1}{c}\psi(x)\right) dx + f(0) \\ g(z) &= \tfrac{1}{2}\phi(z) + \tfrac{1}{2c}\int_0^z \psi(x)dx + A \\ f(z) &= \tfrac{1}{2}\phi(z) - \tfrac{1}{2c}\int_0^z \psi(x)dx + B \end{aligned}$$

where $A + B = 0$ since $\phi(x) = f(x) + g(x)$ hence

$$\begin{aligned} u(x,t) &= \tfrac{1}{2}\phi(x+ct) + \tfrac{1}{2c}\int_0^{x+ct} \psi(z)dz \\ &\quad + \tfrac{1}{2}\phi(x-ct) - \tfrac{1}{2c}\int_0^{x-ct} \psi(z)dz \\ &= \tfrac{1}{2}\phi(x+ct) + \tfrac{1}{2}\phi(x-ct) + \tfrac{1}{2c}\int_{x-ct}^{x+ct} \psi(z)dz \end{aligned}$$

Example 15

$$\begin{cases} u_{xx} + 2u_{xt} - 20u_{tt} & = & 0 \\ \\ u(x,0) & = & \phi(x) \\ u_t(x,0) & = & \psi(x) \end{cases}$$

$$\begin{array}{ll} & a_{22} = -20, \quad a_{12} = 1 \quad a_{11} = 1 \\ \Rightarrow & a_{22} - a_{12}^2 = -20 - 1 < 0 \\ \Rightarrow & \textit{Hyperbolic} \end{array}$$

change of variables $x = \xi$, $t = \xi + (\sqrt{21})\zeta$ *so* $\partial_\xi = \partial_x + \partial_t$ *and* $\partial_\zeta = \sqrt{21}\partial_t$ *so*

$$\partial_{xx}^2 + 2\partial_{xt}^2 - 20\partial_{tt}^2 = (\partial_\xi + \partial_\zeta)(\partial_\xi - \partial_\zeta)$$

and

$$u(\xi, \zeta) = f(\xi - \zeta) + g(\xi + \zeta)$$

and

$$u(x,y) = f(x + \frac{x-t}{\sqrt{21}}) + g(x - \frac{x-t}{\sqrt{21}})$$

4.2.2 Laplace's Equation

The characteristics of an elliptic equation in $\mathbb{R}^2$ are complex for example those for

$$u_{xx} + u_{yy} = (\partial_x + i\partial_y)(\partial_x - i\partial_y)u = 0$$

are $x + iy = \textit{constant}$ and $x - iy = \textit{constant}$. Consider the change of variables

$$\xi = x + iy, \qquad \zeta = x - iy$$

then

$$\begin{cases} \partial_x u & = & \partial_\xi u \frac{\partial \xi}{\partial x} + \partial_\zeta u \frac{\partial \zeta}{\partial x} \\ \partial_y u & = & \partial_\xi u \frac{\partial \xi}{\partial y} + \partial_\zeta u \frac{\partial \zeta}{\partial y} \end{cases}$$

so

$$\partial_x = \partial_\xi + \partial_\zeta, \qquad \partial_y = i\partial_\xi - i\partial_\zeta$$

and

$$\partial_y - i\partial_x = -2i\partial_\zeta, \qquad \partial_y + i\partial_x = 2i\partial_\xi$$

and

$$4i\partial_{\xi\zeta}^2 u = 0 \qquad \Rightarrow \qquad u(\xi, \zeta) = g(\xi) + f(\zeta)$$

so

$$u(x,t) = g(x + iy) + f(x - iy)$$

In particular $e^{k(x+iy)} \pm e^{k(x-iy)} = e^{kx}\left(e^{iky} \pm e^{-iky}\right)$ is a solution for any k (real or complex).

Exercises

Exercise 4.1 *What is the solution of the initial value problem*

$$\left\{\begin{array}{c} u_{tt} - 4u_{xx} = 0 \quad -\infty < x < \infty,\ t > 0 \\ u(x,0) = e^{-x^2}, \quad \partial_t u(x,0) = 0 \quad -\infty < x < \infty \end{array}\right. ?$$

Exercise 4.2

Consider the differential equation

$$u_{xx} - 6u_{xt} - 7u_{tt} = 0.$$

- *What are the characteristics of the equation?*
- *What is the general solution?*
- *Given the initial conditions*

$$\begin{array}{rcl} u(x,0) & = & e^{-x^2}, \quad -\infty < x < \infty \\ u_t(x,0) & = & 0, \quad -\infty < x < \infty \end{array}$$

what is the solution?

Exercise 4.3 *Consider the differential equation*

$$u_{xx} - 8u_{xt} + 15u_{tt} = 0.$$

- *What are the characteristics of the equation?*
- *What is the general solution?*
- *Given the initial conditions*

$$\begin{array}{rcl} u(x,0) & = & 0, \quad -\infty < x < \infty \\ u_t(x,0) & = & \left\{\begin{array}{ll} 1 & 0 < x < 1 \\ 0 & \mathit{otherwise} \end{array}\right. \end{array}$$

what is the solution?

Chapter 5

Energy

5.1 Causality and Energy

Consider the wave equation posed as an initial value problem for:

$$\begin{cases} u_{tt} - c^2 u_{xx} = 0, \quad t > 0, -\infty < x < \infty \\ u(x,0) = \phi(x), \quad -\infty < x < \infty \\ \partial_t u(x,0) = \psi(x), \quad -\infty < x < \infty. \end{cases}$$

Definition 16 *Principle of Causality: An initial condition (ϕ, ψ) at the point $(x_0, 0)$ can effect the solution $u(x,t)$, $t > 0$ only in the set $S(x_0, c) = \{|x - x_0| \leq ct, t > 0\}$. The causality principle in $\mathbb{R}^3$ is known as* Huygens's Principle.

5.1.1 Consequence 1

If $\phi|_{|x|>R} = 0$ and $\psi|_{|x|>R} = 0$ then the solution is $u(x,t) = 0$ for $|x| > R + ct$.

5.1.2 Consequence 2

- If $\phi \equiv 0 \equiv \psi$ then $u = 0$.
- If $\phi_1 = \phi_2$, $\psi_1 = \psi_2$ and

  $$\begin{cases} u_{tt} - c^2 u_{xx} = 0 \\ u|_{t=0} = \phi_1 \quad \partial_t u|_{t=0} = \psi_1 \end{cases} \quad \text{and} \quad \begin{cases} v_{tt} - c^2 v_{xx} = 0 \\ v|_{t=0} = \phi_2 \quad \partial_t v|_{t=0} = \psi_2 \end{cases}$$

  then $u = v$ as $w = u - v$ solves

  $$\begin{cases} w_{tt} - c^2 w_{xx} = 0 \\ w|_{t=0} = 0, \quad \partial_t w|_{t=0} = 0 \end{cases}$$

Definition 17 *: The Domain of Dependence or past history of the point (x_0, t_0) is the region of (x,t) on which the value of $u(x_0, t_0)$ depends.*

In particular from the formula

$$u(x,t)=\frac{1}{2}\phi(x+ct)+\frac{1}{2}\phi(x-ct)+\frac{1}{2c}\int_{x-ct}^{x+ct}\psi(z)dz$$

it follows that the value of u at a point (x_0,t_0) is influenced only by the value of ϕ at (x_0-ct_0) and (x_0+ct_0) and the value of ψ on $[x_0-ct_0,x_0+ct_0]$. Analogously, if $t_1\leq t_0$ then the value of $u(x_0,t_0)$ is only influenced by $u(x_0-c(t_1-t_0))$ and $u(x_0+c(t_1-t_0))$ and by $\partial_t u$ on $[x_0-c(t_1-t_0),x_0+c(t1-t_0)]$.

5.2 The Conservation of Energy

Consider an *infinite string* with constant density ρ and constant tension T so

$$\int(\rho u_{tt}-Tu_{xx})=0$$

the kinetic energy of the string is

$$k_e=\frac{l}{2}\rho\int_{\infty}^{\infty}u_t^2dx$$

the potential energy is

$$P_e=\frac{1}{2}T\int_{\infty}^{\infty}u_x^2dx$$

$$E(t)=k_e+P_e$$

Assume that the initial data ϕ and ψ are zero outside $[-R,R]$. From consequence 1 it follows that $u,\partial_t u\equiv 0$ for $|x|>R+ct$. It is necessary to show that $E(t)=E(0)$, to prove this it is sufficient to show that $\frac{d}{dt}E(t)=0$,

$$\begin{aligned}\tfrac{d}{dt}k_e &= \tfrac{l}{2}\rho\tfrac{d}{dt}\textstyle\int_{\infty}^{\infty}u_t^2dx\\ &= \rho\textstyle\int_{\infty}^{\infty}u_tu_{tt}dx \qquad \text{using } u_{tt}=\tfrac{t}{\rho}u_{xx}\\ &= T\textstyle\int_{\infty}^{\infty}u_tu_{xx}dx\end{aligned}$$

On the other hand

$$\frac{d}{dt}P_e=T\int_{\infty}^{\infty}u_xu_{xt}dx$$

so

$$\begin{aligned}\tfrac{d}{dt}E(t) &= T\textstyle\int_{\infty}^{\infty}(u_xu_{xt}+u_tu_{xx})dx\\ &= T\textstyle\int_{\infty}^{\infty}\partial_x(u_tu_x)dx\\ &= 0 \qquad \lim_{x\to\pm\infty}u_tu_x=0\end{aligned}$$

5.3 The Diffusion Equation

Consider

$$u_t=\kappa u_{xx}\qquad \kappa=\text{constant}\tag{5.1}$$

5.3.1 Maximum Principle

Definition 18 *Maximum Principle: If u solves (5.1) in a rectangle $(0 \leq x \leq l, 0 \leq t \leq T)$ then the maximum value of $u(x,t)$ is assumed either at $t = 0$ or $x = 0$ or $x = l$. The same applies to the minimum value.*

Assume maximum is at an interior point (x_0, T_0) within $R = [0,l] \times [0,T]$ so

$$u_x(x_0,t_0) = u_t(x_0,t_0) = 0 \qquad u_{xx}(x_0,t_0) \leq 0.$$

If $u_{xx}(x_0,t_0) \neq 0$ then $\kappa u_{xx}(x_0,t_0) = u_t(x_0,t_0) < 0$ which is a contradiction. Consider the function $v(x,t) = u(x,t) + \epsilon x^2$ $(\epsilon > 0)$ if the maximum of v is at an interior point (x_0, T_0) then

$$v_x(x_0,t_0) = v_t(x_0,t_0) = 0 \qquad v_{xx}(x_0,t_0) \leq 0$$

and

$$0 \leq v_t - \kappa v_{xx} = u_t - \kappa u_{xx} - 2\epsilon\kappa = -2\epsilon\kappa < 0$$

another contradiction so the maximum of v must be on the boundary. Assume the maximum of u on the boundary is M, then the maximum of v cannot exceed $M + \epsilon l^2$ in the limit as ϵ tends to zero v tends to u so the maximum of u must also be on the boundary.

One immediate consequence of the maximum principle is the uniqueness of the solution, since if u and v are both solutions $w = u - v$ satisfies zero boundary conditions and hence must be zero everywhere.

5.3.2 Energy

Assume $u(l,t) = u(0,t) = 0$

$$E = \frac{1}{2}\int_0^l u^2(x,t)dx$$

then

$$\begin{aligned} \tfrac{d}{dt}E &= \textstyle\int_0^l u_t u dx \\ &= \textstyle\int_0^l \kappa u_{xx} u dx \\ &= \kappa \textstyle\int_0^l \partial_x(uu_x)dx - \kappa\int_0^l u_x^2 dx \\ &= \kappa u u_x|_0^l - \kappa \textstyle\int_0^l u_x^2 dx \\ &\leq 0 \end{aligned}$$

Thus it is possible to prove uniqueness via energy as

$$0 \leq E(t) \leq E(0) = \frac{1}{2}\int_0^l w^2(x,0)dx = 0$$

and

$$E(t) = 0 \quad \Rightarrow \quad w = 0$$

Definition 19 *Stability: In general a system is stable if close initial data generate close solutions.*

Define distance in terms of the $\mathcal{L}_2$ norm

$$\|f-g\|_2 = \left(\int_0^l (f-g)^2 dx\right)^{1/2}$$

Consider two solution u_1 and u_2 of the diffusion equation, with the initial conditions $u_1(x,0)=\phi_1$ and $u_2(x,0)=\phi_2$. Note that $w=u_1-u_2$ solves the equation subject to the initial condition $w|_{t=0}=\phi_1-\phi_2$, then from the energy inequality

$$E(w)^{1/2} = \left(\int_0^l (u_1-u_2)^2 dx\right)^{1/2} \leq \left(\int_0^l (\phi_1-\phi_2)^2 dx\right)^{1/2}$$

So if ϕ_1 and ϕ_2 are close in the sense of the $\mathcal{L}_2$ norm then u_1 and u_2 are close *uniformly in time* in the same norm.

It is possible to define distance in term of the maximum or $\mathcal{L}_\infty$ norm

$$\|f-g\|_\infty = \max_{[0,l]} |f-g|$$

Assume that the solutions u_1 and u_2 satisfy the same boundary conditions

$$u_1(0,t)=u_2(0,t)=g, \qquad u_1(l,t)=u_2(l,t)=f$$

then $w(0,t)=w(l,t)=0$ and by the maximum principle

$$\max w = \max\{\max_{[0,l]}(\phi_1-\phi_2), 0\}$$

and

$$\max\{-w\} = \max\{\max_{[0,l]}(\phi_2-\phi_1), 0\}$$

so

$$\max_{[0,l]} |u_1-u_2|(t) \leq \max_{[0,l]} |\phi_1-\phi_2|$$

for all t.

5.4 Solution of Diffusion Equation

Consider only the initial value problem

$$\left\{\begin{array}{rcl} u_t &=& \kappa u_{xx}, = \frac{dy}{b(x,y)}, t>0 \\ u|_{t=0} &=& \phi(x), = \frac{dy}{b(x,y)} \end{array}\right. \tag{5.2}$$

5.4.1 Properties of the solution

- Translation invariance, if $u(x,t)$ is a solution of (5.2) so is $u(x-y,t)$ for any value y.
- If $u(x,t)$ is a solution of (5.2) then all derivatives of any order are solutions of (5.2).

- A linear combination of solutions of (5.2) is a solution of (5.2).
- If $w(x,t)$ is a solution of (5.2) then for any *smooth* function $g(y)$

$$v(x,t) = \int w(x-y,t)g(y)dy$$

- If $u(x,t)$ is a solution of (5.2), the $u(\lambda x, \lambda^2 t)$ is a solution of (5.2).

is a solution of (5.2)

5.5 Fundamental Solutions (Green's Functions)

Consider the IVP

$$\left\{ \begin{array}{rcl} u_t & = & \kappa u_{xx} \\ u|_{t=0} & = & q(x) = \left\{ \begin{array}{l} 1 \text{ for } x > 0 \\ 0 \text{ for } x < 0 \end{array} \right. \end{array} \right. \tag{5.3}$$

Note that under dilation

$$q(x) \quad \rightarrow \quad q(\lambda x) = \left\{ \begin{array}{l} 1 \text{ for } x > 0 \\ 0 \text{ for } x < 0 \end{array} \right. \quad = q(x)$$

So it may be assumed that the solution Q does not change under dilation

$$Q(x,t) = Q(\lambda x, \lambda^2 t)$$

Assume the solution is of the form

$$Q(x,t) = g\left(\frac{x}{2\sqrt{\kappa t}}\right)$$

Now define $p = \frac{x}{2\sqrt{\kappa t}}$ then

$$\begin{array}{rcl} Q_t & = & \frac{dg}{dp} \cdot \frac{dp}{dt} = q'(p)\left(-\frac{1}{2t}p\right) \\ Q_x & = & \frac{dg}{dp} \cdot \frac{dp}{dx} = q'(p)\left(-\frac{1}{2\sqrt{\kappa t}}\right) \\ Q_{xx} & = & \frac{dQ_x}{dp} \cdot \frac{dp}{dx} = q''(p)\left(\frac{1}{4\kappa t}\right) \end{array}$$

$$0 = Q_t - \kappa Q_{xx} = \frac{1}{t}\left(-\frac{1}{2}pg'(p) - \frac{1}{4}g''(p)\right)$$

which is an ode with a solution

$$g(p) = c_1 \int_0^p e^{-z^2} dz + c_2$$

so

$$Q(x,t) = c_1 \int_0^{\frac{x}{2\sqrt{\kappa t}}} e^{-z^2} dz + c_2 \qquad \forall t > 0$$

To check on the behaviour in the limit as $t \rightarrow 0$,

- If $x > 0$

$$\lim_{t\to 0^+} c_1 \int_0^{\frac{x}{2\sqrt{\kappa t}}} e^{-z^2} dz + c_2 \quad = \quad c_1 \int_0^{\infty} e^{-z^2} dz + c_2 \quad = \quad c_1 \frac{\sqrt{\pi}}{2} + c_2$$

- If $x < 0$

$$\lim_{t\to 0^+} c_1 \int_0^{\frac{x}{2\sqrt{\kappa t}}} e^{-z^2} dz + c_2 \quad == \quad -c_1 \int_{-\infty}^{0} e^{-z^2} dz + c_2 \quad = \quad -c_1 \frac{\sqrt{\pi}}{2} + c_2$$

- But

$$Q(x,0) = \begin{cases} 1 \text{ for } x > 0 \\ 0 \text{ for } x < 0 \end{cases}$$

so

$$\left.\begin{array}{r} c_1 \frac{\sqrt{\pi}}{2} + c_2 = 1 \\ -c_1 \frac{\sqrt{\pi}}{2} + c_2 = 0 \end{array}\right\} \quad \Rightarrow \quad \left\{\begin{array}{c} c_1 = \frac{1}{\sqrt{\pi}} \\ c_2 = \frac{1}{2} \end{array}\right.$$

$$Q(x,t) = \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{\frac{x}{2\sqrt{\kappa t}}} e^{-z^2} dz$$

Now define the *fundamental solution* $S = \frac{\partial Q}{\partial x}$

- S is a solution
- $\lim_{t\to 0^+} \frac{\partial Q}{\partial x}(t) = \delta_x$

Note

$$\begin{aligned} S(x,t) &= \partial_x \left(\tfrac{1}{2} + \tfrac{1}{\sqrt{\pi}} \int_0^{\frac{x}{2\sqrt{\kappa t}}} e^{-z^2} dz \right) \\ &= \tfrac{1}{\sqrt{\pi}} e^{-\frac{x^2}{4\kappa t}} \tfrac{1}{2\sqrt{\kappa t}} \end{aligned}$$

So if $x \neq 0$ then

$$\lim_{t\to 0^+} \frac{1}{\sqrt{2\pi\kappa t}} e^{-\frac{x^2}{4\kappa t}} = 0$$

as an exponential decays faster than a polynomial. If $x = 0$ then

$$\partial_x Q(0,t) = \frac{1}{2\pi\kappa t} \qquad \forall t \neq 0$$

this appears to be singular and needs to be interpreted using the theory of distributions.

If the given initial condition is $u(x,0) = \phi(x)$ s.t. $\lim_{|x|\to\infty} \phi(x) = 0$ then

$$u(x,t) = \int_{-\infty}^{\infty} S((x-y),t)\phi(y)dy \qquad \forall t > 0$$

It can be shown that this solution is unique. An integral of the form

$$h(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy$$

is known as a *convolution* and is denoted by $h = f * g$.

Example 20 *Solve*

$$\begin{aligned} u_t &= \kappa u_{xx} \\ u(x,0) &= x^2 \end{aligned}$$

Define $p = \frac{x-y}{2\sqrt{\kappa t}}$ *then* $dp = -\frac{1}{2\sqrt{\kappa t}} dy$ *then*

$$\begin{aligned} u(x,t) &= \textstyle\int_{-\infty}^{\infty} \frac{1}{2\sqrt{\pi\kappa t}} e^{-\frac{(x-y)^2}{4\kappa t}} y^2 dy \\ &= \textstyle\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} (x - p\sqrt{4\kappa t})^2 dp \\ &= \textstyle\frac{x^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} dp - \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} pdp + \frac{4\kappa t}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} p^2 dp. \end{aligned}$$

(Given

$$\frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} p^2 dp = 1,$$

$$\frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} pdp = 0,$$

$$\frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-p^2} dp = 2)$$

Chapter 6

Separation of Variables

6.1 Solution using Fourier series

Consider the initial-boundary value problem[1]

$$\left\{ \begin{array}{rcl} u_t & = & \kappa u_{xx} \quad \forall x \in [0,l], \quad \forall t > 0 \\ u(x,0) & = & f(x) \quad \forall x \in [0,l] \\ u(0,t) = & 0 & = u(l,t) \quad \forall t > 0 \end{array} \right. \tag{6.1}$$

A solution is required with the following property: u is a product in which the dependence of u on x, t is separated, that is:

$$u(x,t) = X(x)T(t) \tag{6.2}$$

This solution technique is called *separation of variables.* Substituting u into the heat equation,

$$\frac{T'(t)}{\kappa T(t)} = \frac{X''(x)}{X(x)}$$

Since the right hand side depends only on x and the left hand side only on t, both sides are equal to some constant value $-\lambda$. Thus:

$$T'(t) = -\lambda\kappa T(t) \tag{6.3}$$

and

$$X''(x) = -\lambda X(x) \tag{6.4}$$

1. Suppose that $\lambda < 0$. Then there exist real numbers B, C such that

   $$X(x) = Be^{\sqrt{-\lambda}\,x} + Ce^{-\sqrt{-\lambda}\,x}.$$

   From the boundary conditions

   $$X(0) = 0 = X(l)$$

   and therefore $B = 0 = C$ which implies u is identically 0.

[1] The following solution technique for the heat equation was proposed by Joseph Fourier in his treatise *Theorie analytique de la chaleur*, published in 1822.

2. Suppose that $\lambda = 0$. Then there exist real numbers B, C such that

$$X(x) = Bx + C$$

From the boundary conditions, conclude that u is identically 0.

3. Therefore, it must be the case that $\lambda = \beta^2 > 0$. Then there exist real numbers A, B, C such that

$$T(t) = Ae^{-\lambda\kappa t}$$

and

$$X(x) = B\cos(\beta x) + C\sin(\beta x)$$

A, B and C are determined by the data

$$u(0,t) = u(l,t) = 0 \quad \Rightarrow \quad X(0) = X(l) = 0$$

$$X(0) = 0 \quad \Rightarrow \quad B = 0 \quad \Rightarrow \quad X(x) = C\sin(\beta x)$$

$$X(l) = 0 \quad \Rightarrow \quad C\sin(\beta l) = 0 \quad \Rightarrow \quad \beta = \frac{\pi n}{l}$$

for some positive integer n.

So in general from the linearity any finite sum

$$u_k(x,t) = \sum_{n=1}^{k} C_n \sin\left(\frac{n\pi x}{l}\right) e^{-\dfrac{n^2\pi^2\kappa t}{l^2}}$$

is a solution of the PDE and the boundary conditions. To be a solution of the I-BVP it follows that, assuming the infinite sum converges,

$$u(x,t) = \sum_{n=1}^{+\infty} C_n \sin\left(\frac{n\pi x}{l}\right) e^{-\dfrac{n^2\pi^2\kappa t}{l^2}}$$

it must also satisfy the the initial condition so

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

From the theory of Fourier Series it follows that

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

This solves the heat equation in the special case that the dependence of u has the special form (6.2).

6.2 Wave Equation

Consider the initial-boundary value problem

$$\left\{\begin{array}{rcl} u_{tt} - c^2 u_{xx} & = & 0, \quad \forall x \in [0,l], \quad \forall t > 0 \\ u(x,0) & = & \phi(x), \quad \forall x \in [0,l] \\ u_t(x,0) & = & \psi(x), \quad \forall x \in [0,l] \\ u(0,t) = & 0 & = u(l,t) \quad \forall t > 0. \end{array}\right. \tag{6.5}$$

A solution of the form (6.2) leads to

$$\frac{T''(t)}{kT(t)} = c^2 \frac{X''(x)}{X(x)}$$

so

$$\frac{T''(t)}{c^2 T(t)} = \frac{X''(x)}{X(x)} = -\lambda$$

as for the diffusion equation $\lambda \neq 0$. If $0 < \lambda = \beta^2$ then

$$\left\{\begin{array}{rcl} T''(t) & = & -c^2\beta^2 T(t) \\ X''(x) & = & -\beta^2 X(x) \end{array}\right.$$

these are uncoupled ODEs with the solutions

$$\left\{\begin{array}{l} T(t) = A\cos(\beta ct) + B\sin(\beta ct) \\ X(x) = C\cos(\beta x) + D\sin(\beta x) \end{array}\right.$$

A, B, C and D are determined by the data

$$u(0,t) = u(l,t) = 0 \quad \Rightarrow \quad X(0) = X(l) = 0$$

$$X(0) = 0 \quad \Rightarrow \quad C = 0 \quad \Rightarrow \quad X(x) = D\sin(\beta x)$$

$$X(l) = 0 \quad \Rightarrow \quad D\sin(\beta l) = 0 \quad \Rightarrow \quad \beta = \frac{\pi n}{l}$$

So in general from the linearity any finite sum

$$u_k(x,t) = \sum_{n=1}^{k} \left(A_n \cos(\frac{n\pi ct}{l}) + B_n \sin(\frac{n\pi ct}{l}) \right) \sin(\frac{n\pi x}{l})$$

is a solution of the PDE and the boundary conditions. To be a solution of the I-BVP it follows that, assuming the infinite sum converges,

$$u(x,0) = \sum_{n=1}^{\infty} A_n \sin(\frac{n\pi x}{l}) = \phi(x)$$

and

$$\partial_t u(x,0) = \sum_{n=1}^{\infty} B_n \frac{n\pi c}{l} \sin(\frac{n\pi x}{l}) = \psi(x)$$

These are two uncoupled Fourier Series from which it is possible to compute the values of the coefficients A_n and B_n.

6.2.1 Neumann Conditions

The wave equation can be specified subject to different boundary conditions:

$$\begin{cases} u_{tt} - c^2 u_{xx} &= 0, \quad \forall x \in [0,l], \quad \forall t > 0 \\ u(x,0) &= \phi(x), \quad \forall x \in [0,l] \\ u_t(x,0) &= \psi(x), \quad \forall x \in [0,l] \\ u_x(0,t) = & 0 = u_x(l,t) \quad \forall t > 0. \end{cases} \tag{6.6}$$

The ODE BVP for X now becomes

$$\begin{cases} X''(x) &= -\lambda X(x) \\ X'(0) &= X'(l) \ = \ 0 \end{cases}$$

It is no longer guaranteed that $\lambda > 0$

- $\lambda = 0 \quad \Rightarrow \quad X'' = 0 \quad \Rightarrow \quad X(x) = Ax + B \quad \Rightarrow \quad X' = 0 \quad \Rightarrow \quad A = 0$
- $\lambda = \beta^2 > 0$
  The possible solutions are then of the form:
  $$\begin{aligned} X(x) &= C\cos(\beta x) + D\sin(\beta x) \\ X'(x) &= -C\beta\sin(\beta x) + D\beta\cos(\beta x) \\ X'(0) &= D\beta = 0 \quad \Rightarrow \quad D = 0 \\ X'(l) &= -C\beta\sin(\beta l) = 0 \quad \Rightarrow \quad \beta = \tfrac{\pi n}{l} \\ X_n(x) = \cos\left(\tfrac{\pi n}{l}\right) & \end{aligned}$$
- $\lambda < 0$ not possible

Thus returning to $T''(t) = -c^2\lambda T(t)$

- $\lambda = 0$ leads to $T(t) = A + Bt$ so $u(x,t) = A + Bt$
- $\lambda > 0$ from the linearity any finite sum
  $$u_k(x,t) = \sum_{n=1}^{k} \left(A_n \cos(\frac{n\pi ct}{l}) + B_n \sin(\frac{n\pi ct}{l}) \right) \cos(\frac{n\pi x}{l})$$
  is a solution.

So assuming the infinite sum converges the general solution is

$$u(x,t) = \frac{1}{2}A_0 + \frac{1}{2}B_0 t + \sum_{n=1}^{\infty} \left(A_n \cos(\frac{n\pi ct}{l}) + B_n \sin(\frac{n\pi c}{l}) \right) \cos(\frac{n\pi x}{l})$$

The factor $\frac{1}{2}$ is for normalisation and

$$u(x,0) = \frac{1}{2}A_0 + \sum_{n=1}^{\infty} A_n \cos(\frac{n\pi x}{l}) = \phi(x)$$

and

$$\partial_t u(x,0) = \frac{1}{2}B_0 + \sum_{n=1}^{\infty} B_n \frac{n\pi ct}{l} \cos(\frac{n\pi x}{l}) = \psi(x)$$

6.2.2 Mixed Boundary Conditions

If the boundary conditions are

$$u(0,t) = u_x(l,t) = 0$$

A solution of the form (6.2) leads to

$$\begin{cases} X''(x) &= -\lambda X(x) \\ X(0) &= X'(l) \;=\; 0 \end{cases}$$

for which $\lambda_n = (n + \frac{1}{2})^2 \frac{\pi^2}{l^2}$.

6.2.3 Robin Boundary Conditions

If the boundary conditions are

$$\begin{cases} u_x(0,t) - a_0 u(0,t) &= 0 \\ u_x(l,t) + a_l u(l,t) &= 0. \end{cases}$$

A solution of the form (6.2) for the diffusion equation leads to

$$\begin{cases} X'' &= -\lambda X \\ X'(0) - a_0 X(0) &= 0 \\ X'(l) + a_l X(l) &= 0. \end{cases}$$

The solutions are again of the form

$$X(x) = C\cos(\beta x) + D\sin(\beta x)$$

with $\lambda = \beta^2 > 0$, the boundary conditions lead to $D = \frac{a_0 C}{\beta}$ and $(\beta^2 - a_l a_0)\tan(\beta l) = \beta(a_0 + a_l)$, which latter has to be solved for β assuming that $\beta^2 - a_0 a_l \neq 0$. There are different eigenvalues λ depending on the signs of $a_0 a_l$ and $a_0 + a_l$.

6.2.4 Sturm-Liouville Theory

The general problem

$$\begin{cases} \frac{d}{dx}(KX') = -(q + \lambda g)X, \quad 0 \leq x \leq l \\ a_0 X'(0) - b_0 X(0) = 0, \\ a_l X'(l) - b_l X(l) = 0 \end{cases}$$

is known as a *Sturm-Liouville*[2] *problem*, values λ for which there is a non-trivial solution are known as eigenvalues. In general there are no explicit solutions.

[2] After the French mathematician Joseph Liouville (1809-1882) and the Swiss mathematician François Sturm (1803-1855) who collaborated on the solution of boundary value problems.

6.3 Laplace's Equation

It was earlier observed that for Laplace in $\mathbb{R}^2$, $e^{kx}\left(e^{iky} \pm e^{-iky}\right)$ is a solution so $e^{kx}\cos(ky)$ and $e^{kx}\sin(ky)$ are both solutions, other solutions can be obtained interchanging the rôles of x and y.

Assume that a solution can be obtained by multiplying two functions of each one of the two variables only, *i.e.* it is separable, so: $u(x,y) = X(x)Y(y)$. Substitute this product into the PDE:

$$\frac{\partial^2 XY}{\partial x^2} + \frac{\partial^2 XY}{\partial y^2} = 0$$

As $Y(y)$ does not depend on x and $X(x)$ does not depend on y

$$YX'' + XY'' = 0$$

hence

$$\frac{X''}{X} + \frac{Y''}{Y} = 0.$$

The left term depends on x only, while the right term depends on y only. This is only possible if both of them are in fact constant and equal (but with opposite sign). Thus for some constant λ

$$\frac{X''}{X} = -\frac{Y''}{Y} = -\lambda$$

- An x-ODE : $X'' = -\lambda X$
- A y-ODE : $Y'' = \lambda Y$

Consider the boundary value problem

$$\left\{\begin{array}{rcl} u_{xx} + u_{yy} & = & 0, \quad \forall x, y \in (0,l), \\ u(x,0) & = & \phi(x), \quad \forall x \in [0,l] \\ u(x,l) & = & \psi(x), \quad \forall x \in [0,l] \\ u(0,y) = & 0 & = u(l,y), \quad \forall y \in [0,l]. \end{array}\right. \tag{6.7}$$

Considering the first equation and the boundary condition $u(0,y) = 0 = u(l,y)$ becomes $X(0) = 0 = X(l)$ assuming thta $Y(y)$ is not identically zero, it is clear that there are no non-trivial solutions for $\lambda \leq 0$, so assume $\lambda = \beta^2 > 0$.

A solution of the first equation is then

$$X(x) = A\cos(\beta x) + B\sin(\beta x)$$

applying the boundary conditions $X(0) = 0 = X(l)$, gives $A = 0$ and $\beta l = n\pi$. A solution of the second equation is then

$$Y(y) = Ce^{\beta y} + De^{-\beta y}$$

so a solution is

$$u(x,y) = \sin\frac{n\pi x}{l}\left(C_n e^{\frac{n\pi y}{l}} + D_n e^{-\frac{n\pi y}{l}}\right)$$

and hence the general solution is

$$u(x,y)=\sum_{n=1}^{+\infty}\sin\frac{n\pi x}{l}\left(C_n e^{\frac{n\pi y}{l}}+D_n e^{-\frac{n\pi y}{l}}\right)$$

The remaining boundary conditions give, for $y=0$

$$\phi(x)=\sum_{n=1}^{+\infty}\sin\frac{n\pi x}{l}\,(C_n+D_n)$$

so

$$C_n+D_n=\frac{2}{l}\int_0^l\phi(x)\sin\frac{n\pi x}{l}\,dx.$$

and for $y=l$

$$C_n e^{n\pi}+D_n e^{-n\pi}=\frac{2}{l}\int_0^l\psi(x)\sin\frac{n\pi x}{l}\,dx.$$

from which it is possible to determine C_n and D_n.

6.3.1 Boundary Conditions

If the boundary conditions are changed to:

$$\begin{cases} u(0,y) &= \phi(y), \quad \forall y\in[0,l]\\ u(l,y) &= \psi(y), \quad \forall y\in[0,l]\\ u(x,0)= & 0 \quad = u(x,0), \quad \forall x\in[0,l]. \end{cases}$$

The boundary conditions $u(x,0)=0=u(x,0)$ lead to $Y(0)=0=Y(l)$ so there are no non-trivial solutions for $\lambda\geq 0$, so assume $\lambda=-\beta^2<0$. Then

$$Y(y)=A\cos(\beta y)+B\sin(\beta y)$$

and

$$X(x))=Ce^{\beta x}+De^{-\beta x}$$

and the solution is found in terms of a Fourier series for y. If the boundary conditions are

$$\begin{cases} u(x,0) &= \phi_0(x), \quad \forall x\in[0,l]\\ u(x,l) &= \phi_1(x), \quad \forall x\in[0,l]\\ u(0,y) &= \psi_0(y), \quad \forall y\in[0,l]\\ u(l,y) &= \psi_1(y), \quad \forall y\in[0,l] \end{cases}$$

then as the problem is linear we seek $u(x,y)=v(x,y)+w(x,y)$ such that $\nabla^2 v=0$, $\nabla^2 w=0$ and

$$\begin{cases} v(x,0) = \phi_0(x), & w(x,0) = 0, & \forall x\in[0,l]\\ v(x,l) = \phi_1(x), & w(x,l) = 0, & \forall x\in[0,l]\\ v(0,y) = 0, & w(0,y) = \psi_0(y), & \forall y\in[0,l]\\ v(l,y) = 0, & w(l,y) = \psi_1(y), & \forall y\in[0,l] \end{cases}$$

Exercises

Exercise 6.1 *What is the solution of the boundary value problem*

$$\left\{\begin{array}{rcl} u_{xx}+u_{yy} &=& 0, \quad \forall x,y\in(0,1),\\ u(x,0) &=& 0, \quad \forall x\in[0,1]\\ u(x,1) &=& \sin(\pi x), \quad \forall x\in[0,1]\\ u(0,y) &=& 0 \quad \forall y\in[0,1]\\ u(1,y) &=& y(1-y) \quad \forall y\in[0,1] \end{array}\right. \tag{6.8}$$

Exercise 6.2 Derive *the solution of the boundary value problem*

$$\left\{\begin{array}{rcl} u_{xx}+u_{yy} &=& 0, \quad \forall x,y\in(0,1),\\ u(x,0) &=& x, \quad \forall x\in[0,1]\\ u(x,1) &=& x+\sin(\pi x), \quad \forall x\in[0,1]\\ u(0,y) &=& 0 \quad \forall y\in[0,1]\\ u(1,y) &=& 1 \quad \forall y\in[0,1] \end{array}\right. \tag{6.9}$$

Exercise 6.3 *What is the solution of the b.v.p.*

$$\begin{array}{rcl} \nabla^2 u &=& 2, \quad x,y\in(0,1)\\ u(x,0)=u(x,1) &=& 0, \quad x\in[0,1]\\ u(0,y)=u(1,y) &=& 0, \quad y\in[0,1] \end{array} \tag{6.10}$$

Hint: The general solution of $\nabla^2 u = g$ is $u = u_1 + u_2$ where u_1 is a particular solution and u_2 is the general solution of $\nabla^2 u = 0$.

Chapter 7

Fourier Transforms and PDEs

7.1 Diffusion Equation

Consider the initial value problem

$$\left\{\begin{array}{rcl} u_t &=& \kappa u_{xx} \\ u|_{t=0} &=& \phi(x) \qquad (\lim_{|x|\to\infty}\phi(x)=0) \end{array}\right. \tag{7.1}$$

It was shown that the unique solution for this problem is

$$\begin{array}{rcl} u(x,t) &=& \int_{-\infty}^{\infty} S(x-y,t)\phi(y)dy \qquad \forall t>0 \\ &=& S(t)*\phi(x) \end{array}$$

where

$$S(x,t)=\frac{1}{2\sqrt{\pi\kappa t}}e^{-\frac{|x|^2}{4\kappa t}} \qquad \forall t>0$$

and $S(x,t)$ was first derived by solving the IVP

$$\left\{\begin{array}{rcl} Q_t &=& \kappa Q_{xx} \\ Q|_{t=0} &=& q(x)=\left\{\begin{array}{l} 1 \text{ for } x>0 \\ 0 \text{ for } x<0 \end{array}\right. \end{array}\right. \tag{7.2}$$

then $S=\frac{\partial Q}{\partial x}$ and

$$\left\{\begin{array}{rcl} S_t &=& \kappa S_{xx} \\ S|_{t=0} &=& \delta_0 \end{array}\right. \tag{7.3}$$

Now start from (7.3) and construct S using *Fourier Transforms.* The Fourier transform is in the variable x so

$$\mathcal{F}S_t=(\mathcal{F}S)_t=(\hat{S})_t$$

so from (7.2)

$$\left\{\begin{array}{rcl} \hat{S}_t &=& \kappa\widehat{S_{xx}} \\ \hat{S}|_{t=0} &=& \hat{\delta}_0\equiv 1 \end{array}\right.$$

and

$$\widehat{S_{xx}}(\xi)=(i\xi)^2\hat{S}(\xi)=-\xi^2\hat{S}(\xi)$$

so

$$\begin{cases} \hat{S}_t &= \kappa\xi^2\hat{S} \\ \hat{S}|_{t=0} &= \hat{\delta}_0 \equiv 1 \end{cases}$$

For fixed ξ this is an ODE with solution

$$\hat{S}(\xi,t) = e^{-kt\xi^2}$$

as

$$\sqrt{\frac{\pi}{a}}e^{-\frac{1}{4a}\xi^2} = \widehat{e^{-x^2a}}(\xi)$$

hence take $a = (4\kappa t)^{-1}$ then as before

$$S(x,t) = \frac{1}{2\sqrt{\pi\kappa t}}e^{-\frac{|x|^2}{4\kappa t}} \qquad \forall t > 0$$

7.2 Wave Equation

$$\begin{cases} u_{tt} - c^2u_{xx} = 0 \\ u(x,0) = \quad u_t(x,0) = \phi(x) \end{cases}$$

The solution was shown to be

$$u(x,y) = S(t) * \phi(x)$$

where

$$S = \begin{cases} 0 & |x| > ct \\ \frac{1}{2c} & |x| < ct \end{cases}$$

Now start from

$$\begin{cases} S_{tt} - c^2S_{xx} &= 0 \\ S|_{t=0} = 0 \quad S_t|_{t=0} &= \delta_0 \end{cases}$$

and apply Fourier Transforms to give

$$\begin{cases} \hat{S}_{tt} - c^2\widehat{S_{xx}} &= 0 \\ \hat{S}|_{t=0} = 0 \quad \hat{S}_t|_{t=0} &= 1 \end{cases}$$

The differential equation leads to

$$\hat{S}_{tt} + c^2\xi^2\hat{S} = 0$$

which is an ODE with solutions of the form $\hat{S}(\xi,t) = A\sin(c\xi t)$ from the initial conditions $A = \frac{1}{c\xi}$ so

$$\hat{S}(\xi,t) = \frac{1}{c\xi}\sin(c\xi t) = \frac{1}{2ic\xi}\left(e^{ic\xi t} - e^{-ic\xi t}\right)$$

hence

$$S(x,t) = \begin{cases} 0 & |x| > ct \\ \frac{1}{2c} & |x| < ct \end{cases}$$

7.3 Laplace's Equation

To define fundamental solutions, search for functions that solve in the half-plane:

$$\left\{\begin{array}{rcll} u_{xx}+u_{yy} & = & 0 & y>0 \\ u(x,0) & = & \delta(x) & y=0 \end{array}\right. \tag{7.4}$$

Take a Fourier Transform w.r.t. x, define

$$u(\xi,y)=\mathcal{F}_x(u(x,y))(\xi)=\int_{-\infty}^{\infty} e^{-ix\xi}u(x,y)dx$$

then

$$\begin{array}{rcl} u_{yy}(\xi,y) & = & \partial_{yy}\int_{-\infty}^{\infty} e^{-ix\xi}u(x,y)dx \\ & = & \int_{-\infty}^{\infty} e^{-ix\xi}u_{yy}(x,y) \\ & = & \mathcal{F}_x(u_{yy})(\xi) \end{array}$$

and the equation becomes

$$\left\{\begin{array}{rcl} -\xi^2u(\xi,y)+u_{yy}(\xi,y) & = & 0 \\ u(\xi,0) & = & 1 \end{array}\right.$$

For fixed ξ this is an ODE that can be solved w.r.t. y as

$$u(y)=ce^{\alpha y} \qquad \Rightarrow \qquad u_{yy}=c\alpha^2e^{\alpha y}$$

thus

$$-\xi^2e^{\alpha y}+\alpha^2e^{\alpha y}=0 \quad \Rightarrow \quad \alpha^2=\xi^2 \quad \Rightarrow \quad \alpha=\pm\xi$$

thus there are two possible solutions $u_1=ce^{-|\xi|y}$ and $u_2=ce^{|\xi|y}$ and $u_1(0)=u_2(0)=0\Rightarrow c=1$ but $u_2\to\infty$ so it is not possible and hence

$$\begin{array}{rcl} u(x,y) & = & \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ix\xi}e^{-|\xi|y}d\xi \\ & = & \frac{1}{2\pi}\int_{0}^{\infty} e^{\xi(ix-y)}d\xi+\frac{1}{2\pi}\int_{-\infty}^{0} e^{\xi(ix+y)}d\xi \\ & = & \frac{1}{2\pi}\left(\left.\frac{e^{\xi(ix-y)}}{ix-y}\right|_0^\infty+\left.\frac{e^{\xi(ix+y)}}{ix+y}\right|_{-\infty}^0\right) \\ & = & \frac{1}{2\pi}\left(-\frac{1}{ix-y}+\frac{1}{ix+y}\right) \\ & = & \frac{1}{2\pi}\frac{2y}{x^2+y^2} \end{array}$$

So the fundamental solution in this case is

$$S(x,y)=\frac{y}{x^2+y^2}$$

Given the problem

$$\left\{\begin{array}{rcll} u_{xx}+u_{yy} & = & 0 & y>0 \\ u(x,0) & = & \phi(x) & y=0 \end{array}\right.$$

the solution is

$$u(x,y) = \int_{-\infty}^{\infty} S(x-z,y)\phi(z)dz$$

that is $u = S * \phi$.

No known problems in science or engineering lead to initial value problems for Laplace's equation. A well-posed problem for Laplace's equation on a simply connected open region Ω with a closed boundary Γ requires exactly one condition on the function value, the normal derivative or a combination of both at each point on Γ. If all or part of the boundary is at infinitiy, then the solution must be finite there.

Chapter 8

Distributions

If the pointwise definition of a function is replaced by a definition using locally averaged values then it is possible to define Dirac delta functions and to model impulses in mechanics and poles in electromagnetics.

Assume that $\mathcal{X} \subset \mathcal{H}$ where $\mathcal{H}$ is a complete inner product space (*Hilbert*[1] *Space*). For example, given an open set $\Omega \subset \mathbb{R}^n$, if $\mathcal{H} = \mathcal{L}_2(\Omega)$, $\mathcal{X} = C_0^\infty$ and the inner product is $(u,v) = \int_\Omega uv\,d\Omega$ then $\mathcal{X}$ is very much smaller than $\mathcal{H}$ and the *dual space* $\mathcal{X}'$, the *space of distributions on* Ω, is very much larger than $\mathcal{H}$.

A *test function* is any function $f \in C_0^\infty$ with compact support, for example

$$T(x) = \left\{ \begin{array}{ll} ke^{\left(-\frac{1}{1-|x|^2}\right)} & |x| < 1 \\ & \\ 0 & |x| \geq 1 \end{array} \right\} \quad \forall x \in \mathbb{R}^n$$

the constant k is chosen to normalise T. The space of test functions on Ω is denoted by $\mathcal{D}(\Omega)$. Point evaluations and integrals $I_1(f) = f(a)$ and $I_2 = \int_\Omega f\,d\Omega$ are examples are *continuous linear functionals* on $\mathcal{D}$. A continuous linear functional on $\mathcal{D}(\Omega)$ is called a *distribution on* Ω the space of distributions is denoted by $\mathcal{D}'(\Omega)$. The notations $I(f)$ and $\langle I, f\rangle$ will be use interchangeably. Note that the distribution $I(f) = f(a)\,\forall f \in \mathcal{D}$ defines the action of the *Dirac delta function* concentrated at the point a

$$\int_{\mathbb{R}} \delta(x-a) f(x) dx = f(a)$$

Two distributions I_1 and I_2 are equal if $I_1(f) = I_2(f)\,\forall f \in \mathcal{D}$. The *support of a distribution* I is the complement of the largest open set on which $I = 0$. Then, for example the support of the delta function is the point $\{0\}$. A *regular distribution* is defined is, given a locally integrable function f, $I_f(g) = \int_\Omega fg\,d\Omega\,\forall g \in \mathcal{D}$. If the function f has compact support S than the support of the distribution I_f is S. For a regular distribution I_f defined by a locally integrable function f the derivative is defined as the distribution generated by the derivative f' which can be defined formally using integration by parts

$$\langle \frac{dI_f}{dx}, g\rangle = \int_{\mathbb{R}} f' g\,dx = -\int_{\mathbb{R}} f g'\,dx$$

[1] David Hilbert (January 23, 1862, Königsberg, East Prussia (now Kaliningrad, Russia) - February 14, 1943, Göttingen, Germany) was a German mathematician

where it follows from the compactness of the support of f that the boundary terms are zero.

Example 21 *Let $H(x)$ be the* Heaviside *function*

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

so reversing the integration by parts

$$\begin{array}{rcl} \langle \frac{dI_H}{dx}, g\rangle & = & -\int_0^\infty g'\,dx \quad \forall g \in \mathcal{D} \\ & = & -g(x)|_0^\infty \\ & = & g(0) \\ & = & \int_{-\infty}^\infty \delta(x)g(x)\,dx \end{array}$$

so the derivative in the sense of distributions is

$$H'(x) = \delta(x)$$

Example 22 *Consider the function*

$$f(x) = \begin{cases} |x| & |x| < 1 \\ 0 & |x| < 1 \end{cases}$$

so reversing the integration by parts

$$\begin{array}{rcl} \langle \frac{dI_f}{dx}, g\rangle & = & -\int_0^1 xg'\,dx - \int_{-1}^0 (-x)g'\,dx \quad \forall g \in \mathcal{D} \\ & = & -(xg(x))|_0^1 + \int_0^1 g\,dx + (xg(x))|_{-1}^0 - \int_{-1}^0 g\,dx \\ & = & -g(1) + g(-1) + \int_0^1 g\,dx - \int_{-1}^0 g\,dx \\ & = & -\langle \delta(x-1), g(x)\rangle + \langle \delta(x+1), g(x)\rangle \\ & & \qquad + \int_0^1 f'g\,dx + \int_{-1}^0 f'g\,dx + \int_{|x|>1} f'gdx \int_0^\infty g'\,dx \end{array}$$

so the derivative in the sense of distributions is

$$\delta(x+1) - \delta(x-1) + f'(x)$$

where f' denotes the derivative in the classical sense, if it exists.

8.1 Sobolev Spaces

For $\Omega \subset \mathbb{R}^n$ define[2]

$$\mathcal{H}^1(\Omega) = \{u \in \mathcal{L}_2(\Omega) : \partial_{x_i} u \in \mathcal{L}_2(\Omega),\ i = 1, \ldots, n\}$$

where the derivative $\partial_{x_i} u$ is interpreted in the sense of distributions. The inner product on $\mathcal{H}^1(\Omega)$ is defined as

$$(u, v)_1 = \int_\Omega (uv + \nabla u \cdot \nabla v)\,d\Omega$$

[2] Sergei L'vovich Sobolev, born in St. Petersburg 6 October 1908 and died in Moscow 3 January 1989

Theorem 23 (Lemma) *$\mathcal{H}^1(\Omega)$ is a Hilbert space with norm $\|u\|_1 = (u,u)_1^{1/2}$*

The space $\mathcal{H}^1(\Omega)$ is an example of a *Sobolev space.*

Definition 24 *Dirichlet's Principle: The boundary value problem*[3]

$$\begin{cases} \Delta u - u &= 0 \quad \text{in } \Omega \\ u &= g \quad \text{in } \partial\Omega \end{cases}$$

has at most one solution and if a function u is a solution, it minimises the Energy *$E(u) \equiv \|u\|_1$ over all functions that satisfy the boundary condition.*

It is possible to define an inner product

$$(u,v)_m = \sum_{|\alpha|\leq m} \int_{\mathbb{R}^n} \partial^\alpha u \partial^\alpha v \; dx$$

then a norm

$$\|u\|_m = \left(\sum_{|\alpha|\leq m} \int_{\mathbb{R}^n} |\partial^\alpha u| \; dx \right)^{1/2}$$

This defines a Sobolev space $\mathcal{H}^m(\mathbb{R})$ for integer values m, but if the space is defined using Fourier transforms, then for any $s \in \mathbb{R}$ it is possible to define an inner product

$$(u,v)_s = \int (1+|\xi|^2)^s \hat{u}(\xi)\hat{v}(\xi) d\xi \tag{8.1}$$

and a norm

$$\|u\|_s = \left(\int (1+|\xi|^2)^s |\hat{u}(\xi)|^2 d\xi \right)^{1/2} \tag{8.2}$$

in order to define $\mathcal{H}^s(\mathbb{R})$. These two definitions are equivalent from Plancherel's theorem, because $\partial^\alpha u \in \mathcal{L}_2$ iff $\xi^\alpha \in \mathcal{L}_2$ and the observation that there exist constants c and C such that

$$c(1+|\xi|^2)^m \quad \leq \quad \sum_{|\alpha|\leq m} |\xi \alpha|^2 \quad \leq \quad C(1+|\xi|^2)^m$$

[3] The German mathematician Georg Friedrich Bernhard Riemann (November 17, 1826 – July 20, 1866) named this idea Dirichlet's principle in honour of his teacher Dirichlet. Johann Peter Gustav Lejeune Dirichlet (February 13, 1805 - May 5, 1859) was a German mathematician whose family originated in Richelet in Belgium

Chapter 9

Finite Difference Methods

9.1 Forward, backward and central differences

A forward difference is an expression of the form

$$D_+ f(x) = f(x+h) - f(x).$$

A backward difference arises when h is replaced by $-h$:

$$D_- f(x) = f(x) - f(x-h).$$

Finally, the central difference is the average of the forward and backward differences. It is given by

$$D_0 f(x) = \frac{D_+ f(x) + D_- f(x)}{2} = \frac{f(x+h) - f(x-h)}{2}.$$

The derivative of a function f at a point x is defined by the limit

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Assuming that f is continuously differentiable, it is possible to estimate the *truncation error* using *Taylor's series*[1]

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n$$

The Lagrange form of the remainder term states that there exists a number ξ between a and x such that

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}.$$

Then

$$\frac{D_+ f(x)}{h} - f'(x) = \mathcal{O}(h)$$

[1] Brook Taylor (August 18, 1685 - November 30, 1731), although the result was first discovered in 1671 by James Gregory

$$\frac{D_- f(x)}{h} - f'(x) = \mathcal{O}(h)$$

and

$$\frac{D_0 f(x)}{h} - f'(x) = \mathcal{O}(h^2)$$

Similarly the central difference approximation of the second derivative of f is

$$f''(x) \approx \frac{D_+ D_- f(x)}{h^2} = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}.$$

and

$$f''(x) - \frac{D_+ D_- f(x)}{h^2} = \mathcal{O}(h^2)$$

9.2 Some Finite Difference schemes

The original problem in terms of a continuous function $u(x,t)$ is replaced by a numerical problem in terms of a sequence of *discrete values* U_m^n at the mesh points $x_m,\ t_n$ such that $U_m^n \approx u(x_m, t_n)$. To begin with, assume that $t_{n+1} - t_n = k$, known as the *time step* is constant and that $x_{m+1} - x_m = h$, known as the *mesh length* is constant. The Difference operators D_+, *etc.* are used to denote differences with respect to the variable x only. The alternative notation $U(x,t)$ for the numerical solution at fixed values of x and t (using steps of size h and k) in order to emphasise the dependence on the values of x and t rather than the values of m and n the numbers of steps needed. Given a mesh size h and a time step k define a *mesh ratio* $r = \frac{k}{h}$. Then, given the p.d.e

$$\mathcal{L}u = f$$

for example $u_t + au_x = 0$, $U(x_m, t_n) \equiv U_m^n$ is the solution of a difference approximation of the form

$$\mathcal{L}_{hk} U = f_{hk}$$

where

$$\mathcal{L}_{h,k}(U(x,t)) \equiv Q_{-1} U(x, t+k) - \sum_{j=0}^{p} Q_j U(x, t-jk) \tag{9.1}$$

where $Q_n,\ n = -1, 0, \ldots, p$ are difference operators.

Forward Difference Scheme

$$\text{becomes} \quad \left.\begin{aligned} &\tfrac{1}{k}(U_m^{n+1} - U_m^n) + \tfrac{a}{h} D_+ U_m^n = 0 \\ &U_m^{n+1} = (1 - arD_+) U_m^n \end{aligned}\right\} \tag{9.2}$$

Backward Difference Scheme

$$\text{becomes} \quad \left.\begin{aligned} &\tfrac{1}{k}(U_m^{n+1} - U_m^n) + \tfrac{a}{h} D_- U_m^n = 0 \\ &U_m^{n+1} = (1 - arD_-) U_m^n \end{aligned}\right\} \tag{9.3}$$

Central Difference Scheme

$$\text{becomes}\quad \left.\begin{aligned} &\tfrac{1}{k}(U_m^{n+1} - U_m^n) + \tfrac{a}{h}D_0U_m^n = 0 \\ &U_m^{n+1} = (1 - arD_0)U_m^n \end{aligned}\right\} \tag{9.4}$$

Lax-Friedrichs Scheme

$$\text{becomes}\quad \left.\begin{aligned} &\tfrac{1}{k}(U_m^{n+1} - \tfrac{1}{2}(U_{m+1}^n + U_{m-1}^n)) + \tfrac{a}{h}D_0U_m^n = 0 \\ &U_m^{n+1} = \tfrac{1}{2}(U_{m+1}^n + U_{m-1}^n) - arD_0U_m^n \end{aligned}\right\} \tag{9.5}$$

Lax-Wendroff Scheme

$$\text{becomes}\quad \left.\begin{aligned} &\tfrac{1}{k}(U_m^{n+1} - U_m^n) + \tfrac{a}{h}D_0U_m^n - \tfrac{ak}{2h^2}D_+D_-U_m^n = 0 \\ &U_m^{n+1} = (1 - arD_0 + \tfrac{ar^2}{2}D_+D_-)U_m^n \end{aligned}\right\} \tag{9.6}$$

Leap-Frog Scheme

$$\text{becomes}\quad \left.\begin{aligned} &\tfrac{1}{2k}(U_m^{n+1} - U_m^{n-1}) + \tfrac{a}{h}D_0U_m^n = 0 \\ &U_m^{n+1} = U_m^{n-1} - 2arD_0U_m^n \end{aligned}\right\} \tag{9.7}$$

Example 25 *It is possible to apply the schemes to equations written in conservation form, thus given*

$$u_t + (f(u))_x = 0$$

the forward difference scheme would be

$$U_m^{n+1} = U_m^n - r(f(U_{m+1}^n) - f(U_m^n))$$

9.3 Truncation Error

Definition 26 *Truncation Error: If u is the solution of the problem $\mathcal{L}u = f$ and from (9.1) U is the solution of $\mathcal{L}_{hk}U = f_{hk}$ then if*

$$\|\mathcal{L}_{h,k}u(x,t) - f_{hk}\| \leq kC(t)\,(h^{q_1} + k^{q_2})$$

and

$$\|u(x,jk) - U(x,jk)\| \leq C_j\,(h^{q_1} + k^{q_2})$$

the method is accurate of order (q_1, q_2)

Alternatively,assuming that $f_{hk} = f(x,t)$ we could define the truncation error as $\mathcal{L}_{h,k}v - \mathcal{L}v$ for any function v.

Definition 27 *Consistency: Given that $u(x,t)$ a solution of a p.d.e $\mathcal{L}u = f$ and $U(x,t)$ is a solution of a finite difference scheme $\mathcal{L}_{hk}U = f_{hk}$ The scheme is consistent if $\mathcal{L}_{hk}u - f_{hk} \to 0$ as $h, k \to 0$.*

An alternative definition is that the scheme is consistent if, for any function v, $\mathcal{L}_{hk}v - \mathcal{L}v \to 0$ as $h, k \to 0$, assuming that $f_{hk} = f(x,t)$. If the approximation is accurate of order (q_1, q_2) and $q_1, q_2 > 0$ the method is *consistent*.

Forward Difference Scheme Using Taylor series, if $u_t + au_x = g(x,t)$ is denoted by $\mathcal{L}u = f$, then if the forward difference scheme is denoted by $\mathcal{L}_{hk}U = f_{hk}$ it should be written as

$$\frac{1}{k}(U_m^{n+1} - U_m^n) + \frac{a}{h}D_+U_m^n = g_m^n$$

or with $t = nk$ and $x = mh$,

$$\frac{1}{k}(U(x,t+k) - U(x,t)) + \frac{a}{h}D_+U(x,t) = g(x,t)$$

then the truncation error is

$$\begin{aligned}
&\tfrac{1}{k}(u(x,t+k) - u(x,t)) + \tfrac{a}{h}D_+u(x,t) - g(x,t)\\
&= \tfrac{1}{k}(u(x,t) + ku_t(x,t) + \tfrac{k^2}{2}u_{tt}(x,t) + \mathcal{O}(k^3) - u(x,t))\\
&\qquad + a\tfrac{1}{h}(u(x,t) + hu_x(x,t) + \tfrac{h^2}{2}u_{xx}(x,t) + \mathcal{O}(h^3) - u(x,t)) - g(x,t)\\
&= u_t(x,t) + \tfrac{k}{2}u_{tt}(x,t) + \mathcal{O}(k^2) + a(u_x(x,t) + \tfrac{h}{2}u_{xx}(x,t) + \mathcal{O}(h^2)) - g(x,t)\\
&= \tfrac{k}{2}u_{tt}(x,t) + \mathcal{O}(k^2) + a(\tfrac{h}{2}u_{xx}(x,t) + \mathcal{O}(h^2))\\
&= \mathcal{O}(k) + \mathcal{O}(h)
\end{aligned}$$

so the method is order (1,1) and hence it is consistent.

Example 28 *Lax-Wendroff Scheme: Consider the equation $u_t = au_x$ the Lax-Wendroff scheme is*

$$U(x,t+k) = \left(1 + raD_0 + \frac{a^2r^2}{2}D_+D_-\right)U(x,t)$$

if $u_t = au_x + f$ then with

$$g_{m,n} = \frac{1}{2}\left(f(x,t+k) + f(x,t)\right) - \frac{ak}{4h}\left(f(x+h,t) - f(x,t)\right)$$

the method is order (2,2), if $g_{m,n} = f(x,y)$ the scheme is only order (2,1).

9.4 Stability, Consistency and Convergence

9.4.1 Example of Instability

Consider the Cauchy problem

$$\left.\begin{array}{rcl} u_t &=& u_x, \quad t>0 \\ u(x,0) &=& f(x) \end{array}\right\} \quad -\infty < x < \infty \tag{9.8}$$

Using the centred difference scheme, (9.8) can be approximated by

$$\left.\begin{array}{rcl} \frac{1}{k}(U(x,t+k) - U(x,t)) &=& \frac{1}{2h}(U(x+h,t) - U(x-h,t) \\ && \qquad t = t_n = nk, n = 0,1,2,\dots \\ U(x,0) &=& f(x) \end{array}\right\} \begin{array}{l} x = x_m = mh, \\ m = 0, \pm1, \dots \end{array} \tag{9.9}$$

that is

$$U(x,t+k) = (1 + rD_0)U(x,t), \quad U(x,0) = f(x) \tag{9.10}$$

From the inverse transform formula

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

it follows that

$$U(x,t) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{i\xi x}\hat{U}(\xi,t)d\xi, \quad \hat{U}(\xi,0) = \hat{f}(\xi)$$

given that

$$D_0 e^{i\xi x} = 2i\sin(\xi h)e^{i\xi x}$$

it follows by applying the inverse formula to both sides of (9.10), that

$$\hat{U}(\xi,t+k) = (1 + ir\sin(\xi h))\hat{U}(\xi,t),$$

so

$$\hat{U}(\xi,t) = (1 + ir\sin(\xi h))^n\hat{f}(\xi), \quad t = nk$$

Hence

$$\hat{U}(\xi,t) = \kappa^n\hat{f}(\xi)$$

where there are frequencies ξ such that $|\kappa| > 1$ and the growth of those components in $\hat{U}$ is unbounded. Since from Plancherel's theorem $\|\hat{U}(\xi,t)\| = \|U(x,t)\|$, it follows that the growth in U is unbounded so the solution is unstable.

A stable difference formula is obtained if (9.10) is replaced by the Lax-Friedrichs scheme

$$U(x,t+k) = \frac{1}{2}(U(x+h,t) + U(x-h,t)) + rD_0U(x,t), \quad U(x,0) = f(x) \tag{9.11}$$

assuming $0 < r_0 \leq r = k/h \leq 1$. By analogy with the previous results

$$\hat{U}(\xi,t) = (\cos(\xi h) + ir\sin(\xi h))^n\hat{f}(\xi), \quad t = nk$$

it is obvious that

$$|\kappa| = |\cos(\xi h) + ir\sin(\xi h)| \leq 1$$

and the scheme is stable.

9.4.2 Stability

Given that $u(x,t)$ a solution of a p.d.e $\mathcal{L}u = f$ and $U(x,t)$ is a solution of a finite difference scheme $\mathcal{L}_{hk}U = f_{hk}$.

Definition 29 *Stability: A difference approximation of the form*

$$\mathcal{L}_{h,k}(U(x,t)) \equiv Q_{-1}U(x,t+k) - \sum_{j=0}^{p} Q_jU(x,t-jk) = 0 \tag{9.12}$$

where Q_n, $n = -1, 0, \ldots, p$ are difference operators, is stable as $h, k \to 0$ if

$$\|\hat{U}(,t)\| \leq Ke^{\alpha(t-t_0)}\|\hat{U}(,t_0)\|$$

A more readily accessible test is derived from Theorem 30.

Theorem 30 *von Neumann condition: A necessary condition for the stability of the difference approximation is that (9.12) has no exponentially growing solutions, that is all solutions of*

$$det\left(\hat{Q}_{-1}\kappa^{p+1} - \sum_{j=0}^{p} \hat{Q}_j \kappa^j\right) = 0 \tag{9.13}$$

must satisfy $|\kappa| \leq 1$ *(this is not a sufficient condition).*

In a simple application of this von Neumann (Fourier) method it can be assumed that the solution can be expressed as a Fourier series, that is the sum of components such as

$$U(x,t) = \kappa^{t/k} e^{i\xi x}, \qquad x = mh,\ t = nk$$

for different frequencies ξ and the scheme is stable if it only permits solutions with $|\kappa| \leq 1$. In particular $U_m^{n+1} = \kappa U_m^n$ and κ is known as the *amplification factor* and $U_{m+1}^n = e^{i\theta} U_m^n$ where $\theta = h\xi$.

Forward Difference Scheme

$$U_m^{n+1} = (1 - arD_+)U_m^n$$

leads to

$$\begin{aligned} \kappa U_m^n &= U_m^n - ar(e^{i\theta}U_m^n - U_m^n) \\ \kappa &= 1 - ar(e^{i\theta} - 1) \\ &= 1 + ar(1 - \cos\theta) - iar\sin\theta \\ |\kappa|^2 &= (1 + 2ar\sin^2\tfrac{\theta}{2})^2 + a^2r^2\sin^2\theta \\ &= 1 + 4ar\sin^2\tfrac{\theta}{2} + a^2r^2(\sin^2\theta + 4\sin^4\tfrac{\theta}{2}) \\ &= 1 + 4ar\sin^2\tfrac{\theta}{2} + 4a^2r^2\sin^2\tfrac{\theta}{2} \\ &= 1 + 4ar\sin^2\tfrac{\theta}{2}(ar + 1) \end{aligned}$$

so

$$\begin{array}{ccccc} a > 0 & \Rightarrow & |\kappa| > 1 & \Rightarrow & \text{unstable} \\ a < 0, ar + 1 > 0 & \Rightarrow & |\kappa| < 1 & \Rightarrow & \text{stable if } r < \frac{1}{|a|} \end{array}$$

So *upwind schemes* are *conditionally stable* while downwind schemes are always unstable.

Central Difference Scheme

$$U_m^{n+1} = (1 - arD_0)U_m^n$$

leads to

$$\begin{aligned} \kappa U_m^n &= U_m^n - \tfrac{ar}{2}(e^{i\theta}U_m^n - e^{-i\theta})U_m^n \\ \kappa &= 1 - \tfrac{ar}{2}(e^{i\theta} - e^{-i\theta}) \\ &= 1 + iar\sin\theta \\ |\kappa|^2 &= 1 + a^2r^2\sin^2\theta \\ &\geq 1 \end{aligned}$$

So the centred scheme is always unstable.

Lax-Friedrichs Scheme

$$U_m^{n+1} = \frac{1}{2}(U_{m+1}^n + U_{m-1}^n) - arD_0U_m^n$$

leads to

$$\begin{aligned}
\kappa U_m^n &= \tfrac{1}{2}(e^{i\theta}U_m^n + e^{-i\theta})U_m^n - \tfrac{ar}{2}(e^{i\theta}U_m^n - e^{-i\theta})U_m^n \\
\kappa &= \tfrac{1}{2}(e^{i\theta}U_m^n + e^{-i\theta}) - \tfrac{ar}{2}(e^{i\theta} - e^{-i\theta}) \\
&= \cos\theta + iar\sin\theta \\
|\kappa|^2 &= \cos^2\theta + a^2r^2\sin^2\theta \\
&< 1 \qquad \text{if } ar < 1
\end{aligned}$$

So the method is conditionally stable and the condition is the same as for the upwind scheme, namely $r < \frac{1}{a}$.

Leap-Frog Scheme

$$U_m^{n+1} = U_m^{n-1} - 2arD_0U_m^n$$

Leap-Frog is a two step method given $U_m^{n+1} = \kappa U_m^n$ it is also true that $U_m^n = \kappa U_m^{n-1}$ and hence $U_m^{n+1} = \kappa^2 U_m^{n-1}$ which leads to

$$\begin{aligned}
\kappa^2 U_m^{n-1} &= U_m^{n-1} - ar(\kappa e^{i\theta}U_m^{n-1} - \kappa e^{-i\theta}U_m^{n-1}) \\
\kappa^2 &= 1 - arg(e^{i\theta} - e^{-i\theta}) \\
\kappa^2 &= 1 - 2ar\kappa i\sin\theta
\end{aligned}$$

roots of $\kappa^2 + 2ar\kappa i\sin\theta - 1 = 0$ are $\kappa = -ari\sin\theta \pm \sqrt{-a^2r^2\sin^2\theta + 1}$ assuming $ar < 1$ it follows that $|\kappa| = 1$ so the method is stable and there is no decay in the solution.

Definition 31 *Convergence If for any $u(x,t)$ such that $U(x,0)$ converges to $u(x,0)$ as $h,k \to 0$ it follows that $U(x,t)$ converges to $u(x,t)$ for all (x,t) as $h,k \to 0$ then the scheme is convergent.*

In general, convergence is a difficult property to test directly, but convergence is associated with stability for which there is a more straightforward test.

Theorem 32 *Lax-Richtmeyer Equivalence Theorem: Given a properly posed linear initial-value problem and a linear finite-difference approximation to it that satisfies the consistency condition, stability is a necessary and sufficient condition for convergence*

9.4.3 Dissipation and Dispersion

Definition 33 *Dissipation: If all solutions of (9.13) satisfy*

$$|\kappa| \le 1 - \delta|\xi h|^p, \quad 0 \le |\xi h| \le \pi$$

for some constant δ then p is called the order of dissipation.

Example 34 *Leap-Frog: Consider the hyperbolic equation $u_t = au_x$, the leap-frog scheme is*

$$U(x, t+k) = U(x, t-k) + 2raD_0U(x, t)$$

The order is $(2, 2)$ and the characteristic equation is

$$(\kappa^2 - 1) - 2\kappa ira \sin(\xi h) = 0$$

or

$$\kappa = ira \sin(\xi h) \pm \sqrt{1 - r^2a^2 \sin^2(\xi h)}$$

so $|\kappa| = 1$ for $ra \leq 1$ and so the approximation is not dissipative.

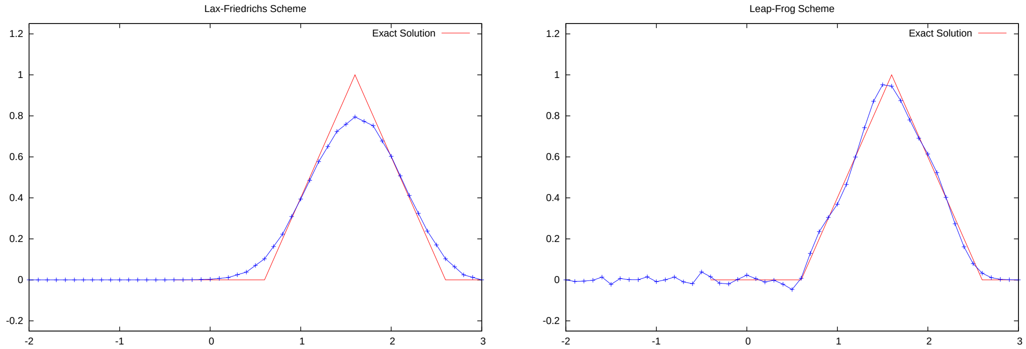
The Lax-Friedrichs[2] scheme is

$$U(x, t+k) = \frac{1}{2}(U(x+h, t) + U(x-h, t)) + kaD_0U(x, t)$$

the characteristic equation is

$$\kappa = \cos(\xi h) + ira \sin(\xi h)$$

this is not dissipative as $|\kappa| = 1$ when $\xi h = \pi$.



(a) Lax-Friedrichs

(b) Leap-Frog

Figure 9.1: Solution of $u_t + u_x = 0$ at $t = 0.8$ with $r = 0.8$

The solution of the problem

$$\left.\begin{array}{rcl} u_t + au_x & = & 0, \quad t > 0 \\ u(x, 0) & = & f(x) \end{array}\right\} \quad -\infty < x < \infty \tag{9.14}$$

satisfies

$$\left.\begin{array}{rcl} \hat{u}_t + i\xi a\hat{u} & = & 0 \\ \hat{u}(\xi, 0) & = & \hat{f} \end{array}\right\} \quad \Rightarrow \quad \hat{u} = e^{-i\xi at}\hat{f}$$

[2] Peter David Lax (born May 1, 1926, Budapest), his PhD supervisor was Kurt Otto Friedrichs (born 28 Sept 1901 in Kiel, died 31 Dec 1982 in New Rochelle, New York)

so in particular

$$\hat{u}(\xi, t+k) = e^{-i\xi ak}\hat{u}(\xi, t)$$

In general a difference solution satisfies

$$\hat{U}(\xi, t) = g(\xi k)^n \hat{f}(\xi)$$

so

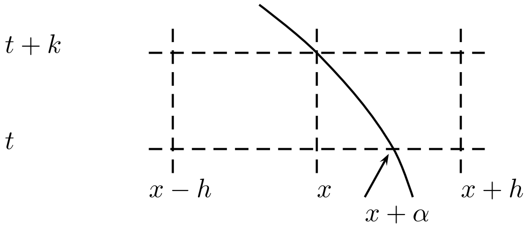
$$\begin{aligned}\hat{U}(\xi, t+k) &= g(\xi ah)\hat{U}(\xi, t)\\ &= |g(\xi ah)|e^{-i\xi\alpha(\xi h)k}\hat{U}(\xi, t)\end{aligned}$$

If $\alpha = a$ for all frequencies ξ then all frequencies are propagated with the correct speed.

Definition 35 *Dispersion: Dispersion is the phenomenon whereby different frenquencies are propagated at different speeds. The quantity $\alpha(\xi h)$ is known as the* phase speed.

9.4.4 Courant Friedrichs Lewy (CFL) Condition

Assume that a first order hyperbolic differential equation has been approximated by a difference equation so that $U(x, t+k)$ is computed in terms of $U(x, t)$, $U(x+h, t)$ and $U(x-h, t)$ in addition assume that the characteristic through $(x, t+k)$ passes through the point $(x+\alpha, t)$[3].



Then the *domain of dependence* (c.f. Definition 17) of the difference solution must include the domain of dependence of the differential equation, if not, the solution U *cannot converge* to u as $h, k \to 0$. Thus the point $(x+\alpha, t)$ must lie between $U(x+h, t)$ and $U(x-h, t)$. Given the differential equation $u_t = au_x$ the characteristics are $x + at =$ constant so the characteristic through $(x, t+k)$ passes through $(x+ak, t)$ so the *CFL condition*[1] is

$$|ak| \leq h \quad \Rightarrow \quad |a|r \leq 1$$

where $r = \frac{k}{h}$. This shows that for $a > 0$ the forward difference scheme

$$U(x, t+k) = (1 - ar)U(x, t) + arU(x+h, t)$$

will converge but

$$U(x, t+k) = (1 + ar)U(x, t) - arU(x-h, t)$$

will converge if $a < 0$. In each case the scheme that will converge is known as an *upwind scheme.*

[3] Hans Lewy born 20 Oct 1904 in Breslau, died 23 Aug 1988 in Berkeley

9.4.5 Stability

The CFL condition is concerned with convergence but it is closely associated with stability problems for the difference schemes. Again consider the first order equation

$$u_t = au_x \tag{9.15}$$

The coefficient a is the propagation velocity and is assumed constant. Then discretising the time derivative using a forward difference and expanding by Taylor's series (centred on x, t) gives

$$u_t = \frac{1}{k}\left(u(x, t+k) - u(x, t)\right) - \frac{k}{2}u_{tt} + \mathcal{O}(k^2) \tag{9.16}$$

using a central difference for the space derivative gives

$$u_x = \frac{1}{2h}\left(u(x+h, t) - u(x-h, t)\right) + \mathcal{O}(h^2) \tag{9.17}$$

Substituting into (9.15) gives

$$U(x, t+k) = U(x, t) + \frac{ak}{2h}\left(U(x+h, t) - U(x-h, t)\right) \tag{9.18}$$

this is first order accurate (in time), but substituting back with (9.16), shows that the difference equation (9.17) is a $\mathcal{O}(k^2 + h^2)$ replacement (*i.e.* order (2,2)) of

$$u_t - au_x = -\frac{k}{2}u_{tt} \tag{9.19}$$

from (9.15) this is equivalent to

$$u_t - au_x = -\frac{k}{2}a^2 u_{xx} \tag{9.20}$$

as the diffusion coefficient in (9.20) is negative this means that the solution is unstable. This explains the derivation of the Lax-Wendroff scheme that incorporates a difference approximation to the term $\frac{k}{2}a^2u_{xx}$. If the centred difference (9.17) is replaced by either a forward difference which, expanding by Taylor's series gives

$$u_x = \frac{1}{h}(u(x+h, t) - u(x, t)) - \frac{h}{2}u_{xx} + \mathcal{O}(h^2) \tag{9.21}$$

or backward difference which, expanding by Taylor's series about (x, t), gives

$$u_x = \frac{1}{h}(u(x, t) - u(x-h, t)) + \frac{h}{2}u_{xx} + \mathcal{O}(h^2) \tag{9.22}$$

These give respectively

$$U(x, t+k) = U(x, t) + a\frac{k}{h}\left(U(x+h, t) - U(x, t)\right) \tag{9.23}$$

and

$$U(x, t+k) = U(x, t) + a\frac{k}{h}\left(U(x, t) - U(x-h, t)\right) \tag{9.24}$$

as second order replacements of respectively

$$u_t - au_x = \frac{a}{2}\,(h - ak)\,u_{xx} \tag{9.25}$$

and

$$u_t - au_x = -\frac{a}{2}\,(h + ak)\,u_{xx} \tag{9.26}$$

Thus for stability, with positive diffusion, the forward difference (9.23) is needed when a is positive and the backward difference (9.24) when a is negative (*i.e.* upwind or upstream differences), in addition to the CFL condition $|a|k \leq h$.

9.5 Finite Volume Methods

Consider the equation

$$u_t + u_x = 0,$$

define

$$\overline{u}(t) = \frac{1}{h}\int_x^{x+h} u(x,t)\;dt$$

then

$$\int_x^{x+h}\int_t^{t+k} (u_t + u_x)\;dtdx = \int_x^{x+h}\int_t^{t+k} u_t\;dtdx\;+\;\int_t^{t+k}\int_x^{x+h} u_x\;dxdt$$

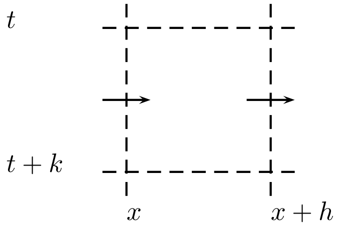
and so

$$\int_x^{x+h} (u(x,t+k) - u(x,t))\;dx\;+\;\int_t^{t+k} (u(x+h,t) - u(x,t))\;dt\;=\;0$$

which can be written as

$$\overline{u}(t+k) = \overline{u}(t) - \int_t^{t+k} u(x+h,t)\;dt + \int_t^{t+k} u(x,t)\;dt$$

the integrals on the right represent flow across the boundaries of the cell (the finite volume), for further details see for example[6].



9.6 Parabolic Equations

Consider the initial boundary value problem

$$\left\{\begin{array}{rcl} u_t &=& u_{xx} \quad 0 < x < l,\ t > 0 \\ u(x,0) &=& f(x) \quad 0 < x < l \\ u(0,t) = g_0, && u(l,t) = g_1 \quad t > 0 \end{array}\right. \tag{9.27}$$

Parabolic equations typically model heat conduction or diffusion processes, so if there is no input into the model, the typical solution exhibits an exponential decay. The proof is not restricted to one space dimension so rewrite (9.27) as

$$\left\{\begin{array}{rcl} u_t &=& \nabla^2 u \quad \text{in } \Omega; t > 0 \\ u &=& f \quad \text{in } \Omega; t = 0 \\ u &=& g \quad \text{on } \Gamma; t > 0 \end{array}\right. \tag{9.28}$$

The space $\mathcal{L}_2(\Omega)$ is defined with the inner product

$$(u,v) = \int_\Omega uv\ d\Omega$$

and hence the norm

$$\|u\| = (u,u)^{1/2} = \left(\int_\Omega u^2\ d\Omega\right)^{1/2}.$$

Theorem 36 *Stability: . Given that u is the solution of (9.28), it follows that if $g = 0$ then*

$$\frac{d}{dt}\|u\| \ < \ 0$$

and hence for any $t \geq 0$, $\|u(.,t)\| \leq \|f\|$.

Proof

For any fixed $t > 0$, using the divergence theorem

$$\begin{array}{rcl} u_t &=& \nabla^2 u \quad \text{in } \Omega, \\ u_t u &=& u\nabla^2 u \quad \text{in } \Omega, \\ \int_\Omega u_t u\ d\Omega &=& \int_\Omega u\nabla^2 u\ d\Omega, \\ \int_\Omega u_t u\ d\Omega &=& -\int_\Omega |\nabla u|^2\ d\Omega, \\ \frac{1}{2}\frac{d}{dt}\int_\Omega u^2\ d\Omega &=& -\int_\Omega |\nabla u|^2\ d\Omega \quad < 0 \end{array}$$

Hence $\|u(.,t)\|$ is a decreasing function of t so

$$\|u(.,t)\| < \|u(.,0)\| = \|f\|$$

Q.E.D.

There are two standard ways of investigating the stability of finite difference approximations

- The matrix method where the scheme can be written as

$$\boldsymbol{U}(t+k) = A\boldsymbol{U}(t) + \boldsymbol{b}$$

  and the stability depends on the eigenvalues of A.

- The Fourier Method (a.k.a. The von Neumann method[4]) is easier to apply, but from the von Neumann condition (Theorem 30), in general it only provides a necessary condition for stability.

In either case, assume that the finite difference equations are solved for two different initial conditions $f_1(x)$ and $f_2(x)$ to give two different solutions U_1 and U_2 the scheme is stable if the difference $w = U_1 - U_2$ remains bounded. The difference w satisfies the same finite difference equations as the solutions U_1 and U_2 so

$$\boldsymbol{w}(t+k) = A\boldsymbol{w}(t)$$

and it remains bounded if $\|A\| \leq 1$.

9.6.1 Crank-Nicolson

A simple *explicit* method for the diffusion equation in (9.27) can be written as

$$\begin{aligned} U(x,t+k) &= U(x,t) + rD_+D_-U(x,t) \\ &= rU(x-h,t) + (1-2r)U(x,y) + rU(x+h,t) \end{aligned}$$

where now $r = \frac{k}{h^2}$. The Crank-Nicolson method[2] for (9.27) is *implicit* and can be written as

$$U(x,t+k) = U(x,t) + \tfrac{r}{2}D_+D_-\left(U(x,t+k) + U(x,t)\right)$$

$$\begin{aligned} -\tfrac{r}{2}U(x-h,t+k) + (1+r)U(x,t+k) - \tfrac{r}{2}U(x+h,t+k) \\ = \tfrac{r}{2}U(x-h,t) + (1-r)U(x,t) + \tfrac{r}{2}U(x+h,t) \end{aligned}$$

Substituting w for U in the explicit scheme gives

$$\begin{aligned} w(x,t+k) &= rw(x-h,t) + (1-2r)w(x,t) + rw(x+h,t) \\ \kappa^{n+1}e^{i\xi x} &= r\kappa^n e^{i\xi(x-h)} + (1-2r)\kappa^n e^{i\xi x} + r\kappa^n e^{i\xi(x+h)} \\ \kappa &= re^{-i\xi h} + (1-2r) + re^{i\xi h} \\ &= 1 + 2r\left(\cos(\xi h) - 1\right) \\ &= 1 - 4r\sin^2\left(\tfrac{\xi h}{2}\right) \end{aligned}$$

hence the method is stable for all frequencies such that

$$-1 \leq 1 - 4r\sin^2\left(\frac{\xi h}{2}\right)$$

[4] John von Neumann first used the method during world war II

and is stable for all frequencies if

$$r \leq \frac{1}{2}$$

In order to apply the matrix method it is necessary to write out the full set of difference equations, including the boundary conditions, in the form $\boldsymbol{U}(t+k) = A\boldsymbol{U}(t) + \boldsymbol{b}$, for problem (9.27) it follows that the explicit scheme with the mesh length $h = l/m$ leads to the $(m-1) \times (m-1)$ matrix

$$A = \begin{pmatrix} 1-2r & r & & & \\ r & 1-2r & r & & \\ & \ddots & \ddots & \ddots & \\ & & r & 1-2r & r \\ & & & r & 1-2r \end{pmatrix}$$

with

$$\boldsymbol{b} = \begin{pmatrix} rU(0,t) \\ 0 \\ \vdots \\ 0 \\ rU(mh,t) \end{pmatrix} \qquad \boldsymbol{U}(t) = \begin{pmatrix} U(h,t) \\ U(2h,t) \\ \vdots \\ U((m-2)h,t) \\ U((m-1)h,t) \end{pmatrix}$$

where $U(0,t)$ and $U(mh,t) \equiv U(l,t)$ are defined by the boundary conditions in (9.27). It is known[5] that the eigenvalues of the $(m-1) \times (m-1)$ matrix

$$A = \begin{pmatrix} a & b & & & \\ c & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & c & a & b \\ & & & c & a \end{pmatrix}$$

are

$$\lambda_s = a + 2\sqrt{bc}\cos\left(\frac{s\pi}{m}\right)$$

where a, b and c may be real or complex. With $h = \frac{l}{m}$ by writing $\xi = \frac{s\pi}{l}$ the condition $\lambda \leq 1$ gives the same condition as for the Fourier method which is necessary and sufficient in this case. for Crank-Nicolson the equation is

$$A_1\boldsymbol{U}(t+k) = A_0\boldsymbol{U}(t) + \boldsymbol{b}$$

where

$$A_0 = \begin{pmatrix} 1-r & \frac{r}{2} & & & \\ \frac{r}{2} & 1-r & \frac{r}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{r}{2} & 1-r & \frac{r}{2} \\ & & & \frac{r}{2} & 1-r \end{pmatrix} \quad \text{and} \quad A_1 = \begin{pmatrix} 1+r & -\frac{r}{2} & & & \\ -\frac{r}{2} & 1+r & -\frac{r}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{r}{2} & 1+r & -\frac{r}{2} \\ & & & -\frac{r}{2} & 1+r \end{pmatrix}$$

[5] A proof is given in [8] and the proof of the eigenvalues for more general symmetric tridiagonal matrices is given in [13].

and $A = A_1^{-1}A_0$ as A_0 and A_1 have the same eigenvectors, the eigenvalues of A are

$$\lambda_s = \frac{1 - 2r\sin^2\left(\frac{\xi h}{2}\right)}{1 + 2r\sin^2\left(\frac{\xi h}{2}\right)}$$

which clearly satisfies $\lambda_s \leq 1$ for all ξ and all $r > 0$. So the stability of the Crank-Nicolson method is unrestricted. For other boundary conditions it may be necessary to bound the eigenvalues using:

Theorem 37 *Gershgorin's first theorem: The largest of the moduli of the eigenvalues cannot exceed the largest sum of the moduli of the elements in any row or column.*

Theorem 38 *Gershgorin's circle theorem: Let P_s be the sum of the moduli of the elements along the s-th row, excluding the diagonal a_{ss}. Then each eigenvalue lies inside or on at least one of the circles $|\lambda - a_{ss}| = P_s$.*

.

9.7 Elliptic Equations

Any numerical solution must preserve the property that the solution depends continuously on the boundary data and cannot have maxima or minima at interior points (*i.e.* Dirichlet's Principle). At the present time, in most practical computation finite element methods are preferred to finite differences. Currently work is focused primarily on solution methods, either fast direct solvers or efficient iterative methods.

9.7.1 Fast Poisson Solver

Consider solving numerically the heat conduction problem

$$\left\{ \begin{array}{rcl} u_{xx} + u_{yy} & = & 0, \quad \forall x, y \in (0,1), \\ u(x,0) & = & g_0(x), \quad \forall x \in [0,1] \\ u(x,1) & = & g_1(x), \quad \forall x \in [0,1] \\ u_x(0,y) = & 0 & = u_x(1,y) \quad \forall y \in [0,1]. \end{array} \right. \tag{9.29}$$

(compare this with (6.7) where the solution was determined in terms is Fourier series in x) This corresponds to computing a temperature distribution on the unit square, with two edges thermally insulated and with prescribed temperatures on the others. A second order accurate finite difference approximation to the p.d.e. is

$$2(1+\alpha)U(x,y) - U(x-h_1,y) - U(x+h_1,y) - \alpha(U(x,y-h_2) + U(x,y+h_2)) = 0 \tag{9.30}$$

where $\alpha = \left(\dfrac{h_1}{h_2}\right)^2$. If $h_1 = h_2 = h$ this reduces to the *5-point Laplacian*

$$4U(x,y) - U(x-h,y) - U(x+h,y) - U(x,y-h) - U(x,y+h) = 0$$

To discretise the problem, use a uniform mesh of $(m_1+2)\times(m_2+2)$ nodes defined by:

$$\begin{array}{lll} x_j = (j-\frac{1}{2})h_1, & h_1 = 1/m_1, & j = 0,\ldots,m_1+1, \\ y_k = kh_2, & h_2 = 1/(m_2+1), & k = 0,\ldots,m_2+1. \end{array}$$

Note that the mesh is centred at the boundaries with Neumann boundary conditions $(x = 0, 1)$. This method clearly becomes complicated if the boundaries are not parallel to the grid, which is one reason why the finite element method with an unstructured grid is often preferred for irregular domains. Order the nodal values row-wise as

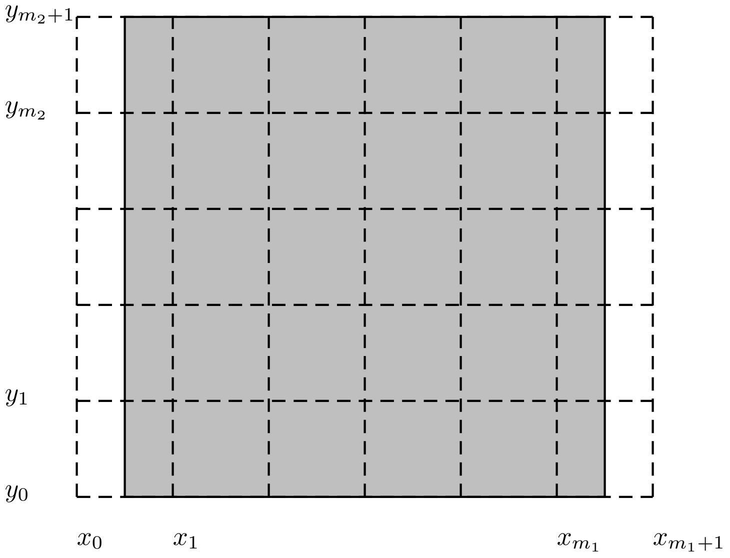
$$\boldsymbol{U}(y) = \begin{pmatrix} U(\frac{h_1}{2}, y) \\ \vdots \\ U(1-\frac{h_1}{2}, y) \end{pmatrix}, \qquad \boldsymbol{U} = \begin{pmatrix} U(h_2) \\ \vdots \\ U(1-h_2) \end{pmatrix}$$

A simple discrete solution satisfies finite difference equations can be written in the form

$$A\boldsymbol{U} = \boldsymbol{b} \tag{9.31}$$

where

$$\boldsymbol{b}(0) = \begin{pmatrix} g_0(x_1) \\ \vdots \\ g_0(x_{m_1}) \end{pmatrix}, \qquad \boldsymbol{b}(1) = \begin{pmatrix} g_1(x_1) \\ \vdots \\ g_1(x_{m_1}) \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}(0) \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \\ \boldsymbol{b}(1) \end{pmatrix}$$



The coefficient matrix is a block-tridiagonal $(m_1 m_2) \times (m_1 m_2)$ matrix with m_2 block rows

$$A = \begin{pmatrix} \tilde{A} & T & & & \\ T & \tilde{A} & T & & \\ & \ddots & \ddots & \ddots & \\ & & T & \tilde{A} & T \\ & & & T & \tilde{A} \end{pmatrix}$$

where for $\tilde{A}$ is the $m_1 \times m_1$ tridiagonal matrix

$$\tilde{A} = \begin{pmatrix} 2\alpha+1 & -1 & & & \\ -1 & 2\alpha+2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2\alpha+2 & -1 \\ & & & -1 & 2\alpha+1 \end{pmatrix}$$

and $T = -\alpha I$. It is known that the eigenvectors of the matrix $\tilde{A}$ have the form $\boldsymbol{v}^{(k)} = (v_1^{(k)}, \ldots, v_{m_1}^{(k)})^T$ where

$$v_l^{(k)} = \begin{cases} \sqrt{\frac{1}{m_1}}, & l = 1, \ldots, m_1; \quad k = 1 \\ \sqrt{\frac{2}{m_1}} \cos\left((k-1)(l-\frac{1}{2})\pi h_1\right), & l = 1, \ldots, m_1; \quad k = 2, \ldots, m_1 \end{cases}$$

Hence

$$\tilde{A}\boldsymbol{v}^{(k)} = \lambda_k \boldsymbol{v}^{(k)}, \quad k = 1, \ldots, m_1$$

where λ_k is the kth eigenvalue of $\tilde{A}$ and

$$V^T \tilde{A} V = \Lambda = \mathrm{diag}(\lambda_k) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{m_1} \end{pmatrix}.$$

Then define the $m_1 \times m_1$ matrix

$$V = \left(\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(m_1)} \right)$$

and the block-diagonal $(m_1 m_2) \times (m_1 m_2)$ matrix with m_2 block rows

$$Q = \begin{pmatrix} V & & \\ & \ddots & \\ & & V \end{pmatrix} = I \otimes V$$

where I is the $m_2 \times m_2$ unit matrix. The *Kroneker (tensor) product* $A \otimes B$ is defined as

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}$$

Then as Q is orthogonal, the system of equations can be written as

$$QQ^T A Q Q^T \boldsymbol{U} = \boldsymbol{b}$$

or

$$\left(Q^T A Q \right) Q^T \boldsymbol{U} = Q^T \boldsymbol{b}$$

the matrix A is therefore replaced by

$$\begin{aligned} Q^T A Q \quad &= \begin{pmatrix} \Lambda & & & \\ & \Lambda & & \\ & & \ddots & \\ & & & \Lambda \end{pmatrix} + \begin{pmatrix} 0 & T & & \\ T & 0 & T & \\ & \ddots & \ddots & \ddots \\ & & T & 0 \end{pmatrix} \\ &= I \otimes \Lambda \; + \; T_2 \otimes I \end{aligned}$$

where the unit matrices are respectively of size m_2 and m_1 and where T_2 is the $m_2 \times m_2$ matrix

$$T_2 = \alpha \begin{pmatrix} & -1 & & & \\ -1 & & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & & -1 \\ & & & -1 & \end{pmatrix}$$

Assuming this corresponds to ordering is initially by rows, reordering by columns leads to the matrix

$$N = \Lambda \otimes I + I \otimes T_2 = \begin{pmatrix} \lambda_1 I + T_2 & & \\ & \ddots & \\ & & \lambda_{m_1} I + T_2 \end{pmatrix}$$

and the $m_2 \times m_2$ matrix $N_j = \lambda_j I + T_2$ is tridiagonal and so a direct solution by Gauss Elimination is straightforward. Computationally, the solution can be split into three phases.

Phase 1:

$$\boldsymbol{v} \leftarrow Q^T \boldsymbol{b}$$

This is multiplication of the row-wise data $\boldsymbol{b}$ by the matrix V, *i.e.* it is a cosine transform of type 2 (see section on discrete transforms). A cosine transform of length n can be performed efficiently in $\mathcal{O}(n \log n)$ operations, hence as it is only the first and last blocks of $\boldsymbol{b}$ that contain non-trivial data, the computational complexity of this phase is $\mathcal{O}(m_1 \log m_1)$.

Phase 2: Solve

$$N\boldsymbol{z} = \boldsymbol{v}$$

This is a sequence of tridiagonal solves on the column-wise data (so a reordering of the data is required), the matrix corresponding to block row j is

$$N_j = \begin{pmatrix} \lambda_j & -\alpha & & & \\ -\alpha & \lambda_j & -\alpha & & \\ & \ddots & \ddots & \ddots & \\ & & -\alpha & \lambda_j & -\alpha \\ & & & -\alpha & \lambda_j \end{pmatrix}$$

The tridiagonal matrices are $m_2 \times m_2$ and hence each system can be solved in $\mathcal{O}(m_2)$ operations. There are m_1 systems, so the overall computational complexity of the second phase is $\mathcal{O}(m_1 m_2)$

Phase 3:

$$\boldsymbol{u} \leftarrow Q\boldsymbol{z}$$

This is multiplication of the row-wise data $\boldsymbol{z}$ by the matrix V^T, *i.e.* it is an inverse cosine transform of type 2, *i.e.* a cosine transform of type 3 as defined later, is required after the data has been reordered back to the original row-wise ordering. I this phase the computation is again $\mathcal{O}(m_1 \log m_1)$ for each block but now there are m_2 non-trivial blocks so the cost for this phase is $\mathcal{O}(m_1 m_2 \log m_1)$

The third phase cost dominates so the total cost is $\mathcal{O}(m_1 m_2 \log m_1)$. In the original form the solution was defined by an $m_2 m_1 \times m_2 m_1$ matrix with $\mathcal{O}(m_1)$ bandwidth hence a full solution using Guass elimination would cost $\mathcal{O}(m_1^3 m_2)$ operations.

9.8 Spectral Methods

Given the equation

$$-u_{xx} - u_{yy} = f.$$

Apply a Discrete Fourier Transform (DFT) (9.42) in both x and y to give

$$j^2 U_{j,k} + k^2 U_{j,k} = F_{j,k}$$

using the coefficients

$$U_{j,k} = \frac{F_{j,k}}{j^2 + k^2}$$

apply an IDFT in each direction to obtain U. Only finite number of terms in both DFT and IDFT are used so the solution is approximate. An efficient solution implments the DFT as an FFT.

9.9 Theory of Gradient Methods

For a more general sparse matrix A, that does not have a narrow regular band structure Guass elimination is even less efficient and it is necessary to use iteration. Unfortunately, simple iterations such as Jacobi or Gauss-Seidel can only converge rapidly for matrices that are strongly diagonally dominant ($|a_{ii}| \gg \sum_{j \neq i} |a_{ij}|$) and matrices that are not strongly diagonally dominant require alternative methods that converge faster. To cut down on the subscripts/superscripts in this section the following simplifying notation is used: $\boldsymbol{x} \equiv \boldsymbol{x}^{(i)}$ is the current approximation and $\boldsymbol{x}_+ \equiv \boldsymbol{x}^{(i+1)}$ is the next iterate, *etc.*, and the exact solution is $\boldsymbol{x}^*$.

At each iteration, given a *search direction* $\boldsymbol{p}$, then the approximate solution is updated as

$$\boldsymbol{x}_+ = \boldsymbol{x} \ + \ \alpha \boldsymbol{p} \tag{9.32}$$

where α is computed by a local minimisation that will be explained below. The residual $\boldsymbol{r} = \boldsymbol{b} \ - \ A\boldsymbol{x}$ is also updated rather than computed explicitly and so from (9.32) multiplying by A

$$A\boldsymbol{x}_+ = A\boldsymbol{x} \ + \ \alpha A\boldsymbol{p}$$

and it follows that

$$\boldsymbol{r}_+ = \boldsymbol{r} \ - \ \alpha A\boldsymbol{p} \tag{9.33}$$

If the matrix A is symmetric and positive definite then the quadratic form $\boldsymbol{x}^T A\boldsymbol{x}$ can be used to define a norm for the vector $\boldsymbol{x}$ as

$$\|\boldsymbol{x}\|_A = \left(\boldsymbol{x}^T A\boldsymbol{x}\right)^{1/2}.$$

When A is symmetric positive definite then A^{-1} is also symmetric positive definite and hence the quadratic form $\boldsymbol{x}^T A^{-1}\boldsymbol{x}$ also defines a norm. For *conjugate gradients*, at each iteration, the parameter α is computed by local minimisation of the residual in terms of such a norm, *i.e.*

$$\min_{\alpha} \ \boldsymbol{r}_+^T A^{-1}\boldsymbol{r}_+ = \min_{\alpha} \ (\boldsymbol{b} \ - \ A\boldsymbol{x}_+)^T A^{-1}(\boldsymbol{b} \ - \ A\boldsymbol{x}_+)$$

where as $A = A^T$ and so $\boldsymbol{x}^T A\boldsymbol{y} = \boldsymbol{y}^T A\boldsymbol{x}$ and from the update formula (9.33)

$$\begin{aligned} \boldsymbol{r}_+^T A^{-1}\boldsymbol{r}_+ &= (\boldsymbol{r} \ - \ \alpha A\boldsymbol{p})^T A^{-1}(\boldsymbol{r} \ - \ \alpha A\boldsymbol{p}) \\ &= \boldsymbol{r}^T A^{-1}\boldsymbol{r} \ - \ 2\alpha \boldsymbol{p}^T\boldsymbol{r} \ + \ \alpha^2 \boldsymbol{p}^T A\boldsymbol{p} \end{aligned}$$

the gradient is zero at minimum hence differentiating w.r.t. α gives

$$\alpha = \frac{(\boldsymbol{b} \ - \ A\boldsymbol{x})^T\boldsymbol{p}}{\boldsymbol{p}^T A\boldsymbol{p}} = \frac{\boldsymbol{r}^T\boldsymbol{p}}{\boldsymbol{p}^T A\boldsymbol{p}} \tag{9.34}$$

The *error* can be written in terms of the *residual* as $\boldsymbol{e} = \boldsymbol{x}^* - \boldsymbol{x} = A^{-1}\boldsymbol{r}$ and the norm to be minimised can be rewritten as

$$\boldsymbol{r}_+^T A^{-1} \boldsymbol{r}_+ = \boldsymbol{e}_+^T A \boldsymbol{e}_+$$

Thus conjugate gradients requires

$$\min_{\alpha} \; \boldsymbol{e}_+^T A \boldsymbol{e}_+ .$$

For any symmetric positive definite matrix B other methods are derived with

$$\min_{\alpha} \; \boldsymbol{e}_+^T B \boldsymbol{e}_+$$

i.e. minimising any error *norm*

$$\|\boldsymbol{e}\|_B = \left(\boldsymbol{e}^T B \boldsymbol{e}\right)^{1/2} .$$

9.9.1 Computing the Search Direction

At each iteration of conjugate gradients it is essential to: compute the value of α from (9.34), update the approximation from (9.32), update the residual from (9.33), but first it is necessary to provide the new search direction $\boldsymbol{p}$.

At each step define the matrix P in which the columns are the search directions, so $P_+ = \left[\begin{array}{cc} P & \boldsymbol{p}_+ \end{array}\right]$. Similarly define R in which the columns are the residuals, so $R_+ = \left[\begin{array}{cc} R & \boldsymbol{r} \end{array}\right]$. Then define span$\{P\}$ as the space spanned by the search directions. If we require that $\boldsymbol{x}_+$ is also optimal in the whole space spanned by the search directions so far, *i.e.* a global minimum not just a local minimum, then writing $\boldsymbol{x}$ as a linear combination of all the search directions gives

$$\boldsymbol{x}_+ = P\boldsymbol{a} \; + \; \alpha \boldsymbol{p}$$

where the coefficients $\boldsymbol{a}$ and α solve the minimisation problem

$$\min_{\boldsymbol{a},\,\alpha} \boldsymbol{r}_+^T A^{-1} \boldsymbol{r}_+ = \min_{\boldsymbol{a},\,\alpha} \; (\boldsymbol{b} \; - \; AP\boldsymbol{a} \; - \; \alpha A\boldsymbol{p})^T A^{-1} (\boldsymbol{b} \; - \; AP\boldsymbol{a} \; - \; \alpha A\boldsymbol{p})$$

then expanding the quadratic form,

$$\begin{aligned} \boldsymbol{r}_+^T A^{-1} \boldsymbol{r}_+ = & \; (\boldsymbol{b} \; - \; AP\boldsymbol{a} \; - \; \alpha A\boldsymbol{p})^T A^{-1} (\boldsymbol{b} \; - \; AP\boldsymbol{a} \; - \; \alpha A\boldsymbol{p}) \\ = & \; \boldsymbol{b}^T A^{-1} \boldsymbol{b} \; - \; 2\alpha \boldsymbol{p}^T \boldsymbol{b} \; + \; \alpha^2 \boldsymbol{p}^T A \boldsymbol{p} \\ & - \; 2\boldsymbol{a}^T P^T \boldsymbol{b} \; + \; \boldsymbol{a}^T P^T A P \boldsymbol{a} \\ & + \; 2\alpha \boldsymbol{a}^T P^T A \boldsymbol{p} \end{aligned}$$

The two minimisations, w.r.t. α and w.r.t. $\boldsymbol{a}$ are uncoupled iff

$$\boldsymbol{a}^T P^T A \boldsymbol{p} \alpha = 0$$

as α and $\boldsymbol{a}$ are arbitrary this condition becomes

$$P^T A\boldsymbol{p} = 0 \tag{9.35}$$

that is, the search directions are *conjugate* (or *A-orthogonal*),

$$\boldsymbol{p}^T A\boldsymbol{q} = 0 \qquad \text{for any } \boldsymbol{p} \neq \boldsymbol{q}$$

The matrix A is symmetric positive definite so we can interpret $(\boldsymbol{x}^T A\boldsymbol{x})^{1/2}$ as a norm and hence $p^T Aq$ as an inner product.

9.9.2 Convergence

The minimisation w.r.t. α leads to (9.34) and the other minimisation

$$\min_{\boldsymbol{a}} \ (\boldsymbol{b} \ - \ AP\boldsymbol{a})^T A^{-1}(\boldsymbol{b} \ - \ AP\boldsymbol{a})$$

after differentiating leads to

$$P^T(\boldsymbol{b} \ - \ AP\boldsymbol{a}) = P^T\boldsymbol{r} = 0. \tag{9.36}$$

If in addition,

$$P_+^T\boldsymbol{r} = \boldsymbol{p}_+^T\boldsymbol{b} - \boldsymbol{p}_+^T AP\boldsymbol{a} = \boldsymbol{p}_+^T\boldsymbol{b} = 0$$

then $\boldsymbol{b} \in \text{span}\{AP\}$ and

$$\begin{aligned} & \boldsymbol{x}^* = A^{-1}\boldsymbol{b} \in \text{span}\{P\} \\ \Rightarrow \quad & \boldsymbol{x}^* = \boldsymbol{x} = PA\boldsymbol{a} \\ \Rightarrow \quad & \boldsymbol{r} = 0 \end{aligned}$$

so that either the iteration has converged or $\boldsymbol{p}^T\boldsymbol{r} \neq 0$.

Given that $\boldsymbol{p}^{(1)} = \boldsymbol{r}^{(0)} = \boldsymbol{b}$ it can be proved by induction from (9.35) and (9.36) that

$$\text{span}\{R\} = \text{span}\{P_+\}.$$

Thus there exists a nonsingular triangular matrix S corresponding to the change in basis $R = P_+S$, so from (9.36),

$$P^T\boldsymbol{r}_+ = 0 = R^T\boldsymbol{r}_+ \tag{9.37}$$

and the *residuals are orthogonal*, *i.e.* $\boldsymbol{r}^T\boldsymbol{s} = 0$.

9.9.3 Recurrence Relation for the search direction

Assume that the update for $\boldsymbol{p}$ can be written in the form

$$\boldsymbol{p}_+ = \boldsymbol{r}_+ \ + \ \beta\boldsymbol{p} \tag{9.38}$$

From (9.38) the conjugacy condition $\boldsymbol{p}_+^T A\boldsymbol{p} = 0$ gives

$$\beta = -\frac{\boldsymbol{r}_+^T A\boldsymbol{p}}{\boldsymbol{p}^T A\boldsymbol{p}} \tag{9.39}$$

The update formulae (9.33) and (9.39) involve 3 different inner products so the efficiency of the method can be improved using the orthogonality of the residuals $\boldsymbol{r}_+^T\boldsymbol{r} = 0$ with

$$\boldsymbol{r}_+ = \boldsymbol{r} \; - \; \alpha A\boldsymbol{p}$$

gives

$$\boldsymbol{r}_+^T\boldsymbol{r}_+ = \; - \; \alpha\boldsymbol{r}_+^T A\boldsymbol{p}.$$

Similarly $\boldsymbol{r}_+^T\boldsymbol{p} = 0$ and so from (9.38)

$$\boldsymbol{r}_+^T\boldsymbol{p}_+ = \boldsymbol{r}_+^T\boldsymbol{r}_+$$

and from (9.33)

$$\boldsymbol{r}^T\boldsymbol{r} = \alpha\boldsymbol{p}^T A\boldsymbol{p}.$$

So (9.39) becomes

$$\beta = -\frac{\boldsymbol{r}_+^T\boldsymbol{r}_+}{\boldsymbol{r}^T\boldsymbol{r}} \tag{9.40}$$

with (9.33) becoming

$$\alpha = \frac{\boldsymbol{r}^T\boldsymbol{r}}{\boldsymbol{p}^T A\boldsymbol{p}}. \tag{9.41}$$

Hence with these modfications to the computations of the scalars **the Basic CG Algorithm** becomes:

```
$\boldsymbol{x}^{(0)} = 0;$          $\boldsymbol{r}^{(0)} = \boldsymbol{p}^{(0)} = \boldsymbol{b}$
$i = 0;$                  $\rho^{(0)} = \; {\boldsymbol{r}^{(0)}}^T\boldsymbol{r}^{(0)}$
while not converged do
            $\boldsymbol{v}^{(i)}$       $= \; A\boldsymbol{p}^{(i)}$
            $\alpha$          $= \; \rho^{(i)}/{\boldsymbol{p}^{(i)}}^T\boldsymbol{v}^{(i)}$
            $\boldsymbol{x}^{(i+1)}$   $= \; \boldsymbol{x}^{(i)} \; + \; \alpha\boldsymbol{p}^{(i)}$
            $\boldsymbol{r}^{(i+1)}$   $= \; \boldsymbol{r}^{(i)} \; - \; \alpha\boldsymbol{v}^{(i)}$
            $\rho^{(i+1)}$   $= \; {\boldsymbol{r}^{(i+1)}}^T\boldsymbol{r}^{(i+1)}$
            $\beta$          $= \; \rho^{(i+1)}/\rho^{(i)}$
            $\boldsymbol{p}^{(i+1)}$   $= \; \boldsymbol{r}^{(i+1)} \; + \; \beta\boldsymbol{p}^{(i)}$
            $i$          $= i + 1$
enddo
```

9.9.4 Preconditioning

The key to rapid convergence of iterative methods is the preconditioning, that is in order to solve $A\boldsymbol{x} = \boldsymbol{b}$, then the iteration is applied to the system

$$\tilde{A}\boldsymbol{x} = \tilde{\boldsymbol{b}}$$

where $\tilde{A} = M^{-1}A$ and $\tilde{\boldsymbol{b}} = M^{-1}\boldsymbol{b}$. Alternatively with $M = M_L M_R$ solve

$$M_L^{-1} A M_R^{-1} \boldsymbol{y} = M_L^{-1}\boldsymbol{b}.$$

The two solutions are connected by $M_R\boldsymbol{x} = \boldsymbol{y}$. If the matrix $M_L^{-1}AM_R^{-1}$ is to be symmetric, when A is symmetric, then $M_L = M_R^T$.
The matrix $M = M_L M_R$ is an approximation to A

9.9.5 Preconditioned Conjugate Gradients (PCG)

The algorithm for PCG requires one linear solve

$$M_L M_R \boldsymbol{z} \;=\; \boldsymbol{r}$$

per iteration, the conjugacy conditions become

$$\begin{aligned} \boldsymbol{r}^{(j)^T} M_R^{-1} M_L^{-1} \boldsymbol{r}^{(i)} &= 0 \\ \boldsymbol{p}^{(j)^T} M_R^{-1} A M_L^{-1} \boldsymbol{p}^{(i)} &= 0 \end{aligned}$$

The derivation is straightforward with the substitutions

$$\begin{aligned} \boldsymbol{r} &\to M_L\boldsymbol{r} \\ \boldsymbol{v} &\to M_L\boldsymbol{v} \\ \boldsymbol{x} &\to M_R^{-1}\boldsymbol{x} \\ \boldsymbol{p} &\to M_R^{-1}\boldsymbol{p} \end{aligned}$$

The PCG Algorithm can be written as:

```
$\boldsymbol{x}^{(0)} = 0;$          $\boldsymbol{r}^{(0)} = \boldsymbol{b}$
$i = 0;$             $\boldsymbol{z}^{(0)} = \boldsymbol{p}^{(0)} = M_R^{-1} M_L^{-1} \boldsymbol{b}$      $\rho^{(0)} = \ {\boldsymbol{r}^{(0)}}^T \boldsymbol{z}^{(0)}$
while not converged do
        $\boldsymbol{v}^{(i)}$       $= A\boldsymbol{p}^{(i)}$
        $\alpha$       $= \rho^{(i)} / {\boldsymbol{p}^{(i)}}^T \boldsymbol{v}^{(i)}$
        $\boldsymbol{x}^{(i+1)}$     $= \boldsymbol{x}^{(i)} \ + \ \alpha \boldsymbol{p}^{(i)}$
        $\boldsymbol{r}^{(i+1)}$     $= \boldsymbol{r}^{(i)} \ - \ \alpha \boldsymbol{v}^{(i)}$
        $\boldsymbol{z}^{(i+1)}$     $= M_R^{-1} M_L^{-1} \boldsymbol{r}^{(i+1)}$     $ie$ Solve $M_L M_R \boldsymbol{z}^{(i+1)} \ = \ \boldsymbol{r}^{(i+1)}$
        $\rho^{(i+1)}$     $= {\boldsymbol{r}^{(i+1)}}^T \boldsymbol{z}^{(i+1)}$
        $\beta$       $= \rho^{(i+1)} / \rho^{(i)}$
        $\boldsymbol{p}^{(i+1)}$     $= \boldsymbol{z}^{(i+1)} \ + \ \beta \boldsymbol{p}^{(i)}$
        $i$       $= i + 1$
enddo
```

The choice of a good preconditioner can have a dramatic effect upon the rate of convergence. Popular choices are

- Diagonal Preconditioning

$$M = \text{diag}(A) \text{ so } M_L = M_R = M^{1/2}$$

- Incomplete Cholesky Factorisation
  - By position: *e.g.* ICCG(0), the *position* of the nonzeros in M is governed by the *position* of the nonzeros in A.
  - By value: Components of M are nonzero if component of L in $A = LL^T$ is large enough

9.9.6 Discrete Transforms

The sums

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n \exp\left(-i \frac{2\pi kn}{N}\right), \qquad k = 0, 1, \ldots, N-1 \tag{9.42}$$

are known as the *discrete Fourier transform* (DFT) of the data $x_n,\ n = 0, 1, \ldots, N-1$. It is a linear, invertible function $F : \mathbb{R}^N \to \mathbb{R}^N$. The *inverse discrete Fourier transform* (IDFT) of the data $X_k,\ k = 0, 1, \ldots, N-1$ gives

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k \exp\left(i \frac{2\pi nk}{N}\right), \qquad n = 0, 1, \ldots, N-1 \tag{9.43}$$

without the factor $\frac{1}{\sqrt{N}}$ the symmetry of formulae (9.42) and (9.43) is lost. The DFT can be implemented in an efficiently so the the transform takes $\mathcal{O}(n \log n)$ arithmetic operations. This is known as a *fast Fourier transform* (FFT), assuming $N = 2^n$, the FFT can be implemented using explicit recursion (as in the FFTW package [4]) a simple MATLAB implementation is below.

```
function xhat=myfft(x,n,z)
if n==1
    xhat=x;
else
    uhat=myfft(x(1:2:n-1), n/2, z(1:2:n-1));
    vhat=myfft(  x(2:2:n), n/2, z(1:2:n-1));
    xhat=[uhat+z(1:n/2).*vhat;uhat+z(n/2+1:n).*vhat];
end
```

The *discrete cosine transform* (DCT) is also a linear, invertible function $F : \mathbb{R}^N \to \mathbb{R}^N$. There are several variants of the DCT with slightly modified definitions. The N real numbers $x_0, \ldots, x_{N-1}$ are transformed into the N real numbers $X_0, \ldots, X_{N-1}$ according to one of the formulae:

DCT-II

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right) k\right]$$

The DCT-II is probably the most commonly used form, and is often simply referred to as “the DCT”.

This transform is exactly equivalent (up to an overall scale factor of 2) to a DFT of $4N$ real inputs of even symmetry where the even-indexed elements are zero. That is, it is half of the DFT of the $4N$ inputs y_n, where $y_{2n} = 0, y_{2n+1} = x_n,\ 0 \leq n < N$, and $y_{4Nn} = y_n,\ 0 < n < 2N$.

DCT-III

$$X_k = \frac{1}{2}x_0 + \sum_{n=1}^{N-1} x_n \cos\left[\frac{\pi}{N} n \left(k + \frac{1}{2}\right)\right]$$

Because it is the inverse of DCT-II (up to a scale factor), this form is sometimes simply referred to as “the inverse DCT” .

Exercises

Exercise 9.1 *Consider the following problem*

$$\begin{cases} u_t = u_x, & 0 < x < 1,\ t > 0, \\ u(0,t) = u(1,t), & t > 0, \\ u(x,0) = f(x), & 0 < x < 1, \end{cases} \tag{9.44}$$

which is approximated by

$$\begin{cases} \frac{1}{k}(U_j^{n+1} - U_j^n) = \frac{1}{2h}(U_{j+1}^{n+1} - U_{j-1}^{n+1}), & j = 0, \dots, N-1,\ n = 0, 1, \dots \\ U_N^n = U_0^n, & n = 0, 1, \dots \\ U_j^0 = f_j, & j = 0, \dots, N-1, \end{cases} \tag{9.45}$$

- *Derive the local truncation error*
- *Show that (9.45) is a consistent approximation to (9.44)*
- *Verify the stability of the method*

Exercise 9.2 *Consider the heat conduction equation*

$$u_t = u_{xx}, \quad 0 < x < 1,\ t > 0. \tag{9.46}$$

1. *It is approximated by the* Leap-Frog Scheme*:*

$$\frac{1}{2k}(U_j^{n+1} - U_j^{n-1}) = \frac{1}{h^2}(U_{j+1}^n - 2U_j^n + U_{j-1}^n), \quad \begin{cases} j = 1, \dots, N-1, \\ n = 1, 2, \dots \end{cases} \tag{9.47}$$

   - *Derive the local truncation error*
   - *Show that (9.47) is a consistent approximation to (9.46)*
   - *Verify that the method is unconditionally* unstable*!*

2. *It is approximated by the* Du Fort-Frankel Scheme*:*

$$\frac{1}{2k}(U_j^{n+1} - U_j^{n-1}) = \frac{1}{h^2}(U_{j+1}^n - (U_j^{n+1} + U_j^{n-1}) + U_{j-1}^n), \quad \begin{cases} j = 1, \dots, N-1, \\ n = 1, 2, \dots \end{cases} \tag{9.48}$$

   - *Derive the local truncation error*
   - *Under what conditions is (9.48) a consistent approximation to (9.46)?*
   - *Verify that the method is unconditionally* stable*!*

Exercise 9.3 *Solve the equation*

$$u_t + (1+\alpha x)u_x = 0$$

on the interval $x \in [-3,3]$ *and* $0 \le t \le 2$ *with the initial data*

$$u(x,0) = \begin{cases} 1-|x| & |x| \le 1, \\ 0 & \textit{otherwise} \end{cases}$$

using the Lax-Friedrichs scheme, use $\alpha = 0.5$ *and* $r = \frac{k}{h} = 1$. *Demonstrate that the instability phenomena occur where* $|(1+\alpha x_m)r| \ > 1$.

Exercise 9.4 *The initial-boundary value problem*

$$\begin{array}{l} u_t + au_x = 0,\ x>0,\ t>0 \\ u(x,0) = g(x),\ u(0,t) = f(t), \end{array} \tag{9.49}$$

is approximated on a grid $(x_j, t_n) = (jh, nk),\ j = 0,\ldots,N;\ t = 0,1,\ldots$.

1. *Derive the truncation error for the following difference approximation*[6]

   $$U_j^{n+1} + U_{j+1}^{n+1} = U_j^n + U_{j+1}^n - ar(U_{j+1}^{n+1} - U_j^{n+1} + U_{j+1}^n - U_j^n)$$

   where $r = \frac{k}{h}$.

2. *Write down the system of equations that must be solved at each time step. How is it solved?*

3. *Verify the system is stable if* $ar < 1$.

4. *How could the method be modified to solve the initial-boundary value problem*

   $$\begin{array}{l} u_t - au_x = 0,\ x<0,\ t>0 \\ u(x,0) = g(x),\ u(0,t) = f(t), \end{array} \tag{9.50}$$

5. *Why is it important that* $x > 0$ *in (9.49) but* $x < 0$ *in (9.50)?*

Chapter 10

Finite Element Methods

10.1 Introduction

The finite-element method was originally introduced in the 1950's as a method to calculate elastic deformations in solids. Later the method has been developed and generalised for all kinds of partial differential equations. It is the dominating technique for solid-mechanics problems such as estimating stresses and strains in elastic material under prescribed loads. CAD (Computer Aided Design) systems typically provides finite-element solvers in a highly integrated fashion. The engineer can typically with a few clicks on the computer screen estimate the deformations and stresses of, say, a machine part during the design. Finite-element methods are also commonly applied to other areas, such as calculations of electromagnetic fields and fluid flows.

In order to provide a brief introduction to the ideas, this note concentrates on a standard model problem for elliptic boundary-value problems, the Poisson problem. Only homogeneous Dirichlet boundary conditions are covered here.

10.2 FEM for the Poisson Problem in Two Space Dimensions

We consider the boundary-value problem

$$\begin{aligned} -\Delta u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \Gamma, \end{aligned} \tag{10.1}$$

where Ω is an open, bounded and connected domain in the plane, and Γ is its boundary. The Laplacian Δ is the sum of second derivatives

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Letting u represent a temperature field, equation (10.1) models steady heat conduction in a homogeneous, isotropic material, such as a metal, in which the temperature is held at zero on the boundary. The function f can be used to model heat sources such as electric heaters embedded in the material.

10.3 Green's Formula

.

We need some definitions and formulae from vector calculus. For any differentiable function v from $\mathbb{R}^m$ to $\mathbb{R}$, the *gradient* is the vector

$$\nabla v = \left(\frac{\partial v}{\partial x_1}, \frac{\partial v}{\partial x_2}, \dots, \frac{\partial v}{\partial x_m} \right),$$

and for any differentiable vector-valued function $\boldsymbol{w} = (w_1, w_2, \dots, w_m)$ from $\mathbb{R}^m$ to $\mathbb{R}^m$, the *divergence* is

$$\nabla \cdot \boldsymbol{w} = \sum_{i=1}^{m} \frac{\partial w_i}{\partial x_i}.$$

By the product rule of differentiation, the formula

$$\nabla \cdot (v \nabla u) = \nabla v \cdot \nabla u + v \Delta u, \tag{10.2}$$

that is,

$$\sum_{i=1}^{m} \frac{\partial}{\partial x_i} \left(v \frac{\partial u}{\partial x_i} \right) = \sum_{i=1}^{m} \frac{\partial v}{\partial x_i} \frac{\partial u}{\partial x_i} + \sum_{i=1}^{m} v \frac{\partial^2 u}{\partial x_i^2},$$

holds for differentiable functions v and twice differentiable functions u.

Also recall the *divergence theorem* (or Gauss' theorem) which identifies the integral of a vector-field divergence over a domain with the integral of the normal component of the field along the boundaries:

$$\int_{\Omega} \nabla \cdot \boldsymbol{w} \, d\Omega = \int_{\Gamma} \boldsymbol{n} \cdot \boldsymbol{w} \, ds, \tag{10.3}$$

Here, $\boldsymbol{n}$ denotes the outward-directed unit normal on Γ. Identity (10.3) holds for functions $\boldsymbol{w}$ and boundaries Γ that are sufficiently smooth.

Combining the divergence theorem (10.3) with formula (10.2) yields *Green's formula*

$$\int_{\Gamma} v \frac{\partial u}{\partial n} \, ds = \int_{\Omega} \nabla v \cdot \nabla u \, d\Omega + \int_{\Omega} v \Delta u \, d\Omega, \tag{10.4}$$

where

$$\frac{\partial u}{\partial n} = \boldsymbol{n} \cdot \nabla u = \sum_{i=1}^{m} n_i \frac{\partial u}{\partial x_i},$$

denotes the directional derivative of u in the normal direction. Green's formula is a generalisation to higher dimensions of the integration-by-parts formula

$$\int_0^1 u' v' \, dx = u'(1) v(1) - u'(0) v(0) - \int_0^1 u'' v \, dx.$$

10.4 The Variational Form

A *classical solution* to the Poisson problem (10.1) is a smooth function u satisfying equation (10.1). The precise requirements for u to be a classical solution is that it should be twice continuously differentiable, and its first and second derivatives should be functions that can be continuously extended up to the boundary. This assures that Green's formula (10.4) can be applied on u. Let v be a smooth function from $\overline{\Omega} = \Omega \cup \Gamma$ to $\mathbb{R}$ such that $v(x) = 0$ for each $x \in \Gamma$. Multiply both sides of equation (10.1) with v, integrate over Ω, and apply Green's formula (10.4) to obtain

$$\begin{array}{ll} \int_\Omega f v \, d\Omega & = -\int_\Omega v \Delta u \, d\Omega \\ & = -\int_\Gamma v \frac{\partial u}{\partial n} \, ds + \int_\Omega \nabla v \cdot \nabla u \, d\Omega = \int_\Omega \nabla v \cdot \nabla u \, d\Omega, \end{array} \tag{10.5}$$

where the fact that v vanishes on the boundary has been used in the last equality. From expression (10.5) immediately follows

Theorem 39 *If u is a classical solution to the Poisson problem (10.1), then u satisfies*

$$\int_\Omega \nabla u \cdot \nabla v \, d\Omega = \int_\Omega f v \, d\Omega, \tag{10.6}$$

for each smooth function v vanishing on the boundary.

Equation (10.6) is called the *variational form* of the Poisson equation. Theorem 39 refers to the original problem (10.1), but the variational form can be used to *define* a function u without reference to the differential equation. For this purpose, we introduce the *function space*

$$V = \left\{ v : \int_\Omega |\nabla v|^2 \, d\Omega < +\infty \text{ and } v|_\Gamma = 0 \right\} , \tag{10.7}$$

where

$$|\nabla v|^2 = \left(\frac{\partial v}{\partial x_1} \right)^2 + \left(\frac{\partial v}{\partial x_2} \right)^2 .$$

The condition

$$\int_\Omega |\nabla v|^2 \, d\Omega < +\infty$$

corresponds in many applications to demanding that the *energy* should be bounded, for instance when the Poisson equation is used to model steady heat conduction. Note that V is a *linear space*, that is, if v, $w \in V$, then $\alpha v + \beta w \in V$ for each α, $\beta \in \mathbb{R}$. The space V is a so-called Sobolev space, and is often denoted $H_0^1(\Omega)$ in the literature.

The variational problem, now formulated without reference to the differential equation (10.1) is

$$\begin{array}{l} \text{Find } u \in V \text{ such that} \\ \int_\Omega \nabla u \cdot \nabla v \, d\Omega = \int_\Omega f v \, d\Omega \qquad \forall v \in V. \end{array} \tag{10.8}$$

Solutions to variational problem (10.8) are called *weak solutions* of the partial differential equation (10.1). From Theorem 39 follows that classical solutions are weak solutions. As the label "weak" suggests, there are weak solutions that are not classical solutions. However, one can show that weak solutions are classical solutions provided that the function f and the boundary Γ are sufficiently smooth.

10.5 The Minimisation Problem

The variational form above is all that is needed to define a finite-element discretisation. However, a classical solution to the particular problem that we consider, equation (10.1), also satisfies a certain *minimisation problem*, that is, the classical solution minimises the quadratic form

$$F(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 \, d\Omega - \int_\Omega fv \, d\Omega.$$

Similarly as was done for the variational problem, we can also consider the problem of minimising F within the function space V without reference to classical solutions, that is, consider the problem:

$$\begin{array}{c} \text{find } u \in V \text{ such that} \\ F(u) \le F(v) \qquad \forall v \in V. \end{array} \tag{10.9}$$

In fact, the variational problem (10.8) and the minimisation problem (10.9) are equivalent:

Theorem 40 *The element $u \in V$ minimises F if and only if it is a solution to the variational problem (10.8)*

Remark 1 *The proof below may appear long, but in essence it is really no more complicated than showing that the parabola $F(x) = \frac{1}{2}x^2 - xf$ has its minimum at $x = f$*

Proof

For any u, $v \in V$, we have

$$\begin{aligned} F(u+v) &= \tfrac{1}{2}\textstyle\int_\Omega |\nabla u + \nabla v|^2 \, d\Omega - \int_\Omega f(u+v)\, d\Omega \\ &= \tfrac{1}{2}\textstyle\int_\Omega \left[|\nabla u|^2 + 2\nabla u \cdot \nabla v + |\nabla v|^2\right] \, d\Omega - \int_\Omega f(u+v)\, d\Omega. \end{aligned} \tag{10.10}$$

(i) Assume that $u \in V$ is a solution to the variational problem (10.8). Then expression (10.10) reduces to

$$\begin{aligned} F(u+v) &= \tfrac{1}{2}\textstyle\int_\Omega |\nabla u|^2 \, d\Omega - \int_\Omega fu \, d\Omega + \tfrac{1}{2}\int_\Omega |\nabla v|^2 \, d\Omega \\ &= F(u) + \underbrace{\frac{1}{2}\int_\Omega |\nabla v|^2 \, d\Omega}_{\ge 0} \ge F(u) \end{aligned} \tag{10.11}$$

for any $v \in V$, which shows that u minimises F.

(ii) Now assume that $u \in V$ minimises F. For any $t \in \mathbb{R}$ and $v \in V$, we define the function $f(t) = F(u + tv)$, that is, by perturbing F away from its minimum. Expression (10.10) yields that

$$\begin{aligned} f(t) &= F(u+tv) \\ &= F(u) + t\left(\textstyle\int_\Omega \nabla u \cdot \nabla v \, d\Omega - \int_\Omega fv \, d\Omega\right) + \tfrac{t^2}{2}\textstyle\int_\Omega |\nabla v|^2 \, d\Omega, \end{aligned} \tag{10.12}$$

that is, f is a second-order polynomial in t with a minimum when the derivative vanishes (since the leading term is non-negative). We also know that the minimum is attained for $t = 0$ since u minimises F. Setting $f'(0) = 0$ yields that

$$\int_\Omega \nabla u \cdot \nabla v \, d\Omega - \int_\Omega fv \, d\Omega = 0, \tag{10.13}$$

for any $v \in V$, that is, u is a solution to the variational problem (10.8).

Q.E.D.

Remark 2 *Variational forms can be defined for practically all elliptic boundary-value problems, but minimisation forms cannot always be defined, for instance when the differential equation contains first-derivative terms.*

Remark 3 *In mechanics application the variational form (10.8) is called* the principle of virtual work*, and the minimisation problem (10.9) is called* the principle of minimum potential energy*.*

Remark 4 *The terminology used here, "variational" for (10.8) and "minimisation" for (10.9), is convenient for our purpose, but is not the only existing. Quite commonly the minimisation problem is called a variational form. In fact, the notion of variational forms was first attached to minimisations of "functionals" like F in the* calculus of variations*.*

10.6 Meshing and Finite-Element Approximation

We introduce a *triangulation* of the domain Ω, that is, Ω will be subdivided into non-overlapping triangles as illustrated in figures 10.1 and 10.3. The triangular corners are called the *nodes* of the triangulation. The *boundary nodes* are the nodes which are located on the boundary, and the *internal nodes* are the nodes which are not boundary nodes. A valid triangulation should not contain "hanging nodes", that is, no node should be located at another triangles side, as in figure 10.2. The "fineness" of the triangulation is characterised by a parameter $h > 0$, the largest length of any of the triangular sides, for instance.

Now define V_h as the space of all functions that are *continuous* on $\overline{\Omega}$, *linear* on each triangle, and *vanishing* on the boundary Γ. The graph of such a function is a surface composed of triangular-shaped planes, as illustrated in figure 10.4.

This space is constructed so that $V_h \subset V$, and we define the *finite-element discretisation* of the Poisson problem (10.1) as

$$\begin{array}{l} \text{Find } u_h \in V_h \text{ such that} \\ \int_\Omega \nabla u_h \cdot \nabla v_h \, d\Omega = \int_\Omega f v_h \, d\Omega \qquad \forall v_h \in V_h. \end{array} \tag{10.14}$$

Note that the discretisation is obtained simply by replacing V with the subspace V_h in the variational form (10.8). This kind of procedure is also called a *Galerkin approximation.*

10.7 The Algebraic Problem

A function in the above defined space V_h is uniquely defined by its values at the *internal* nodes (we already know that the function is zero at the boundary nodes). To see this, it is enough to note that the planar surface of u_h on each triangle is uniquely defined by the values of u_h at the triangular corners. Let N be the number of internal nodes. Using the *basis functions* $\{\phi_j(\boldsymbol{x})\}_{j=1}^N \subset V_h$, each function $u_h \in V_h$ can be written

$$u_h(\boldsymbol{x}) = \sum_{j=1}^{N} u_j \phi_j(\boldsymbol{x}), \tag{10.15}$$

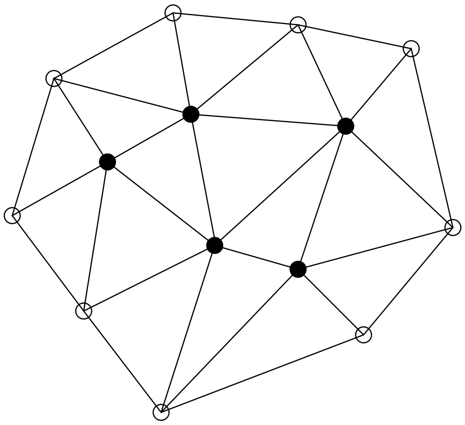


Figure 10.1: A valid triangulation. Internal nodes are marked by solid dots and boundary nodes by circles.

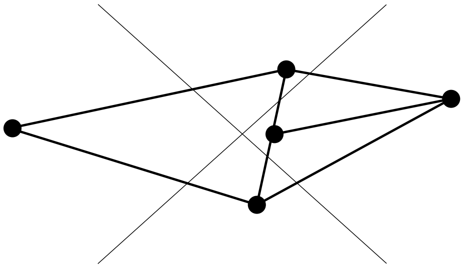


Figure 10.2: Not a valid triangulation: contains hanging nodes.

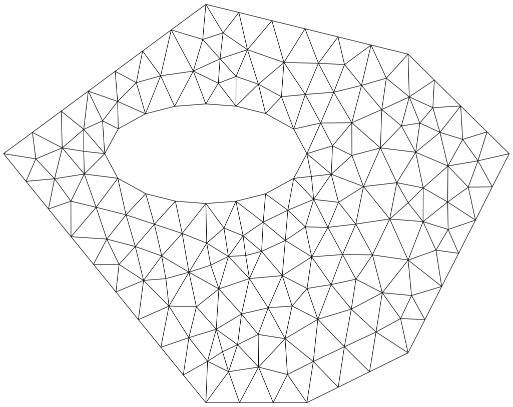


Figure 10.3: A more complicated triangulated domain (note that the domain may contain holes!)

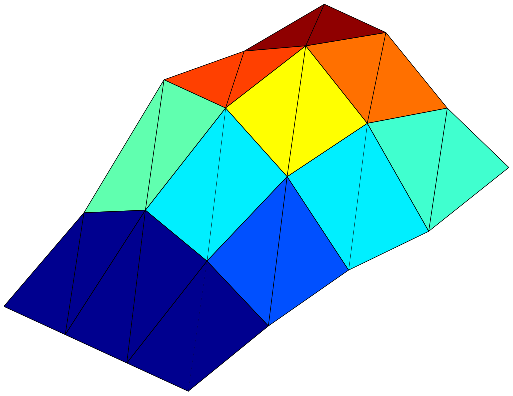


Figure 10.4: The functions in V_h are continuous and linear on each triangle. (The boundary nodes are not included in this picture.)

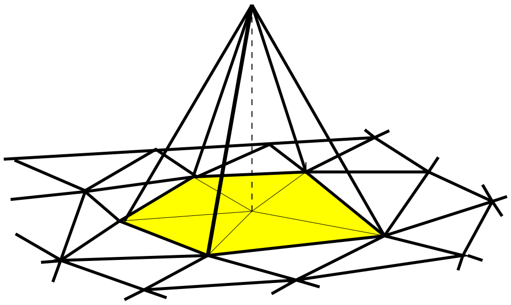


Figure 10.5: The basis function $\phi_j(\boldsymbol{x})$ is equal to one at node j and zero at all other nodes.

where u_j is the value of u_h at note j, and $\phi_j(\boldsymbol{x})$ is the "tent" function depicted in figure 10.5. The function ϕ_j is zero everywhere, except that it raises as a "tent" around node j, that is, $\phi_j \in V$ such that

$$\phi_j(\boldsymbol{x}_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{otherwise,} \end{cases}$$

where $\boldsymbol{x}_k$ is the coordinate of node k. Substituting the expansion (10.15) into equation (10.14) yields that

$$\sum_{j=1}^{N} u_j \int_\Omega \nabla\phi_j \cdot \nabla v_h \, d\Omega = \int_\Omega f v_h \, d\Omega \qquad \forall v_h \in V_h.$$

Since equation (10.7) should hold for each $v_h \in V_h$, it must in particular hold for $v_h = \phi_i$, $i = 1, \ldots, N$, which means that

$$\sum_{j=1}^{N} u_j \int_\Omega \nabla\phi_j \cdot \nabla \phi_i \, d\Omega = \int_\Omega f \phi_i \, d\Omega \qquad i = 1, \ldots, N. \tag{10.16}$$

Problem (10.16) is a system of linear equation in the coefficients u_j, $j = 1, \ldots, N$, that is,

$$A\boldsymbol{u} = \boldsymbol{b}, \tag{10.17}$$

where the matrix A has components

$$A_{ij} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \, d\Omega,$$

and

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} \int_\Omega f\phi_1 \, d\Omega \\ \vdots \\ \int_\Omega f\phi_N \, d\Omega \end{pmatrix}.$$

With a terminology borrowed from solid mechanics, the matrix A is called the *stiffness matrix* and the vector $\boldsymbol{b}$ the *load vector*. This terminology is used also for cases, like heat conduction, when the PDE we are discretising has nothing to do with mechanics!

We conclude that a numerical approximation of the Poisson problem with a finite-element method involves setting up and solving the linear system (10.17).

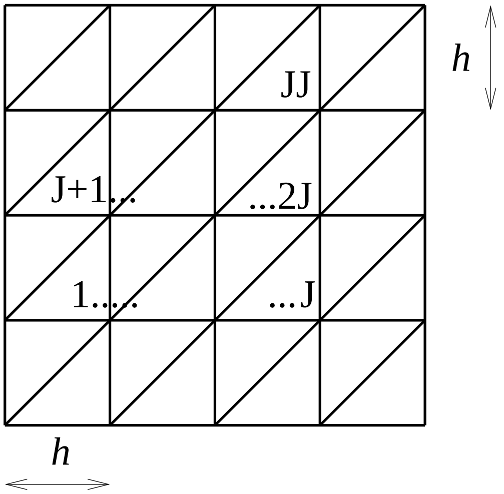


Figure 10.6: A structured meshing of the unit square.

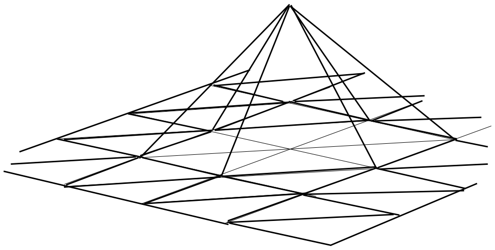


Figure 10.7: A basis function associated with the mesh in figure 10.6

10.8 An Example

Let the domain Ω be the unit square, and consider the *structured mesh* of figure 10.6. There are J internal nodes in both directions and the sides of each triangle are $h = 1/(J+1)$. There is a total of $J^2 = N$ internal nodes, assumed to be numbered in the row-wise direction as indicated in figure 10.6. The basis functions ϕ_i have the shape indicated in figure 10.7. The *support* of each basis function, that is, the nonzero region of the function, consists of the 6 neighbouring triangles that surround node i. Note that this means that most of the stiffness matrix elements

$$A_{ij} = \int_{\Omega} \nabla\phi_i \cdot \nabla\phi_j \, d\Omega$$

are zero. For instance, $A_{i,i+2} = 0$ since there is no overlap in the support for the functions ϕ_i and ϕ_{i+2}; see figure 10.8. In fact, A_{ij} can be nonzero only when i and j are associated with *nearest-neighbouring* nodes (figure 10.9).

To calculate the stiffness-matrix elements, we need to know the gradients of the basis functions,

$$\nabla\phi_i = \left(\frac{\partial\phi_i}{\partial x}, \frac{\partial\phi_i}{\partial y}\right).$$

The gradient is constant at each triangle since ϕ_i is composed of planar surfaces. Letting the x and y directions be oriented in the horizontal and vertical directions, respectively, the values of the gradient at the support of the basis function are indicated in figure 10.10. Note that the basis function is equal to one at the filled dot and equal to zero at the open dots, which means that the gradient can simply be read off as the slope of the "tent" function along the sides of the triangles. With the aid of the gradients given in

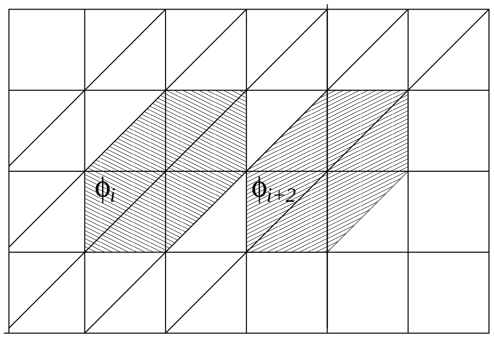


Figure 10.8: There is no overlap in the support for basis functions ϕ_i and ϕ_{i+2}.

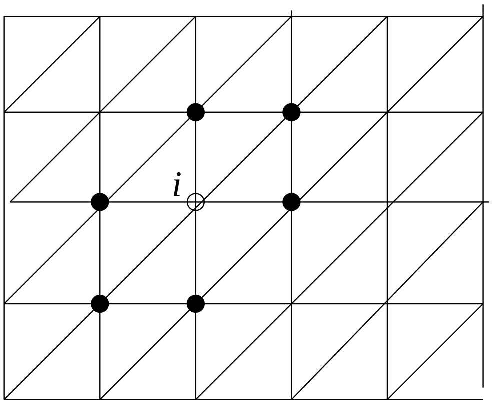


Figure 10.9: The nearest neighbours to node i are the six nodes marked with black dots. Thus, A_{ij} can be nonzero only when j corresponds to one of the black dots.

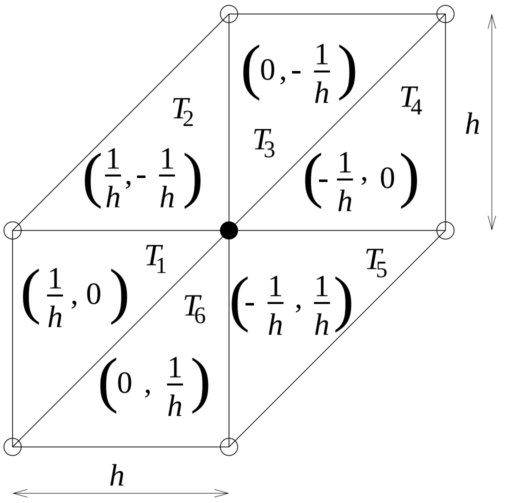

$\left(0, -\frac{1}{h}\right)$
T_4
T_2
h
T_3
$\left(\frac{1}{h}, -\frac{1}{h}\right)$
$\left(-\frac{1}{h}, 0\right)$
T_1
T_5
$\left(\frac{1}{h}, 0\right)$
T_6
$\left(-\frac{1}{h}, \frac{1}{h}\right)$
$\left(0, \frac{1}{h}\right)$
h

Figure 10.10: The gradient of basis function ϕ_i is piecewise constant on each triangle. The x- and y-coordinates are given as the pair $(\cdot, \cdot)$ at each triangle of the support of the function.

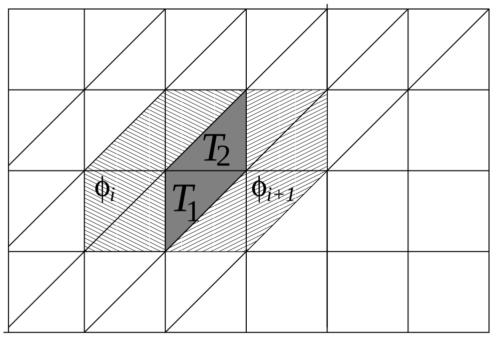


Figure 10.11: The overlap in the support of basis functions ϕ_i and ϕ_{i+1} are the triangles T_1 and T_2.

figure 10.10, we can compute the diagonal elements in the stiffness matrix,

$$\begin{aligned} A_{ii} &= \int_\Omega \nabla\phi_i \cdot \nabla\phi_i \, d\Omega = \textstyle\sum_{i=1}^{6} \int_{T_k} \nabla\phi_i \cdot \nabla\phi_i \, d\Omega \\ &= \tfrac{1}{h^2}|T_1| + 2\tfrac{1}{h^2}|T_2| + \tfrac{1}{h^2}|T_3| + \tfrac{1}{h^2}|T_4| + 2\tfrac{1}{h^2}|T_5| + \tfrac{1}{h^2}|T_6| \\ &= 8\tfrac{1}{h^2}\tfrac{h^2}{2} = 4. \end{aligned}$$

To compute $A_{i,i+1}$, note that $\nabla\phi_i \cdot \nabla\phi_{i+1} \neq 0$ only in two triangles (figure 10.11), thus

$$\begin{array}{lll} \text{on } T_1: & \nabla\phi_i = \left(-\tfrac{1}{h}, \tfrac{1}{h}\right) & \nabla\phi_{i+1} = \left(\tfrac{1}{h}, 0\right) \\ \text{on } T_2: & \nabla\phi_i = \left(-\tfrac{1}{h}, 0\right) & \nabla\phi_{i+1} = \left(\tfrac{1}{h}, -\tfrac{1}{h}\right) \end{array}$$

and thus

$$\begin{aligned} A_{i,i+1} &= \int_\Omega \nabla\phi_i \cdot \nabla\phi_{i+1} \, d\Omega = \textstyle\sum_{k=1}^{2} \int_{T_k} \nabla\phi_i \cdot \nabla\phi_{i+1} \, d\Omega \\ &= -\tfrac{1}{h^2}|T_1| - \tfrac{1}{h^2}|T_2| = -\tfrac{2}{h^2}\tfrac{h^2}{2} = -1. \end{aligned}$$

Similar calculations yield that

$$A_{i,i-1} = A_{i,i+J} = A_{i,i-J} = -1, \qquad A_{i,i+J+1} = A_{i,i-J-1} = 0.$$

Also note that the matrix A is *symmetric*: $A_{ij} = A_{ji}$. Altogether, we obtain the *block triangular structure* (empty space means zeros!)

$$A = \begin{pmatrix} T & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T \end{pmatrix}$$

where T and I are the J-by-J matrices

$$T = \begin{pmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 1. \end{pmatrix},$$

Thus, the ith row of the matrix-vector product $A\boldsymbol{u}$ will be

$$4u_i - u_{i+1} - u_{i-1} - u_{i+J} - u_{i-J}. \tag{10.18}$$

Node $i+1$ and $i-1$ is located to the right and left, respectively, of node i, whereas nodes $i+J$ and $i-J$ are above and below node i. Thus, expression (10.18) is precisely the classical five-point, finite-difference formula. We reach the remarkable conclusion that the finite-element discretisation of the Laplace operator using continuous, piecewise-linear functions on the structured mesh of figure 10.6 reduces to a standard finite-difference formula for the Laplacian. Note, however, that this does not hold in general; finite-element discretisations are not always easy to interpret as a finite-difference method.

10.9 Properties of the Stiffness Matrix

Consider the stiffness matrix A with components

$$A_{ij} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \, d\Omega,$$

which was obtained by discretising the Poisson problem (10.1). This matrix has some very particular properties, which will be discussed in this section: it is *symmetric*, *positive definite*, *sparse*, and *ill conditioned*. All these properties, except the sparsity, reflects the nature of the boundary-value problem (10.1). Some or all of these properties may change if the equation or the boundary conditions are altered. For instance, if an additional term containing first derivatives of u is added to equation (10.1), the stiffness matrix will no longer be symmetric. The sparsity is a consequence of the fact that the chosen piecewise-linear approximations allow a compact basis, the "tent" functions of figure 10.5.

The symmetry of the matrix is immediate,

$$A_{ij} = \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \, d\Omega = \int_\Omega \nabla\phi_j \cdot \nabla\phi_i \, d\Omega = A_{ji}.$$

Moreover, the matrix is *sparse*, since $A_{ij} = 0$ whenever i and j are not nearest neighbours. The number of neighbours to each point does not increase when the mesh is made finer, as long as the mesh refinements are made in a sensible way, see the discussion in section 10.10. Thus, the number of nonzero elements on each row does not increase with the order of the stiffness matrix, that is, the matrix in a sense becomes sparser and sparser with increasing matrix order.

Recall that a real matrix A is *positive definite* if $\boldsymbol{v}^T A \boldsymbol{v} > 0$ whenever $\boldsymbol{v} \neq 0$.

Theorem 41 *The stiffness matrix is positive definite.*

Proof
Let $v_h \in V_h$. Expanding v_h in the "tent" basis functions yields

$$v_h = \sum_{i=1}^{N} v_i \phi_i(\boldsymbol{x}).$$

Setting

$$\boldsymbol{v} = (v_1, v_2, \ldots, v_N)^T,$$

yields that

$$\begin{aligned} \boldsymbol{v}^T A \boldsymbol{v} &= \textstyle\sum_{i=1}^N \sum_{j=1}^N v_i \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \, d\Omega v_j \\ &= \textstyle\int_\Omega \underbrace{\displaystyle\sum_{i=1}^N \nabla(v_i\phi_i)}_{=\nabla v_h} \cdot \underbrace{\displaystyle\sum_{j=1}^N \nabla(v_j\phi_j)}_{=\nabla v_h} \, d\Omega = \textstyle\int_\Omega |\nabla v_h|^2 \, d\Omega \geq 0, \end{aligned} \tag{10.19}$$

with equality if and only if $\nabla v_h = 0$, that is, if v_h is constant. However, since v_h is zero on the boundary (by definition of V_h), it follows that the constant must be zero. Thus expression (10.19) is zero only if $v_h \equiv 0$, that is, when $\boldsymbol{v} = \boldsymbol{0}$. Q.E.D.

One important consequence of Theorem 41 is that equation (10.17) has a unique solution. This follows from the fact that positive-definite matrices are nonsingular: For a *singular* matrix A, there would be nonzero vector $\boldsymbol{v}$ so that $A\boldsymbol{v} = \mathbf{0}$, and thus $\boldsymbol{v}^T A\boldsymbol{v} = 0$. Thus, singular matrices cannot be positive definite, and positive-definite matrices must therefore be nonsingular.

The condition number of the stiffness matrix depends strongly on h. In fact, if the quotient between the size of the smallest and largest triangle in the mesh is kept bounded as the mesh is refined, one can show that the condition number grows like $\mathrm{cond}(A) = \mathcal{O}(h^{-2})$. The stiffness matrix is thus ill conditioned for fine meshes.

10.10 Accuracy

We have shown how to define a finite-element approximation of the Poisson problem (10.1), that this yields the linear system (10.17), and that this has a unique solution. The question how good the finite-element solution is as an approximation of the original problem will be discussed in this section.

For finite-difference discretisations, accuracy questions are usually addressed indirectly by applying the Lax–Richtmyer Theorem. The crucial steps is then to derive truncation errors and to check stability. If the method is consistent, that is, if the truncation error vanishes as the mesh is refined, the method is convergent if and only if it is stable. For finite-element discretisations, this approach is hardly ever used, since it is possible to study the error in the discretisation directly. The easiest and most natural way is to work with *integral norms* of the difference between the weak solution u of problem (10.8) and the finite-element solution u_h of problem (10.14). The $L^2(\Omega)$ norm of a function,

$$|v|_{L^2(\Omega)} = \left(\int_\Omega v^2 \, d\Omega \right)^{1/2},$$

is the analogue for functions of the vector 2-norm. The perhaps most important norm for solutions of the Poisson problem is the *energy norm*

$$|v|_V = \left(\int_\Omega |\nabla v|^2 \, d\Omega \right)^{1/2}, \tag{10.20}$$

that is, the $L^2(\Omega)$-norm of the first derivatives; recall that weak solutions were defined among functions with bounded energy norm (the space of functions defined by (10.7)). The importance of the energy norm is that the finite-element solution is *optimal* in the energy norm. That is, no other function in V_h yields a smaller error in energy norm:

Theorem 42 *Let u be the solution to variational problem (10.8) and u_h the finite-element solution (10.14). Then*

$$|u - u_h|_V \leq |u - v_h|_V \qquad \forall v_h \in V_h, \tag{10.21}$$

Proof
By equation (10.14), the finite-element solution u_h satisfies

$$\int_\Omega \nabla u_h \cdot \nabla v_h \, d\Omega = \int_\Omega f v_h \, d\Omega \qquad \forall v_h \in V_h. \tag{10.22}$$

From equation (10.8) follows that the weak solution u satisfies

$$\int_\Omega \nabla u \cdot \nabla v_h \, d\Omega = \int_\Omega f v_h \, d\Omega \qquad \forall v_h \in V_h, \tag{10.23}$$

since $V_h \subset V$. Subtracting equations (10.22) and (10.23) yields that

$$\int_\Omega \nabla(u - u_h) \cdot \nabla v_h \, d\Omega = 0 \qquad \forall v_h \in V_h. \tag{10.24}$$

Let v_h be an arbitrary element of V_h. Then

$$\begin{aligned} |u-u_h|_V^2 &= \textstyle\int_\Omega |\nabla(u-u_h)|^2 \, d\Omega = \int_\Omega [\nabla(u-u_h)] \cdot [\nabla(u-u_h)] \, d\Omega \\ &= \textstyle\int_\Omega \nabla(u-u_h) \cdot \nabla u \, d\Omega - \underbrace{\int_\Omega \nabla(u-u_h) \cdot \nabla u_h \, d\Omega}_{= 0 \text{ by } (10.24)} \\ &= \textstyle\int_\Omega \nabla(u-u_h) \cdot \nabla u \, d\Omega - \underbrace{\int_\Omega \nabla(u-v_h) \cdot \nabla v_h \, d\Omega}_{= 0 \text{ by } (10.24)} \\ &= \textstyle\int_\Omega \nabla(u-u_h) \cdot \nabla(u-v_h) \, d\Omega \le |u-u_h|_V |u-v_h|_V, \end{aligned} \tag{10.25}$$

where the last inequality follows from the Cauchy–Schwarz inequality. Dividing through with $|u - u_h|_V$ yields the conclusion. Q.E.D.

The optimality property (10.21) does not hold for all elliptic boundary-value problems. For the finite-element solution to be optimal, it is necessary that the variational problem yields a *symmetric* stiffness matrix.

The next step in an analysis of the error is a pure approximation problem. Typically, one considers the *interpolant*, that is, a piecewise-linear function agreeing with u at the node points; note that the interpolant is an element of V_h. The difference between the interpolant and u can be estimated by a type of Taylor expansion. From Theorem 42 follows that the error in the finite-element solution smaller or equal to the error in the interpolant. The precise magnitude of this error depends of course on how fine the mesh is, but it also depends on the *quality* of the mesh. Loosely speaking, one should avoid very thin triangles.

Altogether, estimating the interpolation error and utilising Theorem 42, it can be shown that the error in the finite-element solution is of *second order*, that is,

$$|u_h - u|_{L^2(\Omega)} = O(h^2). \tag{10.26}$$

Note that the norm above is not the energy norm; the error is of *first order* if measured in the energy norm. For estimate (10.26) to hold, assumptions have to be made on the mesh quality and on the smoothness of the solution to the variational problem (10.8). Following conditions are sufficient.

(i) (Mesh quality.) The smallest angle of any of the triangles is bounded below as the mesh is refined. This means that no triangle successively can become infinitely thin.

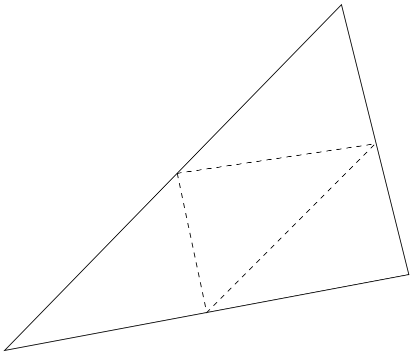


Figure 10.12: A strategy to maintain mesh quality is to subdivide each triangle into four new triangles by joining the edge midpoints.

(ii) (Smoothness.) The boundary of Ω is smooth. Alternatively, the boundary is polygonal and the domain is convex. (If Ω is not polygonal to start with, it is typically approximated with a succession of polygonal domains Ω_h such that $\Omega_h \to \Omega$ as $h \to 0$).

The mesh quality condition above is maintained if the triangles, as the mesh is refined, are subdivided into four triangles in the way indicated in figure 10.12. Refining each triangle in the mesh in this way reduces all triangular sides with a factor 1/2. The error will thus be reduced with a factor 1/4 (for problems on convex domains at least).

Higher accuracy can thus be obtained through refinement of the mesh ("h method"). This should preferably be done *adaptively*, in the parts of the domain where it is needed, to prevent the size of the stiffness matrix to become too large. There are automatic methods for this. Higher accuracy can also be obtained through higher order on the polynomials on each triangle ("p method"). For instance, the error in the sense (10.26) can be improved to *third order* if V_h consists of continuous functions that are *quadratic* on each element.

10.11 Alternative Elements

Quadrilaterals, that is, a geometric figure obtained by connecting four points in the plane by straight lines, can be used to partition the domain instead of triangles, see figure 10.13. In this case will the approximating space V_h contain globally continuous functions who vary linearly along the edges of each quadrilateral. However, the functions will no longer be linear *within* the elements. In the special case when the quadrilaterals are rectangles oriented in the coordinate directions, a function $v_h \in V_h$ will be *bilinear*, that is, of the form

$$v_h(x, y) = a + bx + cy + dxy$$

on each element. The nodal values of v_h (the values of v_h at the four corners of the rectangle) uniquely determine the four coefficients above.

Quadrilaterals and, in particular, rectangular elements yields a regular structure that may give high solution accuracy and allow efficient solutions of the associated linear

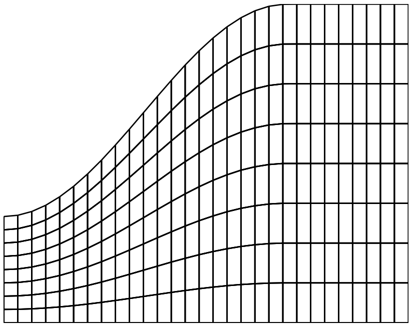


Figure 10.13: A quadrilateral mesh.

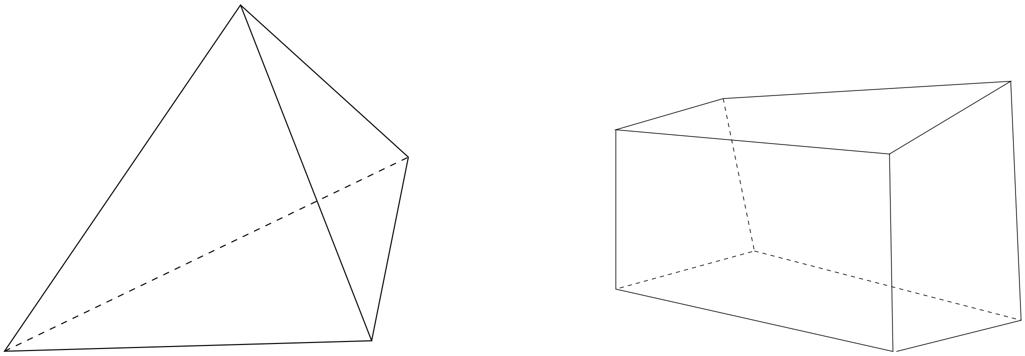


Figure 10.14: Meshes in three space dimensions can be composed of non-overlapping tetrahedra (left) or hexahedra (right).

systems. It is, however, harder to generate such meshes automatically on complicated geometries compared to triangular meshes.

For three space dimensions, triangular and quadrilateral meshes generalise to *tetrahedral* and *hexahedral* meshes (figure 10.14) with advantages and limitations as for corresponding meshes in two space dimensions.

For higher order equations, such as the Euler-Bernoulli model of a bending beam

$$y''''(x) = -f(x)$$

or the biharmonic equation

$$\Delta^2 u = 0$$

it is necessary to use *smoother* elements for which the derivatives are continuous continuous across element boundaries. So, for example, in the one-dimensional beam bending problem it is possible to define $v_h \in V_h$ as *Hermite* piecewise cubic polynomials:

$$\left.\begin{array}{rcl} \phi_0(s) & = & (s-1)^2(2s+1) \\ \phi_1(s) & = & s^2(3-2s) \\ \varphi_0(s) & = & (s-1)^2 s \\ \varphi_1(s) & = & s^2(s-1) \end{array}\right\}$$

to define a cubic polynomial, $p(s)$, $s \in [0,1]$ that interpolates $v(0)$, $\frac{dv(0)}{ds}$, $v(1)$ and $\frac{dv(1)}{ds}$ can be written as

$$p(s) = v(0)\phi_0(s) + v(1)\phi_1(s) + \left(\frac{dv(0)}{ds}\right)\varphi_0(s) + \left(\frac{dv(1)}{ds}\right)\varphi_1(s).$$

The nodal values of v_h are the values of the solution at the nodes and the values of the derivative at the node.

Exercises

Exercise 10.1 *Construct the weak form of the b.v.p.*

$$\begin{array}{rcll} \nabla(a\nabla u) & = & f, & x \in \Omega \subset \mathbb{R}^3 \\ u & = & g, & x \in \Gamma_1 \\ a\partial_n u & = & h, & x \in \Gamma_2 \end{array}$$

where $\Gamma_1 \cup \Gamma_2 = \Gamma$ *the boundary of* Ω *and* $\Gamma_1 \cap \Gamma_2 = \emptyset$.

Exercise 10.2 *Given that* u *is the solution of* $\Delta u = f$ *in* $\Omega = (0,1) \times (-1,1)$*, and on the boundary* Γ*, subject to* $u = \left\{\begin{array}{rl} 1-y^2, & -1 \le y \le 1,\ x=0 \\ y^2, & -1 \le y \le 1,\ x=1 \\ x, & 0 \le x \le 1,\ y=\pm 1 \end{array}\right.$

1. *Derive the variational formulation of this problem.*

2. *Using a uniform grid of triangles and piecewise linear basis functions,* $\phi_i(x,y)$, $i = 1,\ldots,N$ *derive the system of equations* $A\boldsymbol{U} = \boldsymbol{b}$ *that defines the approximate solution*

$$U(x,y) = \sum_{j=1}^{N} U_j \phi_j(x,y)$$

where $\boldsymbol{U} = (U_1,\ldots,U_N)^T$ *in the case when* $f(x,y) = 0$.

Bibliography

[1] R. Courant, K.O Friedrichs, and H. Lewy, *Über die partiellen differenzengleichungen der mathematischen physik*, Math. Anal. **100** (1928), 32–74.

[2] J. Crank and P. Nicolson, *A practical method for numerical integration of solutions of partial differential equations of heat-conduction type*, Proc. Cam. Phil. Soc. **43** (1947), 50–67.

[3] Paul DuChateau, *Introduction to Partial Differential Equations*, `http://www.math.colostate.edu/~pauld/M545.html`, [Online; accessed 22-March-2007].

[4] M. Frigo and S.G. Johnson, *The fastest fourier transform in the west*, Tech. Report MIT-LCS-TR-728, MIT Laboratory for Computer Science, 1997.

[5] Mark S. Gockenbach, *Understanding and Implementing the Finite Element Method*, SIAM, Philadelphia, 2006, ISBN 0-89871-614-4.

[6] Randall J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, CUP, Cambridge, 2002.

[7] Courant R. and Hilbert D., *Methods of Mathematical Physics. Vol. 2, partial differential equations*, Wiley, 1989.

[8] G.D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, OUP, Oxford, 1978, 2nd Edition.

[9] Gigliola Staffilani and Andras Vasy, *Introduction to Partial Differential Equations*, `http://ocw.mit.edu/OcwWeb/Mathematics/18-152Fall-2004/`, [Online; accessed 22-March-2007].

[10] John C. Strikwerda, *Finite Difference Schemes and Partial Differential Equations*, SIAM, Philadelphia, 2004, 2nd edition.

[11] Wikipedia, *Partial Differential Equations*, `http://en.wikipedia.org/wiki/Partial_differential_equations`, [Online; accessed 22-March-2007].

[12] ———, *Rankine-Hugoniot equation*, `http://en.wikipedia.org/wiki/Rankine-Hugoniot_equation`, [Online; accessed 23-June-2008].

[13] Wen-Chyuan Yueh, *Eigenvalues of several tridiagonal matrices*, Applied Mathematics E-Notes (2005), 66–74, [Online; accessed 16-Oct-2008].

Index