Numerical Methods for Solving Partial Differential Equations

Yuexin Liao

October 2023

1 Introduction

Partial Differential Equations, occur frequently in mathematics, natural science and engineering. They are used in many problems, which involve rates of change of functions of several variables. The following involve 2 independent variables:

$$-\nabla^2 u = -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f(x,y) \quad \text{Poisson Equation}$$

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0 \quad \text{Advection Equation}$$

$$\frac{\partial u}{\partial t} - D\frac{\partial^2 u}{\partial x^2} = 0 \quad \text{Heat Equation}$$

$$\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = 0 \quad \text{Wave Equation}$$

Here v, D, c are real positive constants. In these cases, x, y are the space coordinates and t, x are often viewed as time and space coordinates, respectively.

Note. *These are only examples and do not cover all cases. In real scenarios, PDEs usually have 3 or 4 variables.*

2 PDE Classification

PDE is in two independent variables x and y have the form

$$\Phi\left(x, y, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial^2 u}{\partial x^2}, ...\right) = 0$$

where the symbol Φ stands for some functional relationship.

Note. *As we saw with BVP, this is a too general case. Hence, we must define new classes of the general PDE.*

Definition. *The order of a PDE is the order of the highest derivative that appears, i.e., Poisson is 2nd order, Advection eqn is 1st order.*

Most of the mathematical theory of PDE's concerns linear equations of first or second order. After order and linearity (linear or non-linear), the most important classification scheme for PDE involves geometry. We introduce the ideas with the following example.

Example.

$$\alpha(t,x)\frac{\partial u}{\partial t}+\beta\frac{\partial u}{\partial x}=\gamma(t,x) \tag{1}$$

A solution $u(t,x)$ to this PDE defines a surface $\{t,x,u(t,x)\}$ lying over some region of the (t,x)-plane. Consider any smooth path in the (t,x)-plane lying below the solution $\{t,x,u(t,x)\}$.
Such a path has a parameterization $(t(s),x(s))$, where the parameter s measures progress along the path. The rate of change $\frac{du}{ds}$ of the solution is as we travel along the path $(t(s),x(s))$.
The chain rule provides the answer:

$$\frac{dt}{ds}\frac{\partial u}{\partial t}+\frac{dx}{ds}\frac{\partial u}{\partial x}=\frac{du}{ds} \tag{2}$$

Equation (2) holds for an arbitrary smooth path in the (t,x)-plane. Restricting attention to a specific family of paths leads to a useful observation:

$$\frac{dt}{ds}=\alpha(t,x)\ \textit{and}\ \frac{dx}{ds}=\beta(t,x) \tag{3}$$

When the simultaneous validity of (1) and (2) requires that

$$\frac{du}{ds}=\gamma(t,x). \tag{4}$$

Equation (4) defines a family of curves $(t(s),x(s))$. It is called characteristic curves in the plane (t,x).
Equation (4) is an ode called the characteristic equation that the solution must satisfy along only the characteristic curve.
Thus, the original PDE collapses to an ODE along the characteristic curves. Characteristic curves are paths along which information about the solution to the PDE propagates from points where the initial value or boundary values are known.

Then, we consider a second order PDE having the form

$$\alpha(x,y)\frac{\partial^2 u}{\partial x^2}+\beta(x,y)\frac{\partial^2 u}{\partial x\partial y}+\gamma(x,y)\frac{\partial^2 u}{\partial y^2}=\Psi(x,y,u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y}) \tag{5}$$

Along an arbitrary smooth curve $(x(s),y(s))$ in the (x,y)-plane, the gradient $\left(\frac{\partial u}{\partial x},\frac{\partial u}{\partial y}\right)$ of the solution varies according to the chain rule:

$$\frac{dx}{ds}\frac{\partial^2 u}{\partial y\partial x}+\frac{dy}{ds}\frac{\partial^2 u}{\partial y\partial x}=\frac{d}{ds}\left(\frac{\partial u}{\partial x}\right)$$
$$\frac{dx}{ds}\frac{\partial^2 u}{\partial x\partial y}+\frac{dy}{ds}\frac{\partial^2 u}{\partial y^2}=\frac{d}{ds}\left(\frac{\partial u}{\partial y}\right)$$

If the solution $u(x,y)$ is continuously differentiable, then these relationships together with the original PDE yield the following system:

$$\begin{pmatrix}\alpha & \beta & \gamma\\ \frac{dx}{ds} & \frac{dy}{ds} & 0\\ 0 & \frac{dx}{ds} & \frac{dy}{ds}\end{pmatrix}\begin{pmatrix}\frac{\partial^2 u}{\partial x^2}\\ \frac{\partial^2 u}{\partial x\partial y}\\ \frac{\partial^2 u}{\partial y^2}\end{pmatrix}=\begin{pmatrix}\Psi\\ \frac{d}{ds}\left(\frac{\partial u}{\partial x}\right)\\ \frac{d}{ds}\left(\frac{\partial u}{\partial y}\right)\end{pmatrix} \tag{6}$$

By analogy with the first order case, we determine the characteristic curves where the PDE is redundant with the chain rule. This occurs when the determinant of the matrix in (6) vanishes that is when

$$\alpha\left(\frac{dy}{ds}\right)^2-\beta\left(\frac{dy}{ds}\right)\left(\frac{dx}{ds}\right)+\gamma\left(\frac{dx}{ds}\right)^2=0$$

By eliminating the parameter s, we reduce this equation to the equivalent condition

$$\alpha\left(\frac{dy}{dx}\right)^2-\beta\left(\frac{dy}{dx}\right)+\gamma=0$$

Formally solving this quadratic for $\frac{dy}{dx}$, we find

$$\frac{dy}{dx} = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\alpha}$$

This pair of ODE's determine the characteristic curves. From this equation, we divide into 3 classes each defined with respect to $\beta^2 - 4\alpha\gamma$.

1. **HYPERBOLIC**
   $\beta^2 - 4\alpha\gamma > 0$ This gives two families of real characteristic curves.
2. **PARABOLIC**
   $\beta^2 - 4\alpha\gamma = 0$ This gives exactly one family of real characteristic curves.
3. **ELLIPTIC**
   $\beta^2 - 4\alpha\gamma < 0$ This gives no real characteristic equations.

Example. ***The wave equation***

$$c^2 \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0$$

By equating this with our formula for the characteristics, we have

$$\frac{dt}{dx} = \frac{0 \pm \sqrt{0 + 4c^2}}{2} = \pm c$$

This implies that the characteristics are $x + ct = const$ and $x - ct = const$. This means that the effects travel along the characteristics.

Example. ***Laplace equation***

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

From this, we have $-4(1)(1) < 0$ which implies it is elliptic.
This means that information at one point affects all other points.

Example. ***Heat equation***

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0$$

From this, we have $\beta^2 - 4\alpha\gamma = 0$. This implies that the equation is parabolic. Thus, we have:

$$\frac{\partial t}{\partial x} = 0$$

We can also state that hyperbolic and parabolic are Boundary value problems and initial value problems while elliptic problems are boundary value problems.

3 Difference Operators

Through out this part, we will use $U(x_i)$ to denote the exact solution and U_i to denote the numerical (approximate) solution.
For 1-D difference operators, we have:

$$D^+ U_i = \frac{U_{i+1} - U_i}{h_{i+1}}, \quad \text{Forward},$$

$$D^- U_i = \frac{U_i - U_{i-1}}{h_i}, \quad \text{Backward},$$

$$D^0 U_i = \frac{U_{i+1} - U_{i-1}}{x_{i+1} - x_{i-1}}, \quad \text{Centered}.$$

The 2-D Differences Schemes are similar when dealing with the x-direction. We hold the y-direction constant as well as the x-direction constant and then deal with the y-direction.

$$
\begin{aligned}
D_x^+ U_{ij} &= \frac{U_{i+1j} - U_{ij}}{x_{i+1} - x_i}, \quad \text{Forward in the x-direction} \\
D_y^+ U_{ij} &= \frac{U_{ij+1} - U_{ij}}{y_{j+1} - y_j}, \quad \text{Forward in the y-direction} \\
D_x^- U_{ij} &= \frac{U_{ij} - U_{i-1j}}{x_i - x_{i-1}}, \quad \text{Backward in the x direction} \\
D_y^- U_{ij} &= \frac{U_{ij} - U_{ij-1}}{y_j - y_{j-1}}, \quad \text{Backward in the y direction} \\
D_x^0 U_{ij} &= \frac{U_{i+1j} - U_{i-1j}}{x_{i+1} - x_{i-1}}, \quad \text{Centered in the x direction} \\
D_y^0 U_{ij} &= \frac{U_{ij+1} - U_{ij-1}}{y_{j+1} - y_{j-1}}, \quad \text{Centered in the y direction}
\end{aligned}
$$

The second derivatives approximations in the x and y direction are:

$$
\begin{aligned}
\delta_x^2 U_{ij} &= \frac{2}{x_{i+1} - x_{i-1}} \left(\frac{U_{i+1j} - U_{ij}}{x_{i+1} - x_i} - \frac{U_{ij} - U_{i-1j}}{x_i - x_{i-1}} \right), \quad \text{Centered in } x \text{ direction}, \\
\delta_y^2 U_{ij} &= \frac{2}{y_{j+1} - y_{j-1}} \left(\frac{U_{ij+1} - U_{ij}}{y_{j+1} - y_j} - \frac{U_{ij} - U_{ij-1}}{y_j - y_{j-1}} \right), \quad \text{Centered in } y \text{ direction}.
\end{aligned}
$$

The Poisson equation is,

$$
-\nabla^2 U(x,y) = f(x,y), \quad (x,y) \in \Omega = (0,1) \times (0,1), \tag{7}
$$

where ∇ is the Laplacian,

$$
\nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2},
$$

with boundary conditions,

$$
U(x,y) = g(x,y), \quad (x,y) \in \delta\Omega\text{-boundary}.
$$

4 Elliptic equation

4.1 The five point approximation of the Laplacian

To numerically approxiamte the solution of the Poisson Equation (7), the unit square region $\bar{\Omega} = [0,1] \times [0,1] = \Omega \bigcup \partial\Omega$ must be discretised into a uniform grid.

$$
\triangle = \{(x_i, y_j) \in [0,1] \times [0,1] : x_i = ih, y_j = jh\}
$$

for $i = 0, 1, .., N$ and $i = 0, 1, .., N$, where N is a positive constant. The interior nodes of the grid are defined as:

$$
\Omega_h = \{(x_i, y_j) \in \triangle : 1 \leq i, j \leq N - 1\},
$$

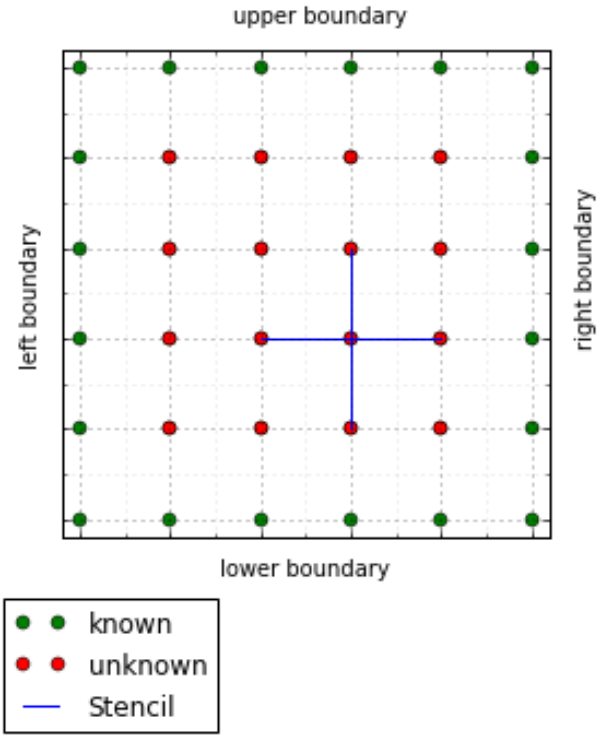
the boundary nodes are

$$
\partial\Omega_h = \{(x_0, y_j), (x_N, y_j), (x_i, y_0), (x_i, y_N) : 1 \leq i, j \leq N - 1\}.
$$

The Poisson Equation (7) is discretised using δ_x^2 the central difference approximation of the second derivative in the x direction

$$
\delta_x^2 = \frac{1}{h^2}(w_{i+1j} - 2w_{ij} + w_{i-1j}),
$$

Figure 1: Graphical representation of the difference equation stencil



and δ_y^2 the central difference approximation of the second derivative in the y direction

$$\delta_y^2 = \frac{1}{h^2}(w_{ij+1} - 2w_{ij} + w_{ij-1}).$$

The gives the Poisson Difference Equation,

$$\begin{aligned} -\nabla_h w_{ij} &= f_{ij} \quad (x_i, y_j) \in \Omega_h, & (8)\\ -(\delta_x^2 w_{ij} + \delta_y^2 w_{ij}) &= f_{ij} \quad (x_i, y_j) \in \Omega_h, & (9)\\ w_{ij} &= g_{ij} \quad (x_i, y_j) \in \partial\Omega_h, & (10) \end{aligned}$$

where w_{ij} is the numerical approximation of U at x_i and y_j. Expanding the the Poisson Difference Equation (10) gives the five point method,

$$-(w_{i-1j} + w_{ij-1} - 4w_{ij} + w_{ij+1} + w_{i+1j}) = h^2 f_{ij}$$

for $i = 1, ..., N-1$ and $j = 1, ..., N-1$, which is depicted in Figure 1 on a 6×6 discrete grid.

Unlike the Parabolic equation, the Elliptic equation cannot be estimated by holding one variable constant and then stepping in that direction. The approximation must be solved at all points at the same instant.

4.1.1 Matrix representation of the five point scheme

The five point scheme results in a system of $(N-1)^2$ equations for the $(N-1)^2$ unknowns. This is depicted in Figure 1 on a $6 \times 6 = 36$ where there is a grid of $4 \times 4 = 16$ unknowns (red) surround by the boundary of 20 known values. The general set of 4×4 equations of the Poisson difference equation on the 6×6 grid where

$$h = \frac{1}{6-1} = \frac{1}{5},$$

can be written as:

$$
\begin{array}{l|lcl}
j=1 & & & \\
i=1 & w_{0,1}+w_{1,0}-4w_{1,1}+w_{1,2}+w_{2,1} & = & \frac{1}{5}^2 f_{11} \\
i=2 & w_{1,1}+w_{2,0}-4w_{2,1}+w_{2,2}+w_{3,1} & = & \frac{1}{5}^2 f_{21} \\
i=3 & w_{2,1}+w_{3,0}-4w_{3,1}+w_{3,2}+w_{4,1} & = & \frac{1}{5}^2 f_{31} \\
i=4 & w_{3,1}+w_{4,0}-4w_{4,1}+w_{4,2}+w_{5,1} & = & \frac{1}{5}^2 f_{41} \\
\hline
j=2 & & & \\
i=1 & w_{0,2}+w_{1,1}-4w_{1,2}+w_{1,3}+w_{2,2} & = & \frac{1}{5}^2 f_{12} \\
i=2 & w_{1,2}+w_{2,1}-4w_{2,2}+w_{2,3}+w_{3,2} & = & \frac{1}{5}^2 f_{22} \\
i=3 & w_{2,2}+w_{3,1}-4w_{3,2}+w_{3,3}+w_{4,2} & = & \frac{1}{5}^2 f_{32} \\
i=4 & w_{3,2}+w_{4,1}-4w_{4,2}+w_{4,3}+w_{5,2} & = & \frac{1}{5}^2 f_{42} \\
\hline
j=3 & & & \\
i=1 & w_{0,3}+w_{1,2}-4w_{1,3}+w_{1,4}+w_{2,3} & = & \frac{1}{5}^2 f_{13} \\
i=2 & w_{1,3}+w_{2,2}-4w_{2,3}+w_{2,4}+w_{3,3} & = & \frac{1}{5}^2 f_{23} \\
i=3 & w_{2,3}+w_{3,2}-4w_{3,3}+w_{3,4}+w_{4,3} & = & \frac{1}{5}^2 f_{33} \\
i=4 & w_{3,3}+w_{4,2}-4w_{4,3}+w_{4,4}+w_{5,3} & = & \frac{1}{5}^2 f_{43} \\
\hline
j=4 & & & \\
i=1 & w_{0,4}+w_{1,3}-4w_{1,4}+w_{1,5}+w_{2,4} & = & \frac{1}{5}^2 f_{14} \\
i=2 & w_{1,4}+w_{2,3}-4w_{2,4}+w_{2,5}+w_{3,4} & = & \frac{1}{5}^2 f_{24} \\
i=3 & w_{2,4}+w_{3,3}-4w_{3,4}+w_{3,5}+w_{4,4} & = & \frac{1}{5}^2 f_{34} \\
i=4 & w_{3,4}+w_{4,3}-4w_{4,4}+w_{4,5}+w_{5,4} & = & \frac{1}{5}^2 f_{44}.
\end{array}
$$

This set of equations can be rearranged by bringing the known boundary conditions $w_{0,j}$, $w_{5,j}$, $w_{i,0}$ and $w_{i,5}$, to the right hand side. This can be written as a 16×16 Matrix equation of the form:

$$
\left(\begin{array}{cccc|cccc|cccc|cccc}
-4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 0 & 0 & 0 & 1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -4 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -4
\end{array}\right) =
$$

$$
\left(\begin{array}{c}
w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{4,1} \\ \hline
w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{4,2} \\ \hline
w_{1,3} \\ w_{2,3} \\ w_{3,3} \\ w_{4,3} \\ \hline
w_{1,4} \\ w_{2,4} \\ w_{3,4} \\ w_{4,4}
\end{array}\right) = -\frac{1}{5}^2 \left(\begin{array}{c}
f_{1,1} \\ f_{2,1} \\ f_{3,1} \\ f_{4,1} \\ \hline
f_{1,2} \\ f_{2,2} \\ f_{3,2} \\ f_{4,2} \\ \hline
f_{1,3} \\ f_{2,3} \\ f_{3,3} \\ f_{4,3} \\ \hline
f_{1,4} \\ f_{2,4} \\ f_{3,4} \\ f_{4,4}
\end{array}\right) + \left(\begin{array}{c}
-w_{1,0} \\ -w_{2,0} \\ -w_{3,0} \\ -w_{4,0} \\ \hline
0 \\ 0 \\ 0 \\ 0 \\ \hline
0 \\ 0 \\ 0 \\ 0 \\ \hline
-w_{1,4} \\ -w_{2,4} \\ -w_{3,4} \\ -w_{4,4}
\end{array}\right) + \left(\begin{array}{c}
-w_{0,1} \\ 0 \\ 0 \\ -w_{5,1} \\ \hline
-w_{0,2} \\ 0 \\ 0 \\ -w_{5,2} \\ \hline
-w_{0,3} \\ 0 \\ 0 \\ -w_{5,3} \\ \hline
-w_{0,4} \\ 0 \\ 0 \\ -w_{5,4}
\end{array}\right)
$$

The horizontal and vertical lines are for display purposes to help indicated each set of the four sets of four equations.

4.1.2 Generalised Matrix form of the discrete Poisson Equation

The generalized form of this matrix of the system of equations for the parabolic case results in $(N-1)$ equations, that are written as an $(N-1)^2 \times (N-1)^2$ square matrix A and the $(N-1)^2 \times 1$ vectors $\mathbf{w}$, $\mathbf{r}$ and $\mathbf{b}$:

$$A\mathbf{w} = -h\mathbf{r} + \mathbf{b}.$$

The matrix can be written as following block tridiagonal structure :

$$
\left(\begin{array}{ccccccc}
T & I & 0 & 0 & . & . & . \\
I & T & I & 0 & 0 & . & . \\
0 & . & . & . & 0 & . & . \\
. & . & . & . & . & . & . \\
. & . & . & 0 & I & T & I \\
. & . & . & . & 0 & I & T
\end{array}\right)
\left(\begin{array}{c}
\mathbf{w}_1 \\ \mathbf{w}_2 \\ . \\ . \\ \mathbf{w}_{N-2} \\ \mathbf{w}_{N-1}
\end{array}\right) = -h^2 \left(\begin{array}{c}
\mathbf{f}_1 \\ \mathbf{f}_2 \\ . \\ . \\ \mathbf{f}_{N-2} \\ \mathbf{f}_{N-1}
\end{array}\right) + \left(\begin{array}{c}
\mathbf{b}_1 \\ \mathbf{b}_2 \\ . \\ . \\ \mathbf{b}_{N-2} \\ \mathbf{b}_{N-1}
\end{array}\right),
$$

where I denotes an $(N-1)\times(N-1)$ identity matrix and T is an $(N-1)\times(N-1)$ tridiagonal matrix of the form:

$$T = \begin{pmatrix} -4 & 1 & 0 & 0 & . & . & . \\ 1 & -4 & 1 & 0 & 0 & . & . \\ 0 & . & . & . & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & 0 & 1 & -4 & 1 \\ . & . & . & . & 0 & 1 & -4 \end{pmatrix},$$

$\mathbf{w}_j$ is an $(N-1)\times 1$ vector of approximations w_{ij},

$$\mathbf{w}_j = \begin{pmatrix} w_{1j} \\ w_{2j} \\ . \\ . \\ w_{N-2j} \\ w_{N-1j} \end{pmatrix}$$

the vector $\mathbf{f}$ is made up of $(N-1)$ vectors of length $(N-1)\times 1$,

$$\mathbf{f}_j = \begin{pmatrix} f_{1j} \\ f_{2j} \\ . \\ . \\ f_{N-2j} \\ f_{N-1j} \end{pmatrix},$$

finally $\mathbf{b}$ is the vector of boundary conditions made up of two $(N-1)$ vectors of length $(N-1)\times 1$,

$$\mathbf{b}_j = \mathbf{b}_{left,right,j} + \mathbf{b}_{top,bottom,j} = -\begin{pmatrix} g_{0j} \\ 0 \\ . \\ . \\ 0 \\ g_{Nj} \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ 0 \end{pmatrix}$$

for $j = 2, .., N-2$, for $j = 1$ and $j = N-1$

$$\mathbf{b}_1 = -\begin{pmatrix} g_{10} \\ 0 \\ . \\ . \\ 0 \\ g_{1N} \end{pmatrix} - \begin{pmatrix} g_{10} \\ g_{20} \\ . \\ . \\ g_{N-20} \\ g_{1N} \end{pmatrix}, \quad \mathbf{b}_{N-1} = -\begin{pmatrix} g_{0N-1} \\ 0 \\ . \\ . \\ 0 \\ g_{NN-1} \end{pmatrix} - \begin{pmatrix} g_{1N} \\ g_{2N} \\ . \\ . \\ g_{N-2N} \\ g_{N-1N} \end{pmatrix}.$$

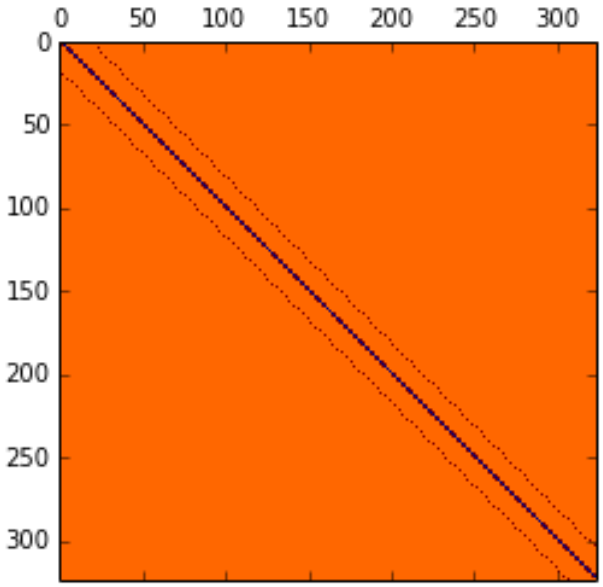
The matrix has a unique solution. For sparse matrices of this form an iterative method is used as it would be to computationally expensive to compute the inverse.

4.1.3 Specific Examples

Our goal in this part is to delve into these three example problems:

1. Homogenous form of the Poisson Equation (Lapalacian),
2. Poisson Equation with zero boundary conditions,
3. Poisson Equation with non-zero boundary conditions.

Figure 2: Graphical representation of the large sparse matrix A for the discrete solution of the Poisson Equation



4.2 Example 1:Homogeneous equation with non-zero boundary

We consider the Homogeneous Poisson Equation (also known as the Laplacian):

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x^2} = 0, \quad (x, y) \in \Omega = (0, 1) \times (0, 1),$$

with boundary conditions:
lower boundary

$$u(x, 0) = \sin(2\pi x),$$

upper boundary

$$u(x, 1) = \sin(2\pi x),$$

left boundary

$$u(0, y) = 2\sin(2\pi y),$$

right boundary

$$u(1, y) = 2\sin(2\pi y).$$

The general difference equation for the Laplacian is of the form

$$-(w_{i-1j} + w_{ij-1} - 4w_{ij} + w_{ij+1} + w_{i+1j}) = 0.$$

Here, $N = 4$, which gives the step-size,

$$h = \frac{1}{4},$$

and

$$x_i = i\frac{1}{4}, \quad y_j = j\frac{1}{4},$$

for $i = 0, 1, 2, 3, 4$ and $j = 0, 1, 2, 3, 4$. This gives the system of 3×3 equations:

$$\begin{array}{l|lll} j = 1 & & & \\ i = 1 & w_{0,1} + w_{1,0} - 4w_{1,1} + w_{1,2} + w_{2,1} & = & \frac{1}{4}^2 0 \\ i = 2 & w_{1,1} + w_{2,0} - 4w_{2,1} + w_{2,2} + w_{3,1} & = & \frac{1}{4}^2 0 \\ i = 3 & w_{2,1} + w_{3,0} - 4w_{3,1} + w_{3,2} + w_{4,1} & = & \frac{1}{4}^2 0 \end{array}$$

$$
\begin{array}{l|rcl}
j=2 & & & \\
i=1 & w_{0,2}+w_{1,1}-4w_{1,2}+w_{1,3}+w_{2,2} & = & \frac{1}{4}^2 0 \\
i=2 & w_{1,2}+w_{2,1}-4w_{2,2}+w_{2,3}+w_{3,2} & = & \frac{1}{4}^2 0 \\
i=3 & w_{2,2}+w_{3,1}-4w_{3,2}+w_{3,3}+w_{4,2} & = & \frac{1}{4}^2 0 \\
\hline
j=3 & & & \\
i=1 & w_{0,3}+w_{1,2}-4w_{1,3}+w_{1,4}+w_{2,3} & = & \frac{1}{4}^2 0 \\
i=2 & w_{1,3}+w_{2,2}-4w_{2,3}+w_{2,4}+w_{3,3} & = & \frac{1}{4}^2 0 \\
i=3 & w_{2,3}+w_{3,2}-4w_{3,3}+w_{3,4}+w_{4,3} & = & \frac{1}{4}^2 0.
\end{array}
$$

This system is then rearranged by bringing the known boundary conditions to the right hand side, to give:

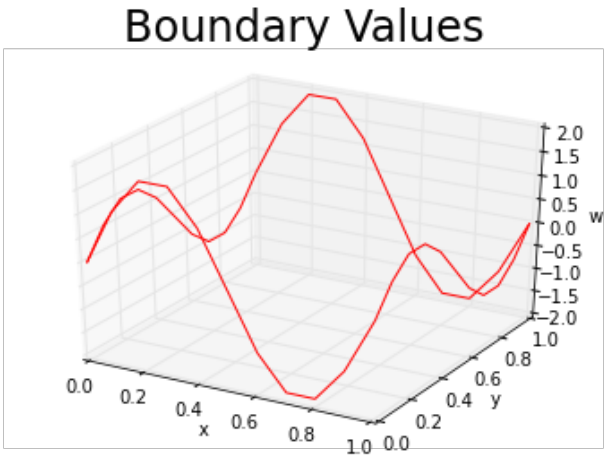
$$
\begin{array}{l|rcl}
j=1 & & & \\
i=1 & -4w_{1,1}+w_{1,2}+w_{2,1} & = & \frac{1}{4}^2 0 - w_{0,1} - w_{1,0} \\
i=2 & w_{1,1}-4w_{2,1}+w_{2,2}+w_{3,1} & = & \frac{1}{4}^2 0 - w_{2,0} \\
i=3 & w_{2,1}-4w_{3,1}+w_{3,2} & = & \frac{1}{4}^2 0 - w_{4,1} - w_{3,0} \\
\hline
j=2 & & & \\
i=1 & w_{1,1}-4w_{1,2}+w_{1,3}+w_{2,2} & = & \frac{1}{4}^2 0 - w_{0,2} \\
i=2 & w_{1,2}+w_{2,1}-4w_{2,2}+w_{2,3}+w_{3,2} & = & \frac{1}{4}^2 0 \\
i=3 & w_{2,2}+w_{3,1}-4w_{3,2}+w_{3,3} & = & \frac{1}{4}^2 0 - w_{4,2} \\
\hline
j=3 & & & \\
i=1 & w_{1,2}-4w_{1,3}+w_{2,3} & = & \frac{1}{4}^2 0 - w_{0,3} - w_{1,4} \\
i=2 & w_{1,3}+w_{2,2}-4w_{2,3}+w_{3,3} & = & \frac{1}{4}^2 0 - w_{2,4} \\
i=3 & w_{2,3}+w_{3,2}-4w_{3,3} & = & \frac{1}{4}^2 0 - w_{4,3} - w_{3,4}
\end{array}
$$

Given the discrete boundary conditions:

Left boundary	**Right boundary**
$x_0 = 0$	$x_4 = 1$
$u(0,y) = 2\sin(2\pi y)$	$u(1,y) = 2\sin(2\pi y)$
$w_{0,0} = 0$	$w_{4,0} = 0$
$w_{0,1} = 2\sin(2\pi y_1) = 2$	$w_{4,1} = 2\sin(2\pi y_1) = 2$
$w_{0,2} = 2\sin(2\pi y_2) = 0$	$w_{4,2} = 2\sin(2\pi y_2) = 0$
$w_{0,3} = 2\sin(2\pi y_3) = -2$	$w_{4,3} = 2\sin(2\pi y_3) = -2$
$w_{0,4} = 2\sin(2\pi y_4) = 0$	$w_{4,4} = 2\sin(2\pi y_4) = 0$

Lower boundary	**Upper boundary**
$y_0 = 0$	$y_4 = 1$
$u(x,0) = \sin(2\pi x)$	$u(x,1) = \sin(2\pi x)$
$w_{0,0} = 0$	$w_{0,4} = 0$
$w_{1,0} = \sin(2\pi x_1) = 1$	$w_{1,4} = \sin(2\pi x_1) = 1$
$w_{2,0} = \sin(2\pi x_2) = 0$	$w_{2,4} = \sin(2\pi x_2) = 0$
$w_{3,0} = \sin(2\pi x_3) = -1$	$w_{3,4} = \sin(2\pi x_3) = -1$
$w_{4,0} = \sin(2\pi x_4) = 0$	$w_{4,4} = \sin(2\pi x_4) = 0$

Figure 3: Sine Wave Boundary Conditions.



The system of equations are written in matrix form:

$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} = \begin{pmatrix} -w_{1,0} \\ -w_{2,0} \\ -w_{3,0} \\ 0 \\ 0 \\ 0 \\ -w_{1,4} \\ -w_{2,4} \\ -w_{3,4} \end{pmatrix} + \begin{pmatrix} -w_{0,1} \\ 0 \\ -w_{4,1} \\ -w_{0,2} \\ 0 \\ -w_{4,2} \\ -w_{0,3} \\ 0 \\ -w_{4,3} \end{pmatrix},$$

where the matrix is $3^2 \times 3^2$ which is graphically represented in Figure 4, where the colours indicated the values in each cell.

For the given boundary conditions the matrix equation is written as :

$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} -2 \\ 0 \\ -2 \\ 0 \\ 0 \\ 0 \\ 2 \\ 0 \\ 2 \end{pmatrix}.$$

Figure 5 shows the approximate solution of the Laplacian Equation for the given boundary conditions and $h = \frac{1}{4}$.

4.2.1 Example 2: non-homogeneous equation with zero boundary

We consider the Poission Equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x^2} = x^2 + y^2 \quad (x, y) \in \Omega = (0, 1) \times (0, 1)$$

with zero boundary conditions: Left boundary:

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

Figure 4: Graphical representation of the matrix

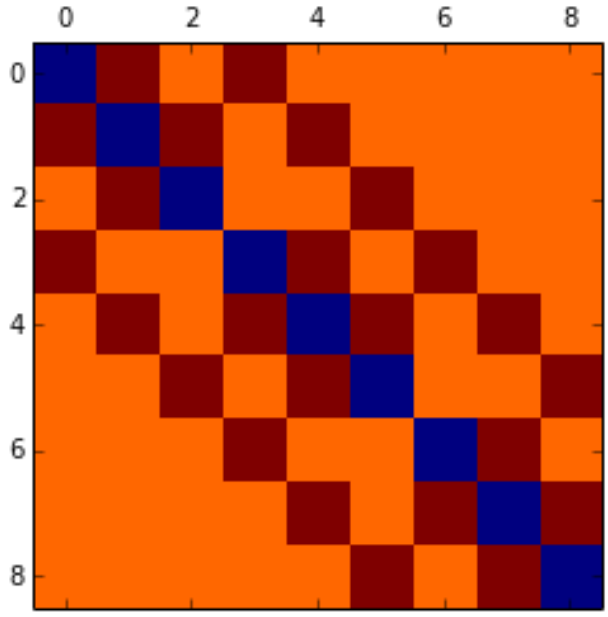
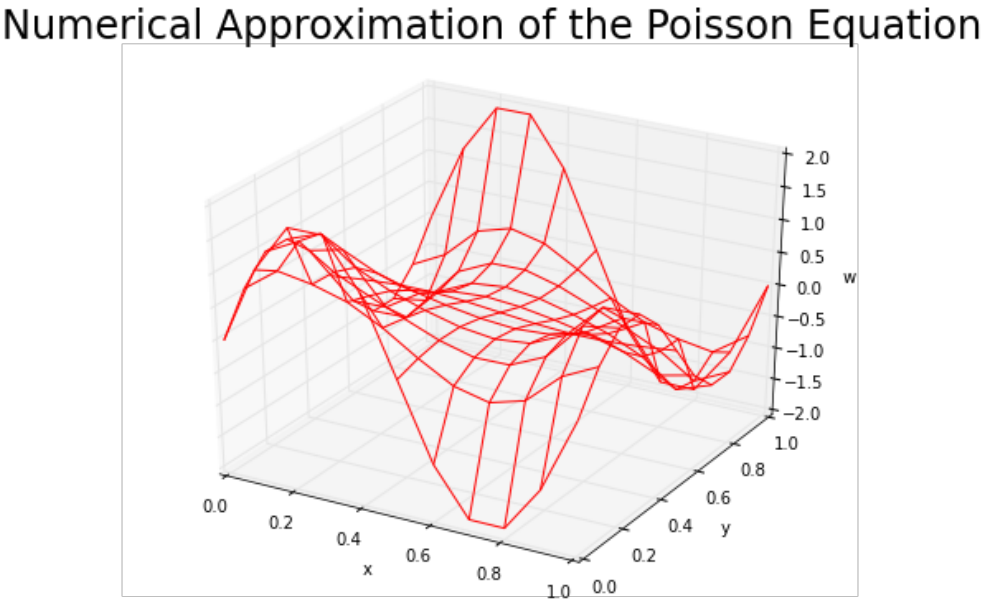


Figure 5: Numerical solution of the homogeneous differential equation



Right boundary:

$$u(x,1) = 0$$

Lower boundary:

$$u(0,y) = 0$$

Upper boundary:

$$u(1,y) = 0.$$

The difference equation is of the form:

$$-(w_{i-1j} + w_{ij-1} - 4w_{ij} + w_{ij+1} + w_{i+1j}) = h^2(x_i^2 + y_j^2).$$

Here, $N = 4$, which gives the step-size,

$$h = \frac{1}{4},$$

and

$$x_i = i\frac{1}{4}, \quad y_j = j\frac{1}{4},$$

for $i = 0,1,2,3,4$ and $j = 0,1,2,3,4$. This gives the system of 3×3 equations:

$$\begin{array}{l|rcl}
j=1 & & & \\
i=1 & w_{0,1} + w_{1,0} - 4w_{1,1} + w_{1,2} + w_{2,1} & = & \frac{1}{4}^2(x_1^2 + y_1^2) \\
i=2 & w_{1,1} + w_{2,0} - 4w_{2,1} + w_{2,2} + w_{3,1} & = & \frac{1}{4}^2(x_2^2 + y_1^2) \\
i=3 & w_{2,1} + w_{3,0} - 4w_{3,1} + w_{3,2} + w_{4,1} & = & \frac{1}{4}^2(x_3^2 + y_1^2)
\end{array}$$

$$\begin{array}{l|rcl}
j=2 & & & \\
i=1 & w_{0,2} + w_{1,1} - 4w_{1,2} + w_{1,3} + w_{2,2} & = & \frac{1}{4}^2(x_1^2 + y_2^2) \\
i=2 & w_{1,2} + w_{2,1} - 4w_{2,2} + w_{2,3} + w_{3,2} & = & \frac{1}{4}^2(x_2^2 + y_2^2) \\
i=3 & w_{2,2} + w_{3,1} - 4w_{3,2} + w_{3,3} + w_{4,2} & = & \frac{1}{4}^2(x_3^2 + y_2^2)
\end{array}$$

$$\begin{array}{l|rcl}
j=3 & & & \\
i=1 & w_{0,3} + w_{1,2} - 4w_{1,3} + w_{1,4} + w_{2,3} & = & \frac{1}{4}^2(x_1^2 + y_3^2) \\
i=2 & w_{1,3} + w_{2,2} - 4w_{2,3} + w_{2,4} + w_{3,3} & = & \frac{1}{4}^2(x_2^2 + y_3^2) \\
i=3 & w_{2,3} + w_{3,2} - 4w_{3,3} + w_{3,4} + w_{4,3} & = & \frac{1}{4}^2(x_3^2 + y_3^2)
\end{array}$$

This system is then rearranged by bringing the known boundary conditions to the right hand side, to give:

$$\begin{array}{l|rcl}
j=1 & & & \\
i=1 & -4w_{1,1} + w_{1,2} + w_{2,1} & = & \frac{1}{4}^2(x_1^2 + y_1^2) - w_{0,1} - w_{1,0} \\
i=2 & w_{1,1} - 4w_{2,1} + w_{2,2} + w_{3,1} & = & \frac{1}{4}^2(x_2^2 + y_1^2) - w_{2,0} \\
i=3 & w_{2,1} - 4w_{3,1} + w_{3,2} & = & \frac{1}{4}^2(x_3^2 + y_1^2) - w_{4,1} - w_{3,0}
\end{array}$$

$$\begin{array}{l|rcl}
j=2 & & & \\
i=1 & w_{1,1} - 4w_{1,2} + w_{1,3} + w_{2,2} & = & \frac{1}{4}^2(x_1^2 + y_2^2) - w_{0,2} \\
i=2 & w_{1,2} + w_{2,1} - 4w_{2,2} + w_{2,3} + w_{3,2} & = & \frac{1}{4}^2(x_2^2 + y_2^2) \\
i=3 & w_{2,2} + w_{3,1} - 4w_{3,2} + w_{3,3} & = & \frac{1}{4}^2(x_3^2 + y_2^2) - w_{4,2}
\end{array}$$

$$\begin{array}{l|rcl}
j=3 & & & \\
i=1 & w_{1,2} - 4w_{1,3} + w_{2,3} & = & \frac{1}{4}^2(x_1^2 + y_3^2) - w_{0,3} - w_{1,4} \\
i=2 & w_{1,3} + w_{2,2} - 4w_{2,3} + w_{3,3} & = & \frac{1}{4}^2(x_2^2 + y_3^2) - w_{2,4} \\
i=3 & w_{2,3} + w_{3,2} - 4w_{3,3} & = & \frac{1}{4}^2(x_3^2 + y_3^2) - w_{4,3} - w_{3,4}.
\end{array}$$

Given the zero boundary conditions

Lower Boundary	**Upper Boundary**
$x_0 = 0$	$x_4 = 1$
$u(0, y) = 0$	$u(1, y) = 0$
$w_{0,0} = 0$	$w_{4,0} = 0$
$w_{0,1} = 0$	$w_{4,1} = 0$
$w_{0,2} = 0$	$w_{4,2} = 0$
$w_{0,3} = 0$	$w_{4,3} = 0$
$w_{0,4} = 0$	$w_{4,4} = 0$

Left Boundary	**Right Boundary**
$y_0 = 0$	$y_4 = 1$
$u(x, 0) = 0$	$u(x, 1) = 0$
$w_{0,0} = 0$	$w_{0,4} = 0$
$w_{1,0} = 0$	$w_{1,4} = 0$
$w_{2,0} = 0$	$w_{2,4} = 0$
$w_{3,0} = 0$	$w_{3,4} = 0$
$w_{4,0} = 0$	$w_{4,4} = 0$

The system of equations can be written in matrix form:

$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} = h^2 \begin{pmatrix} (x_1^2 + y_1^2) \\ (x_2^2 + y_1^2) \\ (x_3^2 + y_1^2) \\ (x_1^2 + y_2^2) \\ (x_2^2 + y_2^2) \\ (x_3^2 + y_2^2) \\ (x_1^2 + y_3^2) \\ (x_2^2 + y_3^2) \\ (x_3^2 + y_3^2) \end{pmatrix}.$$

Substituting values into the right hand side gives the specific matric form:

$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} = \begin{pmatrix} 0.0078125 \\ 0.01953125 \\ 0.0390625 \\ 0.01953125 \\ 0.0312 \\ 0.05078125 \\ 0.0390625 \\ 0.05078125 \\ 0.0703125 \end{pmatrix}.$$

Figure 7 shows the numerical solution of the Poisson Equation with zero boundary conditions.

4.2.2 Example 3: Inhomogeneous equation with non-zero boundary

We consider the Poisson Equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial x^2} = xy, \quad (x, y) \in \Omega = (0, 1) \times (0, 1),$$

with boundary conditions
Right Boundary

$$u(x, 0) = -x^2 + x$$

Figure 6: Sine Wave Boundary Conditions.

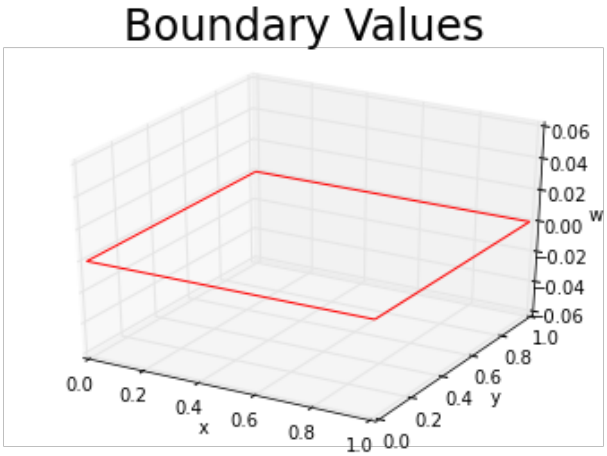
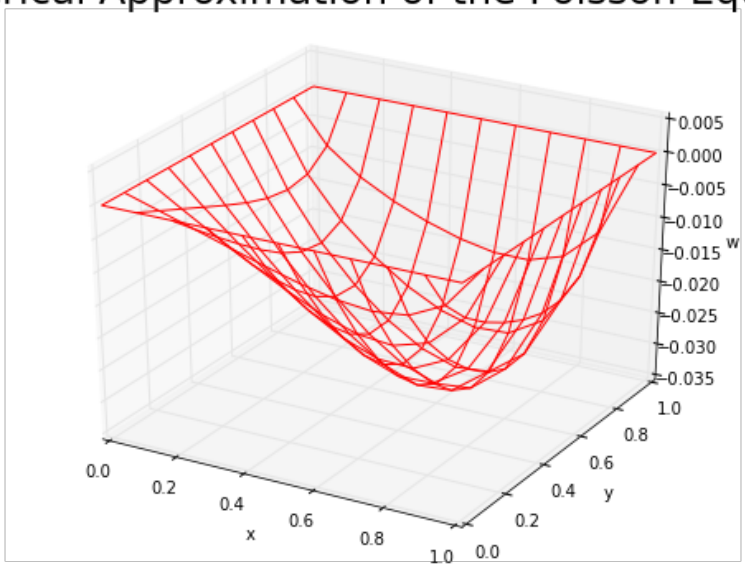


Figure 7: Numerical solution of the differential equation with zero boundary conditions



Left Boundary

$$u(x,1) = x^2 - x$$

Lower Boundary

$$u(0,y) = -y^2 + y$$

Upper Boundary

$$u(1,y) = -y^2 + y.$$

The five point difference equation is of the form

$$-(w_{i-1j} + w_{ij-1} - 4w_{ij} + w_{ij+1} + w_{i+1j}) = h^2(x_i y_j).$$

Here, $N = 4$, which gives the step-size,

$$h = \frac{1}{4},$$

and

$$x_i = i\frac{1}{4}, \quad y_j = j\frac{1}{4},$$

for $i = 0,1,2,3,4$ and $j = 0,1,2,3,4$. This gives the system of 3×3 equations:

$$\begin{array}{l|lcl}
i=1 & w_{0,1} + w_{1,0} - 4w_{1,1} + w_{1,2} + w_{2,1} & = & \frac{1}{4}^2(x_1y_1) \\
i=2 & w_{1,1} + w_{2,0} - 4w_{2,1} + w_{2,2} + w_{3,1} & = & \frac{1}{4}^2(x_2y_1) \\
i=3 & w_{2,1} + w_{3,0} - 4w_{3,1} + w_{3,2} + w_{4,1} & = & \frac{1}{4}^2(x_3y_1)
\end{array}$$

$$\begin{array}{l|lcl}
j=2 & & & \\
i=1 & w_{0,2} + w_{1,1} - 4w_{1,2} + w_{1,3} + w_{2,2} & = & \frac{1}{4}^2(x_1y_2) \\
i=2 & w_{1,2} + w_{2,1} - 4w_{2,2} + w_{2,3} + w_{3,2} & = & \frac{1}{4}^2(x_2y_2) \\
i=3 & w_{2,2} + w_{3,1} - 4w_{3,2} + w_{3,3} + w_{4,2} & = & \frac{1}{4}^2(x_3y_2)
\end{array}$$

$$\begin{array}{l|lcl}
j=3 & & & \\
i=1 & w_{0,3} + w_{1,2} - 4w_{1,3} + w_{1,4} + w_{2,3} & = & \frac{1}{4}^2(x_1y_3) \\
i=2 & w_{1,3} + w_{2,2} - 4w_{2,3} + w_{2,4} + w_{3,3} & = & \frac{1}{4}^2(x_2y_3) \\
i=3 & w_{2,3} + w_{3,2} - 4w_{3,3} + w_{3,4} + w_{4,3} & = & \frac{1}{4}^2(x_3y_3).
\end{array}$$

Rearranging the system such that the known values are on the right hand side:

$$\begin{array}{l|rcl}
j=1 & & & \\
i=1 & -4w_{1,1} + w_{1,2} + w_{2,1} & = & \frac{1}{4}^2(x_1y_1) - w_{0,1} - w_{1,0} \\
i=2 & w_{1,1} - 4w_{2,1} + w_{2,2} + w_{3,1} & = & \frac{1}{4}^2(x_2y_1) - w_{2,0} \\
i=3 & w_{2,1} - 4w_{3,1} + w_{3,2} & = & \frac{1}{4}^2(x_3y_1) - w_{4,1} - w_{3,0}
\end{array}$$

$$\begin{array}{l|rcl}
j=2 & & & \\
i=1 & w_{1,1} - 4w_{1,2} + w_{1,3} + w_{2,2} & = & \frac{1}{4}^2(x_1y_2) - w_{0,2} \\
i=2 & w_{1,2} + w_{2,1} - 4w_{2,2} + w_{2,3} + w_{3,2} & = & \frac{1}{4}^2(x_2y_2) \\
i=3 & w_{2,2} + w_{3,1} - 4w_{3,2} + w_{3,3} & = & \frac{1}{4}^2(x_3y_2) - w_{4,2}
\end{array}$$

$$\begin{array}{l|rcl}
j=3 & & & \\
i=1 & w_{1,2} - 4w_{1,3} + w_{2,3} & = & \frac{1}{4}^2(x_1y_3) - w_{0,3} - w_{1,4} \\
i=2 & w_{1,3} + w_{2,2} - 4w_{2,3} + w_{3,3} & = & \frac{1}{4}^2(x_2y_3) - w_{2,4} \\
i=3 & w_{2,3} + w_{3,2} - 4w_{3,3} & = & \frac{1}{4}^2(x_3y_3) - w_{4,3} - w_{3,4}.
\end{array}$$

The discrete boundary conditions are

Left boundary	**Right boundary**
$x_0 = 0$	$x_4 = 1$
$u(0,y) = -y^2 + y$	$u(1,y) = -y^2 + y$
$w_{0,0} = 0$	$w_{4,0} = 0$
$w_{0,1} = -y_1^2 + y_1 = \frac{3}{16}$	$w_{4,1} = -y_1^2 + y_1 = \frac{1}{16}$
$w_{0,2} = -y_2^2 + y_2 = \frac{1}{4}$	$w_{4,2} = -y_2^2 + y_2 = \frac{1}{4}$
$w_{0,3} = -y_3^2 + y_3 = \frac{3}{16}$	$w_{4,3} = -y_3^2 + y_3 = \frac{3}{16}$
$w_{0,4} = -y_4^2 + y_4 = 0$	$w_{4,4} = -y_4^2 + y_4 = 0$

Lower boundary	**Upper boundary**
$y_0 = 0$	$y_4 = 1$
$u(x,0) = -x^2 + x$	$u(x,1) = x^2 - x$
$w_{0,0} = 0$	$w_{0,4} = 0$
$w_{1,0} = -x_1^2 + x_1 = \frac{3}{16}$	$w_{1,4} = x_1^2 - x_1 = -\frac{3}{16}$
$w_{2,0} = -x_2^2 + x_2 = \frac{1}{4}$	$w_{2,4} = x_2^2 - x_2 = -\frac{1}{4}$
$w_{3,0} = -x_3^2 + x_3 = \frac{3}{16}$	$w_{3,4} = x_3^2 - x_3 = -\frac{3}{16}$
$w_{4,0} = 0$	$w_{4,4} = 0$

The system of equations can be written in 9×9 Matrix form:

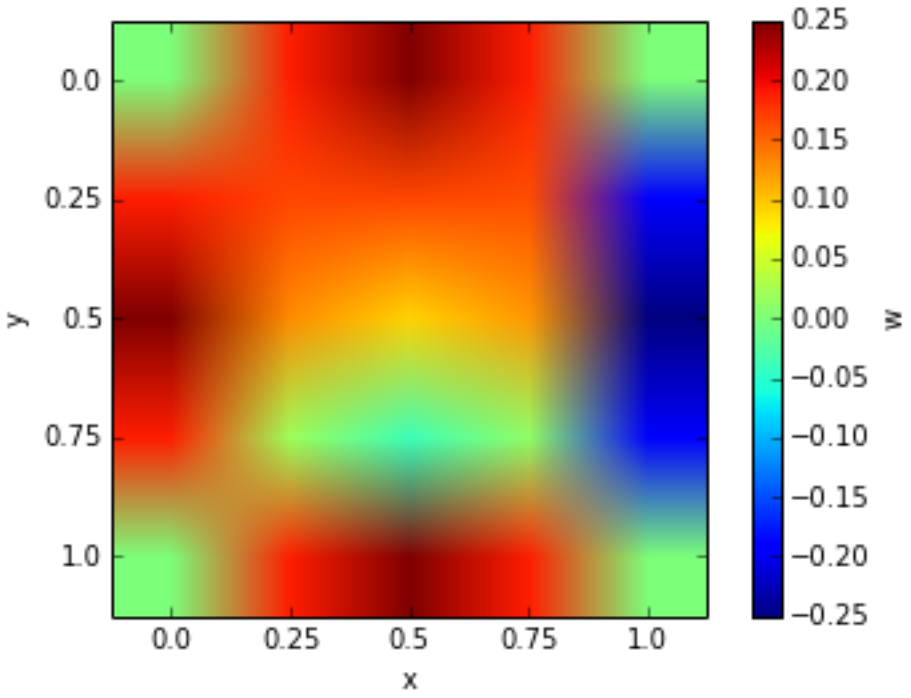
$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} =$$

$$h^2 \begin{pmatrix} (x_1 y_1) \\ (x_2 y_1) \\ (x_3 y_1) \\ (x_1 y_2) \\ (x_2 y_2) \\ (x_3 y_2) \\ (x_1 y_3) \\ (x_2 y_3) \\ (x_3 y_3) \end{pmatrix} + \begin{pmatrix} -w_{1,0} \\ -w_{2,0} \\ -w_{3,0} \\ 0 \\ 0 \\ 0 \\ -w_{1,4} \\ -w_{2,4} \\ -w_{3,4} \end{pmatrix} + \begin{pmatrix} -w_{0,1} \\ 0 \\ -w_{4,1} \\ -w_{0,2} \\ 0 \\ -w_{4,2} \\ -w_{0,3} \\ 0 \\ -w_{4,3} \end{pmatrix},$$

inputting the specific boundary values and the right hand side of the equation gives:

$$\begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix} \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ w_{3,1} \\ w_{1,2} \\ w_{2,2} \\ w_{3,2} \\ w_{1,3} \\ w_{2,3} \\ w_{3,3} \end{pmatrix} =$$

Figure 8: Numerical solution of the differential equation with non-zero boundary conditions



$$\left(\frac{1}{4}\right)^2 \begin{pmatrix} 0.0625 \\ 0.125 \\ 0.1875 \\ 0.125 \\ 0.25 \\ 0.375 \\ 0.1875 \\ 0.375 \\ 0.5626 \end{pmatrix} + \begin{pmatrix} -\frac{3}{16} \\ -\frac{1}{4} \\ -\frac{3}{16} \\ 0 \\ 0 \\ 0 \\ \frac{3}{16} \\ \frac{1}{4} \\ \frac{3}{16} \end{pmatrix} + \begin{pmatrix} -\frac{3}{16} \\ 0 \\ -\frac{3}{16} \\ -\frac{1}{4} \\ 0 \\ -\frac{1}{4} \\ -\frac{3}{16} \\ 0 \\ -\frac{3}{16} \end{pmatrix}.$$

Figure 8 shows the numerical solution of the Poisson Equation with non-zero boundary conditions.

4.3 Consistency and Convergence

We will discuss how well the grid function determined by the five point scheme approximates the exact solution of the Poisson problem.

Definition. *Let L_h denote the finite difference approximation associated with the grid Ω_h having the mesh size h, to a partial differential operator L defined on a simply connected, open set $\Omega \subset R^2$. For a given function $\varphi \in C^\infty(\Omega)$, the truncation error of L_h is*

$$\tau_h(\mathbf{x}) = (L - L_h)\varphi(x)$$

The approximation L_h is consistent with L if

$$\lim_{h \to 0} \tau_h(x) = 0,$$

for all $\mathbf{x} \in D$ and all $\varphi \in C^\infty(\Omega)$. The approximation is consistent to order p if $\tau_h(\mathbf{x}) = O(h^p)$. ∘

While we have seen this definition a few times, it is always interesting to discuss how the terms are denoted and expressed. Nevertheless, the ideas keep the same.

Proposition 1. *The five-point difference analog $-\nabla_h^2$ is consistent to order 2 with $-\nabla^2$.*

Proof. Pick $\varphi \in C^\infty(D)$, and let $(x,y) \in \Omega$ be a point such that $(x \pm h, y), (x, y \pm h) \in \Omega \bigcup \partial\Omega$. By the Taylor Theorem

$$\varphi(x \pm h, y) \quad = \quad \varphi(x,y) \pm h\frac{\partial \varphi}{\partial x}(x,y) + \frac{h^2}{2!}\frac{\partial^2 \varphi}{\partial x^2}(x,y) \pm \frac{h^3}{3!}\frac{\partial^3 \varphi}{\partial x^3}(x,y) + \frac{h^4}{4!}\frac{\partial^4 \varphi}{\partial x^4}(\zeta^{\pm}, y)$$

where $\zeta^{\pm} \in (x-h, x+h)$. Adding this pair of equation together and rearranging , we get

$$\frac{1}{h^2}[\varphi(x+h,y) - 2\varphi(x,y) + \varphi(x-h,y)] - \frac{\partial^2 \varphi}{\partial x^2}(x,y) = \frac{h^2}{4!}\left[\frac{\partial^4 \varphi}{\partial x^4}(\zeta^+, y) + \frac{\partial^4 \varphi}{\partial x^4}(\zeta^-, y)\right]$$

By the intermediate value theorem

$$\left[\frac{\partial^4 \varphi}{\partial x^4}(\zeta^+, y) + \frac{\partial^4 \varphi}{\partial x^4}(\zeta^-, y)\right] = 2\frac{\partial^4 \varphi}{\partial x^4}(\zeta, y),$$

for some $\zeta \in (x-h, x+h)$. Therefore,

$$\delta_x^2(x,y) = \frac{\partial^2 \varphi}{\partial x^2}(x,y) + \frac{h^2}{2!}\frac{\partial^4 \varphi}{\partial x^4}(\zeta, y)$$

Similar reasoning shows that

$$\delta_y^2(x,y) = \frac{\partial^2 \varphi}{\partial y^2}(x,y) + \frac{h^2}{2!}\frac{\partial^4 \varphi}{\partial y^4}(x, \eta)$$

for some $\eta \in (y-h, y+h)$. We conclude that $\tau_h(x,y) = (\nabla - \nabla_h)\varphi(x,y) = O(h^2)$. • □

Note. *Consistency does not guarantee that the solution to the difference equations approximates the exact solution to the PDE.*

Definition. *Let $L_h w(\mathbf{x}_j) = f(\mathbf{x}_j)$ be a finite difference approximation, defined on a grid mesh size h, to a **PDE** $LU(\mathbf{x}) = f(\mathbf{x})$ on a simply connected set $D \subset R^n$. Assume that $w(x,y) = U(x,y)$ at all points (x,y) on the boundary $\partial\Omega$. The finite difference scheme converges (or is convergent) if*

$$\max_j |U(\mathbf{x}_j) - w(\mathbf{x}_j)| \to 0 \text{ as } h \to 0.$$

For the five point scheme there is a direct connection between consistency and convergence. Underlying this connection is an argument based on the following principle:

Theorem. *(DISCRETE MAXIMUM PRINCIPLE). If $\nabla_h^2 V_{ij} \geq 0$ for all points $(x_i, y_j) \in \Omega_h$, then*

$$\max_{(x_i,y_j)\in\Omega_h} V_{ij} \leq \max_{(x_i,y_j)\in\partial\Omega_h} V_{ij}.$$

If $\nabla_h^2 V_{ij} \leq 0$ for all points $(x_i, y_j) \in \Omega_h$, then

$$\min_{(x_i,y_j)\in\Omega_h} V_{ij} \geq \min_{(x_i,y_j)\in\partial\Omega_h} V_{ij}.$$

In other words, a grid function V for which $\nabla_h^2 V$ is nonnegative on Ω_h attains its maximum on the boundary $\partial\Omega_h$ of the grid. Similarly, if $\nabla_h^2 V$ is nonpositive on Ω_h, then V attains its minimum on the boundary $\partial\Omega_h$.

Proof. The proof is by contradiction. We argue for the case $\nabla_h^2 V_{ij} \geq 0$, reasoning for the case $\nabla_h^2 V_{ij} \leq 0$ begin similar.
Assume that V attains its maximum value M at an interior grid point (x_I, y_J) and that $\max_{(x_i,y_j)\in\partial\Omega_h} V_{ij} < M$. The hypothesis $\nabla_h^2 V_{ij} \geq 0$ implies that

$$V_{IJ} \leq \frac{1}{4}(V_{I+1J} + V_{I-1J} + V_{IJ+1} + V_{IJ-1})$$

This cannot hold unless

$$V_{I+1J} = V_{I-1J} = V_{IJ+1} = V_{IJ-1} = M.$$

If any of the corresponding grid points $(x_{I+1}, y_L), (x_{J-1}, y_L), (x_I, y_{L+1}), (x_I, y_{L-1})$ lies in $\partial\Omega_h$, then we have reached the desired contradiction.
Otherwise, we continue arguing in this way until we conclude that $V_{I+iJ+j} = M$ for some point $(x_{I+iJ+j}) \in \partial\Omega$, which again gives a contradiction. • □

This leads to some interesting results:

Proposition 2. *1. The zero grid function (for which $U_{ij} = 0$ for all $(x_i, y_j) \in \Omega_h \bigcup \partial\Omega_h$ is the only solution to the finite difference problem*

$$\nabla_h^2 U_{ij} = 0 \text{ for } (x_i, y_j) \in \Omega_h,$$

$$U_{ij} = 0 \text{ for } (x_i, y_j) \in \partial\Omega_h.$$

2. For prescribed grid functions f_{ij} and g_{ij}, there exists a unique solution to the problem

$$\nabla_h^2 U_{ij} = f_{ij} \text{ for } (x_i, y_j) \in \Omega_h,$$

$$U_{ij} = g_{ij} \text{ for } (x_i, y_j) \in \partial\Omega_h.$$

Definition. *For any grid function $V : \Omega_h \bigcup \partial\Omega_h \to R$,*

$$||V||_\Omega = \max_{(x_i, y_j) \in \Omega_h} |V_{ij}|,$$

$$||V||_{\partial\Omega} = \max_{(x_i, y_j) \in \partial\Omega_h} |V_{ij}|.$$

∘

Lemma 1. *If the grid function $V : \Omega_h \bigcup \partial\Omega_h \to R$ satisfies the boundary condition $V_{ij} = 0$ for $(x_i, y_j) \in \partial\Omega_h$, then*

$$||V_|||_\Omega \leq \frac{1}{8} ||\nabla_h^2 V||_\Omega.$$

Proof. Let $\nu = ||\nabla_h^2 V||_\Omega$. Clearly for all points $(x_i, y_j) \in \Omega_h$,

$$-\nu \leq \nabla_h^2 V_{ij} \leq \nu \tag{11}$$

Now we define $W : \Omega_h \bigcup \partial\Omega_h \to R$ by setting $W_{ij} = \frac{1}{4}[(x_i - \frac{1}{2})^2 + (y_j - \frac{1}{2})^2]$, which is nonnegative. Also $\nabla_h^2 W_{ij} = 1$ and that $||W||_{\partial\Omega} = \frac{1}{8}$. The inequality (11) implies that, for all points $(x_i, y_j) \in \Omega_h$,

$$\nabla_h^2 (V_{ij} + \nu W_{ij}) \geq 0$$

$$\nabla_h^2 (V_{ij} - \nu W_{ij}) \leq 0$$

By the discrete minimum principle and the fact that V vanishes on $\partial\Omega_h$

$$V_{ij} \leq V_{ij} + \nu W_{ij} \leq \nu ||W||_{\partial\Omega}$$

$$V_{ij} \geq V_{ij} - \nu W_{ij} \geq -\nu ||W||_{\partial\Omega}$$

Since $||W||_{\partial\Omega} = \frac{1}{8}$

$$||V_|||_\Omega \leq \frac{1}{8}\nu = \frac{1}{8} ||\nabla_h^2 V||_\Omega$$

• □

Finally we prove that the five point scheme for the **Poisson equation** is convergent.

Theorem. *Let U be a solution to the* ***Poisson equation*** *and let w be the grid function that satisfies the discrete analog*

$$-\nabla_h^2 w_{ij} = f_{ij} \quad \textit{for } (x_i, y_j) \in \Omega_h,$$
$$w_{ij} = g_{ij} \quad \textit{for } (x_i, y_j) \in \partial\Omega_h.$$

Then there exists a positive constant K such that

$$||U - w||_\Omega \leq KMh^2$$

where

$$M = \left\{ \left|\left| \frac{\partial^4 U}{\partial x^4} \right|\right|_\infty, \left|\left| \frac{\partial^4 U}{\partial x^3 \partial y} \right|\right|_\infty, ..., \left|\left| \frac{\partial^4 U}{\partial y^4} \right|\right|_\infty \right\}$$

The statement of the theorem assumes that $U \in C^4(\bar{\Omega})$. This assumption holds if f and g are smooth enough.

Proof. Following from the proof of the Proposition we have

$$(\nabla_h^2 - \nabla^2)U_{ij} = \frac{h^2}{12}\left[\frac{\partial^4 U}{\partial x^4}(\zeta_i, y_j) + \frac{\partial^4 U}{\partial y^4}(x_i, \eta_j)\right]$$

for some $\zeta \in (x_{i-1}, x_{i+1})$ and $\eta_j \in (y_{j-1}, y_{j+1})$. Therefore,

$$-\nabla_h^2 U_{ij} = f_{ij} - \frac{h^2}{12}\left[\frac{\partial^4 U}{\partial x^4}(\zeta_i, y_j) + \frac{\partial^4 U}{\partial y^4}(x_i, \eta_j)\right].$$

If we subtract from this the identity equation $-\nabla_h^2 w_{ij} = f_{ij}$ and note that $U - w$ vanishes on $\partial\Omega_h$, we find that

$$\nabla_h^2 (U_{ij} - w_{ij}) = \frac{h^2}{12}\left[\frac{\partial^4 U}{\partial x^4}(\zeta_i, y_j) + \frac{\partial^4 U}{\partial y^4}(x_i, \eta_j)\right].$$

It follows that

$$||U - w||_\Omega \leq \frac{1}{8}||\nabla_h^2 (U - w)||_\Omega \leq KMh^2.$$

□

5 Hyperbolic equations

Content

We have a first-order scalar equation

$$\begin{array}{ll} \frac{\partial U}{\partial t} = -a\frac{\partial U}{\partial x} & x \in R \;\; t > 0 \\ U(x,0) = U_0(x) & x \in R \end{array} \tag{12}$$

where a is a positive real number. Its solution is given by

$$U(x,t) = U_0(x - at) \quad t \geq 0$$

and represents a traveling wave with velocity a. The curves $(x(t), t)$ in the plane (x, t) are the characteristic curves. They are the straight lines $x(t) = x_0 + at$, $t > 0$. The solution of (12) remains constant along them. For the more general problem

$$\begin{array}{ll} \frac{\partial U}{\partial t} + a\frac{\partial U}{\partial x} + a_0 = f & x \in R \;\; t > 0 \\ U(x,0) = U_0(x) & x \in R \end{array} \tag{13}$$

where a, a_0 and f are given functions of the variables (x, t), the characteristic curves are still defined as before. In this case the solutions of (13) satisfy along the characteristics the following differential equation

$$\frac{du}{dt} = f - a_0 u \text{ on } (x(t), t)$$

5.1 The Wave Equation

We consider the second-order hyperbolic equation

$$\frac{\partial^2 U}{\partial t^2} - \gamma \frac{\partial^2 U}{\partial x^2} = f \quad x \in (\alpha, \beta), \quad t > 0 \tag{14}$$

with initial data

$$U(x,0) = u_0(x) \text{ and } \frac{\partial U}{\partial t}(x,0) = v_0(x), \quad x \in (\alpha, \beta)$$

and boundary data

$$U(\alpha, t) = 0 \text{ and } U(\beta, t) = 0, \quad t > 0$$

In this case, U may represent the transverse displacement of an elastic vibrating string of length $\beta - \alpha$, fixed at the endpoints and γ is a coefficient depending on the specific mass of the string and its tension. The spring is subject to a vertical force of density f.

The functions $u_0(x)$ and $v_0(x)$ denote respectively the initial displacement and initial velocity of the string.
The change of variables

$$\omega_1 = \frac{\partial U}{\partial x}, \quad \omega_2 = \frac{\partial U}{\partial t}$$

We transform this into

$$\frac{\partial \hat{\omega}}{\partial t} + A \frac{\partial \hat{\omega}}{\partial x} = \mathbf{0}$$

where

$$\hat{\omega} = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$$

Since the initial conditions are $\omega_1(x,0) = u_0'(x)$ and $\omega_2(x,0) = v_0(x)$.

Note. *Notice that replacing $\frac{\partial^2 u}{\partial t^2}$ by t^2, $\frac{\partial^2 u}{\partial x^2}$ by x^2 and f by 1, the wave equation becomes*

$$t^2 - \gamma^2 x^2 = 1$$

which represents an hyperbola in (x,t) plane. Proceeding analogously in the case of the heat equation we end up with

$$t - x^2 = 1$$

which represents a parabola in the (x,t) plane. Finally, for the Poisson equation we get

$$x^2 + y^2 = 1$$

which represents an ellipse in the (x,y) plane.
Due to the geometric interpretation above, the corresponding differential operators are classified as hyperbolic, parabolic and elliptic.

5.2 Finite Difference Method for Hyperbolic equations

As always we discretise the domain by space-time finite difference. To this aim, the half-plane $\{(x,t) : -\infty < x < \infty, t > 0\}$ is discretised by choosing a spatial grid size Δx, a temporal step Δt and the grid points (x_j, t^n) as follows

$$x_j = j\Delta x \quad j \in Z, \quad t^n = n\Delta t \quad n \in N$$

and let

$$\lambda = \frac{\Delta t}{\Delta x}.$$

5.3 Discretisation of the scalar equation

Here are some explicit methods

- Forward Euler/centered method:

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}a(u_{j+1}^n - u_{j-1}^n),$$

- Lax-Friedrichs method,

$$u_j^{n+1} = \frac{u_{j+1}^n + u_{j-1}^n}{2} - \frac{\lambda}{2}a(u_{j+1}^n - u_{j-1}^n),$$

- Lax-Wendroff method,

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}a(u_{j+1}^n - u_{j-1}^n) + \frac{\lambda^2}{2}a^2(u_{j+1}^n - 2u_j^n + u_{j-1}^n),$$

- Upwind method,

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{\lambda}{2}|a|(u_{j+1}^n - 2u_j^n + u_{j-1}^n).$$

The last three methods can be obtained from the forward Euler/centered method by adding a term proportional to a numerical approximation of a second derivative term. Hence, they can be written in the equivalent form

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}a(u_{j+1}^n - u_{j-1}^n) + \frac{1}{2}k\frac{(u_{j+1}^n - 2u_j^n + u_{j-1}^n)}{(\Delta x)^2}$$

where k is an artificial viscosity term.
An example of an implicit method is the backward Euler/ centered scheme

$$u_j^{n+1} + \frac{\lambda}{2}a(u_{j+1}^{n+1} - u_{j-1}^{j+1}) = u_j^n$$

5.4 Consistency

A numerical method is convergent if

$$\lim_{\Delta t, \Delta x \to 0} \max_{j,n} |U(x_j, t^n) - w_j^n|$$

The local truncation error at x_j, t^n is defined as

$$\tau_j^n = L(U_j^n)$$

the truncation error is

$$\tau(\Delta t, \Delta x) = \max_{j,n} |\tau_j^n|$$

When $\tau(\Delta t, \Delta x)$ goes to zero as Δt and Δx tend to zero independently is said to be consistent.

5.5 Stability

5.5.1 Courant Freidrich Lewy Condition

A method is said to be stable if, for any T here exist a constant $C_T > 0$ and δ_0 such that

$$||\mathbf{u}^n||_\Delta \leq C_T ||\mathbf{u}^0||_\Delta$$

for any n such that $n\Delta t \leq T$ and for any $\Delta t, \Delta x$ such that $0 < \Delta t \leq \delta_0, 0 < \Delta x \leq \delta_0$. We have denoted by $||.||_\Delta$ a suitable discrete norm.

Forward Euler/centered

$$u_j^{n+1} = u_j^n - \frac{\lambda}{2}a(u_{j+1}^n - u_{j-1}^n)$$

Truncation error

$$O(\Delta t, (\Delta x)^2)$$

For an explicit method to be stable, we need

$$|a\lambda| = \left| a\frac{\delta t}{\delta x} \right| \leq 1$$

this is known as the Courant Freidrich Lewy condition.
Using Von Neumann stability analysis we can show that the method is stable under the Courant Freidrich Lewy condition.

5.5.2 test Neumann stability for the Forward Euler

$$u_j^n = e^{i\beta j\Delta x}\xi^n$$

where

$$\xi = e^{\alpha\Delta t}$$

It is sufficient to show

$$|\xi| \leq 1$$

$$\xi^{n+1}e^{i\beta(j)\Delta x} = \xi^n e^{i\beta(j)\Delta x} + \frac{\lambda}{2}a(\xi^n e^{i\beta(j+1)\Delta x} - \xi^n e^{i\beta(j-1)\Delta x})$$

$$\xi = 1 - \frac{\lambda}{2}a(e^{i\beta\Delta x} - e^{-i\beta\Delta x})$$

$$\xi = 1 - i\frac{\lambda}{2}a(2sin(\beta\Delta x))$$

$$\xi = 1 - i\lambda a(sin(\beta\Delta x))$$

$$|\xi| = \sqrt{1 + (\lambda a(sin(\beta\Delta x)))^2}$$

Hence

$$\xi > 1$$

Therefore, the method is unstable for the Courant Freidrich Lewy.

5.5.3 test Neumann stability for the Lax-Friedrich

$$\xi^{n+1}e^{i\beta(j)\Delta x} = \frac{\xi^n e^{i\beta(j+1)\Delta x} + \xi^n e^{i\beta(j-1)\Delta x}}{2} + \frac{\lambda}{2}a(\xi^n e^{i\beta(j+1)\Delta x} - \xi^n e^{i\beta(j-1)\Delta x})$$

$$\xi = \frac{e^{i\beta\Delta x} - e^{-i\beta\Delta x}}{2} + \frac{\lambda}{2}a(e^{i\beta\Delta x} - e^{-i\beta\Delta x})$$

$$\xi = \frac{1+\lambda a}{2}e^{i\beta\Delta x} + \frac{1-\lambda a}{2}e^{-i\beta\Delta x}$$

$$\xi = \cos(\beta\Delta x) + i\lambda a\sin(\beta\Delta x)$$

$$|\xi|^2 \leq (\cos(\beta\Delta x))^2 + (a\lambda)^2(\sin(\beta\Delta x))^2$$

Hence

$$\xi < 1$$

for $a\lambda \leq 1$.

Example. *These methods can be applied to the Burgers equation*

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0$$

as there it is a non-trivial non-linear hyperbolic equation. Taking initial condition

$$u(x,0) = u_0(x) = \begin{cases} 1, & x_0 \leq 0, \\ 1-x, & 0 \leq x_0 \leq 1, \\ 0, & x_0 \geq 1. \end{cases}$$

the characteristic line issuing from the point $(x_0, 0)$ *is given by*

$$x(t) = x_0 + tu_0(x_0) = \begin{cases} x_0 + t, & x_0 \leq 0 \\ x_0 + t(1-x_0), & 0 \leq x_0 \leq 1, \\ x_0 & x_0, \geq 1. \end{cases}$$

6 Parabolic equations

Content

We will look at the **Heat equation** as our sample parabolic equation.

$$\frac{\partial U}{\partial T} = K\frac{\partial^2 U}{\partial X^2} \text{ on } \Omega$$

and

$$U = g(x,y) \text{ on the boundary } \delta\Omega$$

this can be transformed without loss of generality by a non-dimensional transformation to

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2} \tag{15}$$

with the domain

$$\Omega = \{(t,x)|\ 0 \leq t, 0 \leq x \leq 1\}.$$

6.1 An explicit method for the Heat Equation

The difference equation of the differential Heat Equation (15) is of the form:

$$\frac{w_{ij+1} - w_{ij}}{t_{j+1} - t_j} = \frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h^2} \tag{16}$$

when approaching this we have divided up the area into two uniform meshes one in the x direction and the other in the t-direction. We define $t_j = jk$ where k is the step size in the time direction.
We define $x_i = ih$ where h is the step size in the space direction.
w_{ij} denotes the numerical approximation of U at (x_i, t_j).
Rearranging the equation we get

$$w_{ij+1} = rw_{i-1j} + (1-2r)w_{ij} + rw_{i+1j} \tag{17}$$

where $r = \frac{k}{h^2}$.
This gives the formula for the unknown term w_{ij+1} at the $(ij+1)$ mesh points in terms of terms along the j th time row.
Hence we can calculate the unknown pivotal values of w along the first row $t = k$ or $j = 1$ in terms of the known boundary conditions.

Example. *In this case, we look at a rod of unit length with each end in ice.*
The rod is heat insulated along its length, so that temp changes occur through heat conduction along its length and heat transfer at its ends, where w denotes temp.
Simple case
Given that the ends of the rod are kept in contact with ice and the initial temp distribution is non dimensional form is

1. $U = 2x$ *for* $0 \leq x \leq \frac{1}{2}$
2. $U = 2(1-x)$ *for* $\frac{1}{2} \leq x \leq 1$

In other words we are seeking a numerical solution of

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$$

which satisfies

1. *U=0 at x=0 for all* $t > 0$ *(the boundary condition)*
2. $U = 2x$ *for* $0 \leq x \leq \frac{1}{2}$ *for t=0* $U = 2(1-x)$ *for* $\frac{1}{2} \leq x \leq 1$ *for t=0 (the initial condition)*

The problem is symmetric with respect to $x = 0.5$ *so we need the solution only for* $0 \leq x \leq \frac{1}{2}$

Case 1 *Let* $h = \frac{1}{10}$ *and* $k = \frac{1}{1000}$ *so that* $r = \frac{k}{h^2} = \frac{1}{10}$ *difference equation (17) becomes*

$$w_{ij+1} = \frac{1}{10}(w_{i-1j} + 8w_{ij} + w_{i+1j})$$

to solve for w_{51} *we have*

$$w_{51} = \frac{1}{10}(w_{40} + 8w_{50} + w_{60}) = \frac{1}{10}(0.8 + 8 * 1 + 0.8) = 0.96$$

j/x	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*
0	*0*	*0.2*	*0.4*	*0.6*	*0.8*	*1.0*	*0.8*
1	*0*	*0.2*	*0.4*	*0.6*	*0.8*	*0.96*	*0.8*

The analytical solution of the PDE satisfying these conditions is

$$U = \frac{8}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} sin(\frac{1}{2} n\pi) sin(\pi x) e^{-n^2 \pi^2 t}$$

Comparison to the solution with the difference solution it is reasonably accurate.
Case 2 *Let* $h = \frac{1}{10}$ *and* $k = \frac{1}{200}$ *so that* $r = \frac{k}{h^2} = \frac{1}{2}$ *difference equation (17) becomes*

$$w_{ij+1} = \frac{1}{2}(w_{i-1j} + w_{i+1j})$$

This method also gives an acceptable approximation to the solution of the PDE. ***Case 3*** *Let* $h = \frac{1}{10}$ *and* $k = \frac{1}{100}$ *so that* $r = \frac{k}{h^2} = 1$ *difference equation (17) becomes*

$$w_{ij+1} = w_{i-1j} - w_{ij} + w_{i+1j}$$

Considered as a solution to the ***PDE*** *this is meaningless although it is the correct solution of the difference equation with respect to the initial conditions and the boundary conditions.*
This will be discussed later.

6.2 Crank Nicholson Implicit method

Since the implicit method requires that $k \leq \frac{1}{2}h^2$ a new method was needed which would work for all finite values of r.
They considered the partial differential equation as being satisfied at the midpoint $\{ih, (j+\frac{1}{2})k\}$ and replace $\frac{\delta^2 U}{\delta x^2}$ by the mean of its finite difference approximations at the jth and (j+1)th time levels. In other words they approximated the equation

$$\left(\frac{\delta U}{\delta t}\right)_{i,j+\frac{1}{2}} = \left(\frac{\delta^2 U}{\delta x^2}\right)_{i,j+\frac{1}{2}}$$

by

$$\frac{w_{i,j+1} - w_{ij}}{k} = \frac{1}{2}\left\{\frac{w_{i+1j+1} - 2w_{ij+1} + w_{i-1j+1}}{h^2} + \frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h^2}\right\}$$

giving

$$-rw_{i-1j+1} + (2+2r)w_{ij+1} - rw_{i+1j+1} = rw_{i-1j} + (2-2r)w_{ij} + rw_{i+1j} \tag{18}$$

with $r = \frac{k}{h^2}$.
In general the LHS contains 3 unknowns and the RHS 3 known pivotal values. Figure. See notes.
If there are N intervals mesh points along each row then for j=0 and i=1,..,N it gives N simultaneous equations for N unknown pivotal values along the first row.

Example.

$$\frac{\delta U}{\delta t} = \frac{\delta^2 U}{\delta x^2} \quad 0 < x < 1 \quad t > 0$$

where

- $U = 0$, $x = 0$ *and* $x = 1$ $t \geq 0$;
- $U = 2x$ $0 \leq x \leq \frac{1}{2}$ *and* $t = 0$;
- $U = 2(1-x)$ $\frac{1}{2} \leq x \leq 1$ *and* $t = 0$;

Choosing $h = \frac{1}{10}$ *and* $k = \frac{1}{100}$ *and* $r = 1$, *while all finite values of r are valid a large value of h will lead to inaccurate solution.*
(18) becomes

$$-w_{i-1j+1} + 4w_{ij+1} - w_{i+1j+1} = w_{i-1j} - w_{i+1j}$$

Due to symmetry $U_{4j} = U_{6j}$ *so we only need consider* $w_{0j}, ..., w_{5j}$ *and for notation simplicity we drop the j when dealing with a specific case.*
Considering when j=0.

$$-0 + 4w_1 - w_2 = 0 + 0.4$$
$$-w_1 + 4w_2 - w_3 = 0.2 + 0.6$$
$$-w_2 + 4w_3 - w_4 = 0.4 + 0.8$$
$$-w_3 + 4w_4 - w_5 = 0.6 + 1.0$$
$$-2w_4 + 4w_5 = 0.8 + 0.8$$

using the Thomas algorithm we get $w_1 = 0.1989$, $w_2 = 0.3956$, $w_3 = 0.5834$, $w_4 = 0.7381$ *and* $w_5 = 0.7691$, *and so forth.*
This yields a good numerical solution. $\diamond$

6.3 The Theta Method

The Theta Method is a generalization of the Crank-Nicholson method and expresses our partial differential equation as

$$\frac{w_{i,j+1} - w_{ij}}{k} = \left\{ \theta \left(\frac{w_{i+1j+1} - 2w_{ij+1} + w_{i-1j+1}}{h^2} \right) + (1-\theta) \left(\frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h^2} \right) \right\} \tag{19}$$

- when $\theta = 0$ we get the explicit scheme,
- when $\theta = \frac{1}{2}$ we get the Crank-Nicholson scheme,
- and $\theta = 1$ we get fully implicit backward finite difference method.

The equations are unconditionally valid for $\frac{1}{2} \leq \theta \leq 1$. For $0 \leq \theta < \frac{1}{2}$ we must have

$$r \leq \frac{1}{2(1-2\theta)}$$

.

6.3.1 Derivative Boundary Conditions

Boundary conditions expressed in terms of derivatives occur frequently.

Example.

$$\frac{\partial U}{\partial x} = H(U - v_0) \quad at\ x = 0$$

where H is a positive constant and v_0 is the surrounding temp.

How do we deal with this type of boundary condition?

1. By using forward difference for $\frac{\partial U}{\partial x}$, we have

$$\frac{w_{1j} - w_{0j}}{h_x} = H(w_{0j} - v_0)$$

   where $h_x = x_1 - x_0$. This gives us one extra equation for the temp w_{ij}.

2. If we wish to represent $\frac{\partial U}{\partial x}$ more accurately at x=0, we use a central difference formula. It is necessary to introduce a fictitious temp w_{-1j} at the external mesh points $(-h_x, jk)$. The temp w_{-1j} is unknown and needs another equation. This is obtained by assuming that the heat conduction equation is satisfied at the end points. The unknown w_{-1j} can be eliminated between these equations.

Example. *Solve for the equation*

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$$

satisfying the initial condition

$$U = 1 \text{ for } 0 \leq x \leq 1 \text{ when } t = 0$$

and the boundary conditions

$$\frac{\partial U}{\partial x} = U \ \text{ at } x = 0 \text{ for all } t$$

$$\frac{\partial U}{\partial x} = -U \ \text{ at } x = 1 \text{ for all } t.$$

Case 1 *Using forward difference approximation for the derivative boundary condition and the explicit method to approximate the* ***PDE****.*
Our difference equation is,

$$\frac{w_{i,j+1} - w_{ij}}{k} = \frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h^2}$$

$$w_{ij+1} = w_{ij} + r(w_{i-1j} - 2w_{ij} + w_{i+1j}) \tag{20}$$

where $r = \frac{k}{h_x^2}$.
At i=1, (20) is,

$$w_{1j+1} = w_{1j} + r(w_{0j} - 2w_{1j} + w_{2j}) \tag{21}$$

The boundary condition at x=0 is $\frac{\partial U}{\partial x} = U$ *in terms of forward difference this is*

$$\frac{w_{1j} - w_{j0}}{h_x} = w_{0j}$$

rearranging

$$w_{0j} = \frac{w_{1j}}{1 + h_x} \tag{22}$$

Using (22) and (21) to eliminate we get,

$$w_{1j+1} = \left(1 - 2r + \frac{r}{1 + h_x}\right) w_{1j} + r w_{2j}.$$

It will be proven, that the scheme is valid for $0 \leq r \leq \frac{1}{2}$. *Choose* $h_s = \frac{1}{10}$ *and* $k = \frac{1}{400}$ *such that* $r = \frac{1}{4}$.
The equations become

$$w_{1j+1} = \frac{8}{11} w_{1j} + \frac{1}{4} w_{2j}$$

$$w_{0j+1} = \frac{10}{11} w_{1j+1}$$

$$w_{ij+1} = \frac{1}{4}(w_{i-1j} + 2w_{ij} + w_{i+1j}) \quad i = 2, 3, 4$$

and

$$w_{5j+1} = \frac{1}{4}(2w_{4j} + 2w_{5j}) \quad \textit{by symmetry}$$

Solving with an initial guess of U=1, at j=0 we have

$$w_{11} = \frac{8}{11} 1 + \frac{1}{4} 1 = 0.9773$$

$$w_{01} = \frac{10}{11} 0.9773 = 0.8884$$

$$w_{i1} = \frac{1}{4}(1 + 2 + 1) = 1 \quad i = 2, 3, 4$$

and

$$w_{5j+1} = \frac{1}{4}(2 + 2) = 1 \quad \textit{by symmetry}$$

Case 2 *Using central difference approximation for the derivative boundary condition and the explicit method to approximate the* ***PDE****.*
Our difference equation is as in (20).
At i=0 we have

$$w_{0j+1} = w_{0j} + r(w_{-1j} - 2w_{0j} + w_{1j}) \tag{23}$$

The boundary condition at x=0, in terms of central differences can be written as

$$\frac{w_{1j} - w_{-1j}}{2h_x} = w_{0j} \tag{24}$$

Using (24) and (23) to eliminate the fictitious term w_{-1j} we get,

$$w_{0j+1} = w_{0j} + 2r((-1 - h_x)w_{0j} + w_{1j})$$

As before let $h_x = 0.1$. Then at x=1 out difference equation becomes

$$w_{10j+1} = w_{10j} + r(w_{9j} - 2w_{10j} + w_{11j})$$

and the boundary condition is

$$\frac{w_{11j} - w_{9j}}{2h_x} = w_{10j}$$

Eliminating the fictitious term w_{11j} we have

$$w_{10j+1} = w_{10j} + 2r(w_{9j} - (1 + h_x)w_{10j})$$

Choosing $r = \frac{1}{4}$ we have

$$w_{1j+1} = \frac{1}{2}(0.9w_{0j} + w_{1j})$$

$$w_{ij+1} = \frac{1}{4}(w_{i-1j} + 2w_{ij} + w_{i+1j}) \quad i = 1, 2, 3, 4$$

and

$$w_{5j+1} = \frac{1}{4}(2w_{4j} + 2w_{5j}) \quad \textit{by symmetry}$$

With the initial condition U=1 and k=0.025, at j=0 we have

$$w_{01} = \frac{1}{2}(0.9 + 1) = 0.95$$

$$w_{i1} = \frac{1}{4}(1 + 2 + 1) = 1 \quad i = 1, 2, 3, 4$$

and

$$w_{5j+1} = \frac{1}{4}(2 + 2) = 1 \quad \textit{by symmetry}$$

While this method yields more accurate results, both are acceptable.

Case 3 *Using central difference approximation for the derivative boundary condition and the Crank-Nicholson method to approximate the **PDE**.*
The difference equation is,

$$\frac{w_{i,j+1} - w_{ij}}{k} = \frac{1}{2}\left\{\frac{w_{i+1j+1} - 2w_{ij+1} + w_{i-1j+1}}{h^2} + \frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h^2}\right\}$$

giving

$$-rw_{i-1j+1} + (2 + 2r)w_{ij+1} - rw_{i+1j+1} = rw_{i-1j} + (2 - 2r)w_{ij} + rw_{i+1j} \tag{25}$$

with $r = \frac{k}{h^2}$.
The boundary condition at x=0, in terms of central differences can be written as

$$\frac{w_{1j} - w_{-1j}}{2h_x} = w_{0j}$$

Rearranging we have

$$w_{-1j} = w_{1j} - 2h_x w_{0j} \tag{26}$$

and

$$w_{-1j+1} = w_{1j+1} - 2h_x w_{0j+1} \tag{27}$$

Let j=0 and i=0 the difference equation becomes

$$-rw_{-11} + (2 + 2r)w_{01} - rw_{11} = rw_{-10} + (2 - 2r)w_{00} + rw_{10} \tag{28}$$

Using, (26), (27) and (28) we can eliminate the fictious terms w_{-1j} and w_{-1j+1}. Also due to symmetry around $\frac{1}{2}$ we have $w_{4j} = w_{6j}$. Choosing $h_x = 0.1$, $k = 0.01$ and $r = 1$, our equations become

$$2.1w_{0,j+1} - w_{1j+1} = -0.1w_{0j} + w_{1j}$$

$$-w_{i-1j+1} + 4w_{ij+1} - w_{i+1j+1} = w_{i-1j} + w_{i+1j} \quad i = 1, 2, 3, 4$$

$$-w_{4j+1} + 2w_{5j+1} = w_{4j}$$

For the time step j=0 we have

$$2.1w_0 - w_1 = 0.9$$

$$-w_{i-1} + 4w_i - w_{i+1} = 2 \quad i = 1, 2, 3, 4$$

$$-w_4 + 2w_5 = 1$$

This method yields very good results.

6.4 Local Truncation Error and Consistency

6.4.1 Local Truncation

Let $F_{ij}(w)$ represent the difference equation approximating the **PDE** at the ijth point with exact solution w.
If w is replaced by U at the mesh points of the difference equation where U is the exact solution of the PDE, the value of $F_{ij}(U)$ is the local truncation error T_{ij} in at the $ijth$ mesh point.
Using Taylor expansions it is easy to express T_{ij} in terms of h_x and k and partial derivatives of U at (ih_x, jk). Although U and its derivatives are generally unknown, it is worthwhile because it provides a method for comparing the local accuracies of different difference schemes approximating the **PDE**.

Example. *The local truncation error of the classical explicit difference approach to*

$$\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} = 0$$

with

$$F_{ij}(w) = \frac{w_{ij+1} - w_{ij}}{k} - \frac{w_{i+1j} - 2w_{ij} + w_{i-1j}}{h_x^2} = 0$$

is

$$T_{ij} = F_{ij}(U) = \frac{U_{ij+1} - U_{ij}}{k} - \frac{U_{i+1j} - 2U_{ij} + U_{i-1j}}{h_x^2} = 0$$

By Taylors expansions, we have

$$\begin{aligned}
U_{i+1j} &= U((i+1)h_x, jk) = U(x_i + h, t_j) \\
&= U_{ij} + h_x \left(\frac{\partial U}{\partial x}\right)_{ij} + \frac{h_x^2}{2}\left(\frac{\partial^2 U}{\partial x^2}\right)_{ij} + \frac{h_x^3}{6}\left(\frac{\partial^3 U}{\partial x^3}\right)_{ij} + ... \\
U_{i-1j} &= U((i-1)h_x, jk) = U(x_i - h, t_j) \\
&= U_{ij} - h_x \left(\frac{\partial U}{\partial x}\right)_{ij} + \frac{h_x^2}{2}\left(\frac{\partial^2 U}{\partial x^2}\right)_{ij} - \frac{h_x^3}{6}\left(\frac{\partial^3 U}{\partial x^3}\right)_{ij} + ... \\
U_{ij+1} &= U(ih_x, (j+1)k) = U(x_i, t_j + k) \\
&= U_{ij} + k \left(\frac{\partial U}{\partial t}\right)_{ij} + \frac{k^2}{2}\left(\frac{\partial^2 U}{\partial t^2}\right)_{ij} + \frac{k^3}{6}\left(\frac{\partial^3 U}{\partial t^3}\right)_{ij} + ...
\end{aligned}$$

Substitution into the expression for T_{ij} gives

$$\begin{aligned}
T_{ij} &= \left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{ij} + \frac{k}{2}\left(\frac{\partial^2 U}{\partial t^2}\right)_{ij} - \frac{h_x^2}{12}\left(\frac{\partial^4 U}{\partial x^4}\right)_{ij} \\
&\quad + \frac{k^2}{6}\left(\frac{\partial^3 U}{\partial t^3}\right)_{ij} - \frac{h_x^4}{360}\left(\frac{\partial^6 U}{\partial x^6}\right)_{ij} + ...
\end{aligned}$$

But U is the solution to the differential equation so

$$\left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{ij} = 0$$

the principal part of the local truncation error is

$$\frac{k}{2}\left(\frac{\partial^2 U}{\partial t^2}\right)_{ij} - \frac{h_x^2}{12}\left(\frac{\partial^4 U}{\partial x^4}\right)_{ij}.$$

Hence

$$T_{ij} = O(k) + O(h_x^2)$$

•

6.5 Consistency and Compatibility

It is sometimes possible to approximate a parabolic or hyperbolic equation with a finite difference scheme that is stable but which does not converge to the solution of differential equation as the mesh lengths tend to zero. Such a scheme is called inconsistent or incompatible.
This is useful when considering the theorem which states that is a linear finite difference equation is consistent with a properly posed linear IVP then stability guarantees convergence of w to U as the mesh lengths tend to zero.

Definition. *Let $L(U) = 0$ represent the **PDE** in the independent variables x and t with the exact solution U.*
Let $F(w) = 0$ represent the approximate finite difference equation with exact solution w.
Let v be a continuous function of x and t with sufficient derivatives to enable $L(v)$ to be evaluated at the point (ih_x, jk). Then the truncation error $T_{ij}(v)$ at (ih_x, jk) is defined by

$$T_{ij}(v) = F_{ij}(v) - L(v_{ij})$$

*If $T_{ij}(v) \to 0$ as $h \to 0$, $k \to 0$ the difference equation is said to be consistent or compatible with the with the **PDE**.* ∘

Looking back at the previous example it follows that the classical explicit approximation to

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$$

is consistent with the difference equation.

Example. *The equation*

$$\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} = 0$$

is approximated by the difference equation

$$\frac{w_{ij+1} - w_{ij-1}}{2k} - \frac{w_{i+1j} - 2(\theta w_{ij+1} + (1-\theta)w_{ij-1}) + w_{i-1j}}{h_x^2}$$

has a truncation error of

$$\begin{aligned} T_{ij} \quad = \quad & \left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{ij} + \\ & \left\{\frac{k^2}{6}\frac{\partial^3 U}{\partial t^3} - \frac{h_x^2}{12}\frac{\partial^4 U}{\partial x^4} + (2\theta - 1)\frac{r}{h_x^2}\frac{\partial U}{\partial t} - \frac{k^2}{h_x^2}\frac{\partial^2 U}{\partial t^2}\right\}_{ij} + \\ & O(\frac{k^3}{h_x^2}, h_x^4, k^4) \end{aligned}$$

Case1 $k = rh$ *As* $h \to 0$

$$T_{ij} = F_{ij}(U) \to \left\{ \frac{k^2}{6}\frac{\partial^3 U}{\partial t^3} - \frac{h_x^2}{12}\frac{\partial^4 U}{\partial x^4} + (2\theta - 1)\frac{r}{h_x^2}\frac{\partial U}{\partial t} - \frac{k^2}{h_x^2}\frac{\partial^2 U}{\partial t^2} \right\}_{ij}$$

When $\theta \neq \frac{1}{2}$ *the third term tends to infinity.*
When $\theta = \frac{1}{2}$ *the limiting values of* T_{ij} *is*

$$\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} - r^2\frac{\partial^2 U}{\partial t^2} = 0$$

Hence the difference equation is always inconsistent with $\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} = 0$ *when* $k = rh$. ***casew*** $k = rh^2$ *As* $h \to 0$

$\diamond$

6.6 Convergence and Stability

Definition. *By convergence we mean that the results of the method approach the analytical solution as* k *and* h_x *tends to zero.* $\circ$

Definition. *By stability we mean that errors at one stage of the calculations do not cause increasingly large errors as the computations are continued.* $\circ$

6.6.1 Analytical treatment of convergence

Assuming no round off error the only difference between the exact and the numerical result is error. Consider the equation

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2} \quad 0 < x < 1 \quad t > 0$$

Let U_{ij} represent the exact solution and w_{ij} be the numerical solution of the difference equation. Assuming no roundoff error the only difference between the two will be the error e_{ij}.

Example. *Explicit finite difference approximation is*

$$\frac{w_{ij+1} - w_{ij}}{k} = \frac{w_{i-1j} - 2w_{ij} + w_{i+1j}}{h_x^2}$$

at the mesh points

$$w_{ij} = U_{ij} - e_{ij} \quad w_{ij+1} = U_{ij+1} - e_{i+1j}$$

We substitute it into the difference equation

$$e_{ij+1} = re_{i-1j} + (1 - 2r)e_{ij} + re_{i+1j} + U_{ij+1} - U_{ij} + r(2U_{ij} - U_{i-1j} - U_{i+1j})$$

By Taylors Theorem, we have

$$\begin{aligned}
U_{i+1j} &= U(x_i + h, t_j) &&= U_{ij} + h_x^2\left(\frac{\partial U}{\partial x}\right)_{ij} + \frac{h_x^2}{2}\frac{\partial^2 U}{\partial x^2}(x_i + \theta_1 h, t_j) \\
U_{i-1j} &= U(x_i + h, t_j) &&= U_{ij} - h_x^2\left(\frac{\partial U}{\partial x}\right)_{ij} + \frac{h_x^2}{2}\frac{\partial^2 U}{\partial x^2}(x_i - \theta_2 h, t_j) \\
U_{i+1j} &= U(x_i, t_j + k) &&= U_{ij} + k\frac{\partial U}{\partial t}(x_i, t_j + \theta_3 k)
\end{aligned}$$

Where $0 < \theta_1, \theta_2, \theta_3 < 1$. *Substituting it into the original equation gives*

$$e_{ij+1} = re_{i-1j} + (1 - 2r)e_{ij} + re_{i+1j} + k\left\{\frac{\partial U}{\partial x}(x_i, t_j + \theta_3 k) - \frac{\partial^2 U}{\partial x^2}(x_i - \theta_4 h, t_j)\right\}$$

Where $-1 < \theta_4, \theta_3 < 1$.
This is a difference equation for e_{ij} *which we need not solve. Let* E_j *denote the max value along the jth time row and M the maximum modulus of the expression in* $\{\}$.
When $r \leq \frac{1}{2}$, *all coefficients of e in the eqn are positive or zero so*

$$\begin{aligned} |e_{ij+1}| &\leq r|e_{i-1j}| + (1-2r)|e_{ij}| + r|e_{i+1j}| + kM \\ &\leq rE_j + (1-2r)E_j + rE_j + kM \\ &= E_j + kM \end{aligned}$$

Also

$$E_{j+1} \leq E_j + kM \leq (E_{j-1} + kM) + kM \leq ... \leq E_0 + (j+1)kM = t_{j+1}M$$

Since we are dealing with non derivative boundary conditions $E_0 = 0$. *When* $h \to 0$ $k = rh^2$ *also tends to 0 and M tends to*

$$\left(\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2}\right)_{ij}$$

Since the numerical method is consistent the value of M and therefore $E_{j+1} \to 0$.
As $|U_{ij} - w_{ij}| \leq E_j$ *we can say* $w \to U$ *as h tends to 0 when* $r \leq \frac{1}{2}$.
When $r > \frac{1}{2}$ *it can be shown that the complimentary function tends to* ∞ *as h tend to 0.*
$\diamond$

This can also be applied to other methods. Another approach is to look at the matrix form.

6.6.2 A more Analytical Argument

Let the solution domain of the **PDE** be the finite rectangle $0 \leq x \leq 1$ and $0 \leq t \leq T$ and subdivide it into a uniform rectangular mesh by the lines $x_i = ih$ for $i = 0$ to N and $t_j = jk$ for $j = 0$ to J it will be assumed that h is related to k by some relationship such as $k = rh$ or $k = rh^2$ with $r > 0$ and finite so that as $h \to 0$ as $k \to 0$.
Assume that the finite difference equation relating the mesh point values along the $(j+1)$th and jth row is

$$b_{i-1}w_{i-1j+1} + b_i w_{ij+1} + b_{i+1}w_{i+1j+1} = c_{i-1}w_{i-1j} + c_i w_{ij} + c_{i+1}w_{i+1j}$$

where the coefficients are constant. If the boundary values at $i = 0$ and N for $j > 0$ are known these $(N-1)$ equations for $i = 1$ to $N-1$ can be written in matrix form.

$$\begin{pmatrix} b_1 & b_2 & 0 & . & . & . & . \\ b_1 & b_2 & b_3 & 0 & . & . & . \\ 0 & b_2 & b_3 & b_4 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & b_{N-3} & b_{N-2} & b_{N-1} \\ . & . & . & . & . & b_{N-2} & b_{N-1} \end{pmatrix} \begin{pmatrix} w_{1j+1} \\ w_{2j+1} \\ . \\ . \\ w_{N-2j+1} \\ w_{N-1j+1} \end{pmatrix}$$

$$= \begin{pmatrix} c_1 & c_2 & 0 & . & . & . & . \\ c_1 & c_2 & c_3 & 0 & . & . & . \\ 0 & c_2 & c_3 & c_4 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & c_{N-3} & c_{N-2} & c_{N-1} \\ . & . & . & . & . & c_{N-2} & c_{N-1} \end{pmatrix} \begin{pmatrix} w_{1j} \\ w_{2j} \\ . \\ . \\ w_{N-2j} \\ w_{N-1j} \end{pmatrix} + \begin{pmatrix} c_0 w_{0j} - b_0 w_{0j+1} \\ 0 \\ . \\ . \\ 0 \\ c_N w_{Nj} - b_N w_{Nj+1} \end{pmatrix}$$

Which can be written as

$$B\mathbf{w}_{j+1} = C\mathbf{u}_j + \mathbf{d}_j$$

Where B and C are of order $(N-1)$ $\mathbf{w}_j$ denotes a column vector and $\mathbf{d}_j$ denotes a column vector of boundary values.
Hence

$$\mathbf{w}_{j+1} = B^{-1}C\mathbf{w}_j + B^{-1}\mathbf{d}_j.$$

Expressed in a more conventional manner as

$$\mathbf{w}_{j+1} = A\mathbf{w}_j + \mathbf{f}_j$$

Where $A = B^{-1}C$ and $\mathbf{f}_j = B^{-1}\mathbf{d}_j$. Applied recursively this leads to

$$\begin{aligned}
\mathbf{w}_{j+1} &= A\mathbf{w}_j + \mathbf{f}_j = A(A\mathbf{w}_{j-1} + \mathbf{f}_{j-1}) + \mathbf{f}_j \\
&= A(A\mathbf{w}_{j-1} + \mathbf{f}_{j-1}) + \mathbf{f}_j \\
&= A^2\mathbf{w}_{j-1} + A\mathbf{f}_{j-1} + \mathbf{f}_j \\
&\cdot \\
&\cdot \\
&\cdot \\
&= A^{j+1}\mathbf{w}_0 + A^j\mathbf{f}_0 + A^{j-1}\mathbf{f}_1 + ... + \mathbf{f}_j
\end{aligned}$$

where $\mathbf{w}_0$ is the vector with initial values and $\mathbf{f}_0, \mathbf{f}_1, ..., \mathbf{f}_j$ are vectors of known boundary values. When we are concerned with a numerical solution, the constant vectors can be eliminated by investigating the error. Perturb the initial solution $\mathbf{w}_0$ to $\hat{\mathbf{w}}_0$. The solution at the jth time row will be given by

$$\hat{\mathbf{w}}_j = A^j\hat{\mathbf{w}}_0 + A^{j-1}\mathbf{f}_0 + A^{j-1}\mathbf{f}_1 + ... + \mathbf{f}_j$$

If the error $\mathbf{e}$ is defined by

$$\mathbf{e} = \hat{\mathbf{w}} - \mathbf{w}$$

It follows that

$$\mathbf{e}_j = \hat{\mathbf{w}}_j - \mathbf{w}_j = A^j(\hat{\mathbf{w}}_0 - \mathbf{w}_0) = A^j e_0$$

In other words, a perturbation e_0 of the initial values will propagate according to the equation

$$e_j = Ae_{j-1} = A^2e_{j-2} = ... = A^j e_0$$

Hence, for compatible matrix and vector norms

$$||e_j|| \leq ||A^j||||e_0||$$

Lax and Ritchmyer define the difference scheme to be stable when there exists a positive number M independent of j, k such that

$$||A^j|| \leq M$$

The clearly limits the amplification of any initial perturbation and of any arbitrary initial rounding errors.

$$||\mathbf{e}_j|| \leq M||\mathbf{e}_0||$$

Since

$$||A^j|| = ||AA^{j-1}|| \leq ||A||||A^{j-1}|| \leq ... \leq ||A||^j$$

It follows that the Lax-Ritchmyer Definition of stability is satisfied by

$$||A|| \leq 1$$

This is the necessary and sufficient condition for the difference equation to be stable then the solution of the **PDE** does not increase as $t \to T$.
When the condition is satisfied it follows automatically that the spectral radius

$$\rho(A) \leq 1$$

since $\rho(A) \leq ||A||$. If, however $\rho(A) \leq 1$ it does not imply that $||A|| \leq 1$.

Example. *Consider the classical explicit equation*

$$w_{ij+1} = rw_{i-1j} + (1-2r)w_{ij} + rw_{i+1j} \quad i = 1, .., N-1$$

for which A is

$$\begin{pmatrix} 1-2r & r & 0 & . & . & . & . \\ r & 1-2r & r & 0 & . & . & . \\ 0 & r & 1-2r & r & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & r & 1-2r & r \\ . & . & . & . & . & r & 1-2r \end{pmatrix}$$

where $r = \frac{k}{h_x^2} > 0$ *and it is assumed that the boundary values* w_{0j} *and* w_{Nj} *are known for* $j = 1, 2, ...$ *When* $1 - 2r \geq 0$ *then* $0 \leq r \leq \frac{1}{2}$ *and*

$$||A||_\infty = r + 1 - 2r + r = 1.$$

When $1 - 2r < 0$ *then* $r > \frac{1}{2}$ *then* $|1 - 2r| = 2r - 1$ *and*

$$||A||_\infty = r + 2r - 1 + r = 4r - 1 > 1$$

there fore the scheme is unstable for $r > \frac{1}{2}$ *and stable for* $0 < r \leq \frac{1}{2}$.
Alternatively since A is real and symmetric

$$||A||_2 = \rho(A) = \max_j |\mu_j|$$

where $|\mu_j|$ *is the jth eigenvalue of A. Now A can be written as*

$$\begin{pmatrix} 1 & 0 & 0 & . & . & . & . \\ 0 & 1 & 0 & 0 & . & . & . \\ 0 & 0 & 1 & 0 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & 0 & 1 & 0 \\ . & . & . & . & . & 0 & 1 \end{pmatrix} + r \begin{pmatrix} -2 & 1 & 0 & . & . & . & . \\ 1 & -2 & 1 & 0 & . & . & . \\ 0 & 1 & -2 & 1 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & 1 & -2 & 1 \\ . & . & . & . & . & 1 & -2 \end{pmatrix} = I_{N-1} + rT_{N-1}$$

where I_{N-1} *is the unit matrix of order N-1 and* T_{N-1} *is an* $(N-1) \times (N-1)$ *matrix whose eigen values* λ_j *are given by*

$$\lambda_j = -4sin^2(\frac{j\pi}{2N})$$

Hence the eigenvalues of A are

$$\mu_j = 1 - 4rsin^2(\frac{j\pi}{2N})$$

therefore the equation will be stable when

$$||A||_2 = \max_s |1 - 4rsin^2(\frac{s\pi}{2N})| \leq 1$$

ie

$$-1 \leq 1 - 4rsin^2(\frac{s\pi}{2N}) \leq 1 \quad s = 1, ..N-1$$

the left hand side of the inequality gives

$$r \leq \frac{1}{2}sin^2(\frac{(N-1)\pi}{2N})$$

as $h \to 0 N \to \infty$ *and* $sin^2(\frac{(N-1)\pi}{2N}) \to 1$. *Hence* $r \leq \frac{1}{2}$.
Therefore it is only stable when $0 < r \leq \frac{1}{2}$. $\diamond$

Example. *The Crank Nicholson equation*

$$-rw_{i-1j+1} + (2+2r)w_{ij+1} - rw_{i+1j+1} = rw_{i-1j} + (2-2r)w_{ij} + rw_{i+1j} \quad i = 1, .., N-1$$

In matrix form

$$\begin{pmatrix} 2+2r & -r & 0 & . & . & . & . \\ -r & 2+2r & -r & 0 & . & . & . \\ 0 & -r & 2+2r & -r & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & -r & 2+2r & -r \\ . & . & . & . & . & -r & 2+2r \end{pmatrix} \begin{pmatrix} w_{1j+1} \\ w_{2j+1} \\ . \\ . \\ w_{N-2j+1} \\ w_{N-1j+1} \end{pmatrix}$$

$$= \begin{pmatrix} 2-2r & r & 0 & . & . & . & . \\ r & 2-2r & r & 0 & . & . & . \\ 0 & r & 2-2r & r & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & r & 2-2r & r \\ . & . & . & . & . & r & 2-2r \end{pmatrix} \begin{pmatrix} w_{1j} \\ w_{2j} \\ . \\ . \\ w_{N-2j} \\ w_{N-1j} \end{pmatrix} + \mathbf{b}_j$$

where $\mathbf{b}_j$ *is a vector of known boundary conditions. This can be written as*

$$(2I_{N-1} - rT_{N-1})\mathbf{w}_{j+1} = (2I_{N-1} + rT_{N-1})\mathbf{w}_j + \mathbf{b}_j$$

from which it follows that A is of the form

$$A = (2I_{N-1} - rT_{N-1})^{-1}(2I_{N-1} + rT_{N-1})$$

T_{N-1} *has the eigenvalues* $\lambda_s = -4sin^2(\frac{s\pi}{2N})$ *for* $s = 1, ..., N-1$. *It follows that the eigenvalues of A are*

$$\mu_s = \frac{2 - 4rsin^2(\frac{s\pi}{2N})}{2 + 4rsin^2(\frac{s\pi}{2N})}.$$

Thus

$$||A||_\infty = \rho(A) = \max_s \left| \frac{2 - 4rsin^2(\frac{s\pi}{2N})}{2 + 4rsin^2(\frac{s\pi}{2N})} \right| < 1$$

Therefore, Crank-Nicholson is unconditionally stable. $\diamond$

6.7 Stability by the Fourier Series method (compared to Newmann's method)

This method uses a Fourier series to express $w_{pq} = w(ph_x, qk)$ which is

$$w_{pq} = e^{i\beta x}\xi^q$$

where $\xi = e^{\alpha k}$ in this case i denotes the complex number $i = \sqrt{-1}$ and for values of β needed to satisfy the initial conditions. ξ is known as the amplification factor. The finite difference equation will be stable if $|w_{pq}|$ remains bounded for all q as $h \to 0, k \to 0$ and all β.
If the exact solution does not increase exponentially with time then a necessary and sufficient condition is that

$$|\xi| \leq 1$$

Example. *Investigating the stability of the fully implicit difference equation*

$$\frac{1}{k}(w_{pq+1} - w_{pq}) = \frac{1}{h_x^2}(w_{p-1q+1} - 2w_{pq+1} + w_{p+1q+1})$$

approximating $\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}$ *at* (ph_x, qk). *Substituting* $w_{pq} = e^{i\beta x}\xi^q$ *into the difference equation*

$$e^{i\beta ph}\xi^{q+1} - e^{i\beta ph}\xi^q = r\{e^{i\beta(p-1)h}\xi^{q+1} - 2e^{i\beta ph}\xi^{q+1} + e^{i\beta(p+1)h}\xi^{q+1}\}$$

where $r = \frac{k}{h_x^2}$. *Divide across by* $e^{i\beta(p)h}\xi^q$ *leads to*

$$\begin{aligned} \xi - 1 &= r\xi(e^{i\beta(-1)h} - 2 + e^{i\beta h}) \\ &= r\xi(2cos(\beta h) - 2) \\ &= -4r\xi(sin^2(\beta\frac{h}{2}) \end{aligned}$$

Hence

$$\xi = \frac{1}{1 + 4rsin^2(\frac{\beta h}{2})}$$

$0 < \xi \leq 1$ *for all* $r > 0$ *and all* β *therefore the equation is unconditionally stable.* •

7 Variational Methods

Variational methods are based on the fact that the solutions of some Boundary Value Problems,

$$\begin{aligned} &-(p(x)u^{'}(x))^{'} + q(x)u(x) = g(x, u(x)) \\ &u(a) = \alpha, \quad u(b) = \beta, \end{aligned} \tag{29}$$

under the assumptions that,

$$\begin{aligned} &p \in C^1[a, b], && p(x) \geq p_0 > 0, \\ &q \in C^1[a, b], && q(x) \geq 0, \\ &g \in C^1([a, b] \times R), && g_u(x, u) \leq \lambda_0 \end{aligned} \tag{30}$$

then if $u(x)$ is the solution of (29), it can be written in the form $y(x) = u(x) - l(x)$ with

$$l(x) = \alpha\frac{b - x}{b - a} + \beta\frac{a - x}{a - b}, \quad l(a) = \alpha, \quad l(b) = \beta,$$

and y is the solution of a boundary value problem

$$\begin{aligned} &-(p(x)y^{'}(x))^{'} + q(x)y(x) = f(x), \\ &y(a), = 0 \quad y(b) = 0, \end{aligned} \tag{31}$$

with zero boundary values. Without loss of generality we can just consider problems of the form (31), is known as the:

Classical Problem

$$\begin{aligned} -(p(x)u^{'}(x))^{'} + q(x)u(x) = f(x), \\ u(a) = 0, \quad u(b) = 0. \end{aligned}$$

The assumptions on the Classical Problem can be relaxed such that $f \in L_2([0, 1])$, such that

$$u(x) \in D_L = \{u \in C^2[a, b] \mid u(a) = 0, u(b) = 0\}.$$

Convolving the Classical Problem (D) with the function $v(x)$ gives the problem

$$\int_a^b [-(p(x)u^{'}(x))^{'} + q(x)u(x)]v(x)dx = \int_a^b f(x)v(x)dx,$$

where $v \in D_L$. Integrating by parts gives the simplified problem gives the:

Weak Form Problem

$$\int_a^b [p(x)u^{'}(x)v^{'}(x) + q(x)u(x)v(x)]dx = \int_a^b f(x)v(x)dx.$$

It is sufficient to solve the weak form of the Classical Problem. From the Weak Form, we have two definitions:

Definition. ***(Bilinear Form)***

$$a(u,v) = \int_a^b [p(x)u^{'}(x)v^{'}(x) + q(x)u(x)v(x)]dx;$$

Definition.

$$(f,v) = \int_a^b f(x)v(x)dx,$$

where $f \in L_2([a,b])$.

From these definitions the weak form of the ODE problem is then given by

$$a(u,v) = (f,v),$$

where $u \in D_L$ is the solution to the Classical Problem.
The Weak Form of the problem can be equivalently written in a *Variational or Minimisation* form of the problem is given by,

Variational/Minimization form:

$$F(v) = \frac{1}{2}a(v,v) - (f,v).$$

where $f \in L_2([a,b])$. This gives the problem

$$F(u) \leq F(v), \qquad \text{all } v \in D_L$$

such that the function u that minimizes F over D_L.

Theorem. *We have the following relationships between the solutions to the three problems* ***Classical Problem, Weak Form*** *and* ***Minimization Form****.*

1. *If the function* u *solves* ***Classical Problem****, then* u *solves* ***Weak Form****.*
2. *The function* u *solves* ***Weak Form*** *if and only if* u *solves* ***Minimization Form****.*
3. *If* $f \in C([0,1])$ *and* $u \in C^2([0,1])$ *solves* ***Weak Form****, then* u *solves* ***Classical Problem****.*

Proof. We consider to prove this by three steps.

1. Let u be the solution to **Classical Problem**; then u solves **Weak Form** is obvious, since the Weak Form (W) derives directly from **Classical Problem**.
2. (a) Show **Weak Form** $\Rightarrow$ **Minimization Form**.
   Let u solve **Weak Form**, and define $v(x) = u(x) + z(x)$, $u, z \in D_L$. By linearity
   $$\begin{aligned} F(v) &= \tfrac{1}{2}a(u+z,u+z) - (f,u+z) \\ &= F(u) + \tfrac{1}{2}a(z,z) + a(u,z) - (f,z) \\ &= F(u) + \tfrac{1}{2}a(z,z) \end{aligned}$$
   which implies that $F(v) \geq F(u)$, and therefore u solves **Minimization Form** .
   (b) Show **Weak Form** $\Leftarrow$ **Minimization Form**.
   Let u solve **Minimization Form** and choose $\varepsilon \in R$, $v \in D_L$. Then $F(u) \leq F(u + \varepsilon v)$, since $u + \varepsilon v \in D_L$. Now $F(u + \varepsilon v)$ is a quadratic form in ε and its minimum occurs at $\varepsilon = 0$ ie
   $$\begin{aligned} 0 &= \tfrac{dF(u+\varepsilon v)}{d\varepsilon}|_{\varepsilon=0} \\ &= a(u,v) - (f,v), \end{aligned}$$
   it follows that u solves the **Weak Form**.
3. Then, proof is trivial.

□

7.1 Ritz -Galerkin Method

This is a classical approach which we exploit to fined "discrete" approximation to the problem **Weak Form / Minimization Form**. We look for a solution u_S in a finite dimensional subspace S of D_L such that u_S is an approximation to the solution of the continuous problem,

$$u_S = u_1\phi_1 + u_2\phi_2 + ... + u_n\phi_n.$$

Discrete Weak Form (W_S):

Find $u_S \in S = span\{\phi_1, \phi_2, ..., \phi_n\}, \quad n < \infty$ such that

$$a(u_S, v) = (f, v),$$

$$u \approx u_S = u_1\phi_1 + u_2\phi_2 + ... + u_n\phi_n.$$

Similarly, we have:

Discrete Variational/Minimization form (M_S):

Find $u_S \in S = \text{span}\{\phi_1, \phi_2, ..., \phi_n\}, \quad n < \infty$ that satisfies

$$F(u_S) \leq F(v) \qquad \text{all } v \in S,$$

where

$$F(v) = \frac{1}{2}a(v, v) - (f, v).$$

$v \in D_L$.

Theorem. *Given $f \in L_2([0, 1])$, then (W_S) has a unique solution.*

Proof. We write $u_S = \sum_1^n u_j\phi_j(x)$ and look for constants u_j, $j = 1, ..., n$ to solve the discrete problem. We define

$$A = \{A_{ij}\} = \{a(\phi_i, \phi_j)\} = \int_a^b [p(x)\phi_i^{'}\phi_j^{'} + q(x)\phi_i\phi_j]dx$$

and

$$\bar{F} = \{F_j\} = \{(f, \phi_j)\} = \{\int_a^b f\phi_i dx\}$$

Then, we require

$$a(u_S, v) = a(\sum_1^n u_j\phi_j(x), v) = (f, v) \text{ all } v \in S$$

Hence, for each basis function $\phi_i \in S$ we must have,

$$a(u_S, \phi_i) = a(\sum_1^n u_j\phi_j(x), \phi_i) = (f, \phi_i) \text{ all } i = 1, ..., n \in S$$

this gives the matrix,

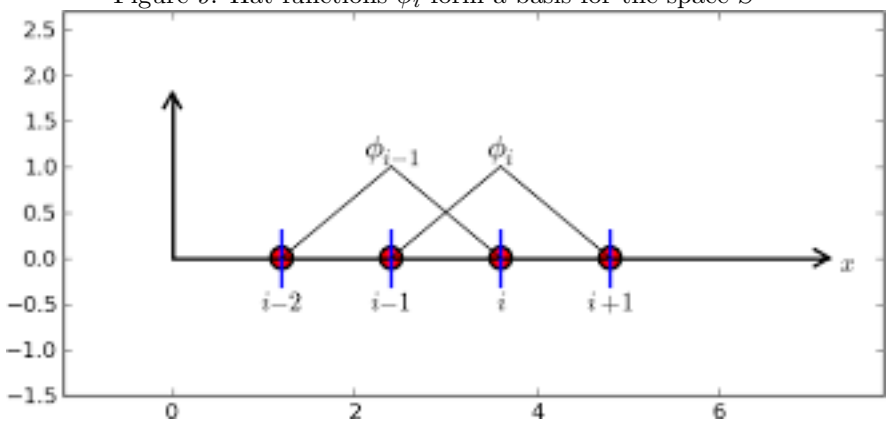
$$\begin{bmatrix} a(\phi_1, \phi_1) & ... & a(\phi_n, \phi_1) \\ . & . & . \\ . & . & . \\ . & . & . \\ a(\phi_1, \phi_n) & ... & a(\phi_n, \phi_n) \end{bmatrix} \begin{bmatrix} u_1 \\ . \\ . \\ . \\ u_n \end{bmatrix} = \begin{bmatrix} (f, \phi_1) \\ . \\ . \\ . \\ (f, \phi_n) \end{bmatrix}$$

which can be written as,

$$A\bar{u} = \bar{F}$$

Hence, u_S is found by the solution to a matrix equation. We now show existence and uniqueness of the solution to the algebraic problem. We show by contradiction that A is full-rank, i.e. that the only solution

Figure 9: Hat functions ϕ_i form a basis for the space S^h



to $A\bar{u} = 0$ is $\bar{u} = 0$.
Suppose that there exists a vector $\bar{v} = \{v_j\} \neq 0$ such that $A\bar{v} = 0$ and construct $v(x) = \sum_1^n v_j\phi_j \in S$. Then, we have:

$$\begin{aligned} A\bar{v} = 0 \quad &\Leftrightarrow \quad \textstyle\sum_j a(\phi_j, \phi_k)v_j = a(v, \phi_k) = 0 \text{ all } k \\ &\Leftrightarrow \quad \textstyle\sum_k a(v, \phi_k)v_j = a(v, \sum v_k\phi_k) = a(v, v) = 0 \\ &\Leftrightarrow \quad v = 0 \end{aligned}$$

Therefore, it can lead to a contradiction. □

Classically, in the Ritz-Galerkin method, the basis functions are chosen to be continuous functions over the entire interval $[a, b]$, for example, $\{\sin(mx), \cos(mx)\}$ give us trigonometric polynomial approximations to the solutions of the ODEs.

7.2 Finite Element

We choose the basis functions $\{\phi_i\}_1^n$ to be piecewise polynomials with compact support. In the simplest case ϕ_i is linear. We divide the region in to n intervals or "elements",

$$a = x_0 < x_1 < ... < x_n = b$$

and let E_i denote the element $[x_{i-1}, x_i]$, $h_i = x_i - x_{i-1}$.

Definition. *Let $S^h \subset D$ be the space of functions such that $v(x) \in [0, 1]$, $v(x)$ is linear on E_i and $v(a) = v(b) = 0$ ie*

$$S^h = \{v(x) : \text{ piecewise linear on } [0, 1], v(a) = v(b) = 0\}$$

The basis functions $\phi_i(x)$ for S^h are defined such that $\phi_i(x)$ is linear on E_i, E_{i+1} and $\phi_i(x_j) = \delta_{ij}$.
We now show that the hat functions ϕ_i form a basis for the space S^h (Figure 9).

Lemma 2. *The set of functions $\{\phi_i\}_i^n$ is a basis for the space S^h.*

Proof. We show first that the set $\{\phi_i\}_1^n$ is linearly independent. If $\sum_1^n c_i\phi_i(x) = 0$ for all $x \in [a, b]$, then taking $x = x_j$, implies $c_j = 0$ for each value of j, and hence the functions are independent.
To show $S^h = \text{span}\{\phi_i\}$, we only need to show that

$$v(x) = v_I = \sum v_j\phi_j, \text{ all } v(x) \in S^h$$

This is proved by construction. Since $(v - v_I)$ is linear on $[x_{i-1}, x_i]$ and $v - v_I = 0$ at all points x_j, it follows that $v = v_I$ on E_i. □

We now consider the matrix $A\hat{u} = \hat{F}$ in the case where the basis functions are chosen to be the "hat functions". In this case, the elements of A can be found. We have

$$\phi_i = 0, \phi_i^{'} = 0, \text{ for } x \notin [x_{i-1}, x_{i+1}) = E_i \bigcup E_{i+1},$$

where

$$\phi_i = \frac{x - x_{i-1}}{x_i - x_{i-1}} = \frac{1}{h_i}(x - x_{i-1}), \quad \phi_i^{'} = \frac{1}{h_i}, \text{ on } E_i.$$

and

$$\phi_i = \frac{x_{i+1} - x}{x_{i+1} - x_i} = \frac{1}{h_{i+1}}(x_{i+1} - x), \quad \phi_i^{'} = \frac{-1}{h_{i+1}}, \text{ on } E_{i+1}.$$

Therefore, we have the elements of the matrix A

$$\begin{aligned} A_{i,i} &= \int_{x_{i-1}}^{x_i} \tfrac{1}{h_i^2} p(x)dx + \int_{x_i}^{x_{i+1}} \tfrac{1}{h_{i+1}^2} p(x)dx \\ &\quad + \int_{x_{i-1}}^{x_i} \tfrac{1}{h_i^2}(x - x_{i-1})^2 q(x)dx + \int_{x_i}^{x_{i+1}} \tfrac{1}{h_{i+1}^2}(x_{i+1} - x)^2 q(x)dx, \\ A_{i,i+1} &= \int_{x_i}^{x_{i+1}} \tfrac{-1}{h_{i+1}^2} p(x)dx + \int_{x_i}^{x_{i+1}} \tfrac{1}{h_{i+1}^2}(x_{i+1} - x)(x - x_i) q(x)dx, \\ A_{i,i-1} &= \int_{x_{i-1}}^{x_i} \tfrac{-1}{h_i^2} p(x)dx + \int_{x_i}^{x_{i+1}} \tfrac{1}{h_i^2}(x_i - x)(x - x_{i-1}) q(x)dx, \end{aligned}$$

and

$$F_i = \int_{x_{i-1}}^{x_i} \tfrac{1}{h_i}(x - x_{i-1}) f(x)dx + \int_{x_i}^{x_{i+1}} \tfrac{1}{h_{i+1}}(x_{i+1} - x) f(x)dx.$$

7.3 Error bounds of Finite Element methods

Lemma 3. *Assume u_S solves $(\mathbf{W_S})$. Then*

$$a(u - u_S, w) = 0, \text{ for all } x \in S$$

Proof. Given that

$$a(u_S, w) = (f, w),$$

and

$$a(u, w) = (f, w),$$

for all $w \in S$. Since a is bilinear, taking the differences gives

$$a(u - u_S, w) = 0.$$

□

The error bounds we are interested in will be in term of the energy norm,

$$||v||_E = [a(v, v)]^{\frac{1}{2}}$$

for all $v \in D_L$. The function satisfies the properties:

$$||\alpha v||_E = \alpha ||a||_E, \quad ||v + z||_E \leq ||v||_E + ||z||_E$$

To show u_S is the best fit we show that

$$||u - u_S||_E = \min_{v \in S} ||u - v||_E$$

Proof. By the Cauchy -Schwarz Lemma, we have $|a(u,v)| \leq ||u||_E||v||_E$. Let $w = u_S - v \in S$. Using the previous lemma, we obtain

$$\begin{aligned} ||u-u_S||_E^2 &= a(u-u_s, u-u_s) \\ &\leq a(u-u_s, u-u_s) + a(u-u_s, w) \\ &\leq a(u-u_s, u-u_s+w) = a(u-u_s, u-v) \\ &\leq ||u-u_s||_E||u-v||_E. \end{aligned}$$

If $||u-u_S||_E = 0$, then the theorem holds. Otherwise

$$\min ||u-v||_E \leq ||u-u_S|| \leq \min ||u-v||_E,$$

the result follows. □

Theorem. *Error bounds*

$$||u-u_S||_E \leq Ch||u^{''}||_\infty$$

where C is a constant.

Proof. Firstly, from the previous theorem, we have that

$$||u-u_S||_E = \min_{v\in S} ||u-v||_E \leq ||u-u_I||_E$$

We look for a bound on $||u-u_I||_E$, where

$$u_I(x) = \sum_j \bar{u_j}\phi_j, \quad \bar{u_j} = u(x_j).$$

We assume that

$$u_S(x) = \sum_j u_j\phi_j$$

where $\mathbf{u} = \{u_j\}$ solves $A\mathbf{u} = \mathbf{F}$. We define $e = u - u_I$. Since $u_I \in S$ implies that u_I is piecewise linear, then $u_I^{''} = 0$. Therefore $e^{''} = u^{''}$. Looking at the subinterval $[x_i, x_{i+1}]$.
The Schwarz inequality yields the estimate

$$\begin{aligned} (e)^2 &\leq \int_{x_i}^x 1^2 d\xi \int_{x_i}^x (e'(\xi))^2 d\xi \\ &\leq (x-x_i)\int_{x_i}^x (e'(\xi))^2 d\xi \\ &\leq h_i \int_{x_i}^{x_{i+1}} (e'(\xi))^2 d\xi \end{aligned}$$

and thus

$$||e||_\infty^2 \leq h_i \int_{x_i}^{x_{i+1}} (e'(\xi))^2 d\xi \leq h_i^2 ||e^{'}||_\infty^2$$

Similarly,

$$\begin{aligned} (e^{'})^2 &\leq \int_{x_i}^x 1^2 d\xi \int_{x_i}^x (e^{''}(\xi))^2 d\xi \\ &\leq (x-x_i)\int_{x_i}^x (e''(\xi))^2 d\xi \\ &\leq h_i \int_{x_i}^{x_{i+1}} (e^{''}(\xi))^2 d\xi \end{aligned}$$

and thus

$$||e^{'}||_\infty^2 \leq h_i \int_{x_i}^{x_{i+1}} (e^{''}(\xi))^2 d\xi \leq h_i^2 ||e^{''}||_\infty^2$$

Finally we also have

$$\begin{aligned} a(e,e) = & \int_{x_i}^{x_{i+1}} (p(x)[e^{'}]^2 + q(x)[e(x)]^2)dx \\ & \leq ||p||_\infty \int_{x_i}^{x_{i+1}} [e^{'}]^2 + ||q||_\infty \int_{x_i}^{x_{i+1}} [e(x)]^2 dx \\ & \leq ||p||_\infty h_i^2 ||e^{''}||_\infty^2 + ||q||_\infty h_i^2 ||e^{'}||_\infty^2 \\ & \leq ||p||_\infty h_i^2 ||e^{''}||_\infty^2 + ||q||_\infty h_i^4 ||e^{'}||_\infty^2 \\ & \leq C h_i^2 ||u^{''}||_\infty^2 \end{aligned}$$

$$||u - u_S||_E = \min_{v \in S} ||u - v||_E \leq ||u - u_I||_E \leq Ch||u^{''}||_\infty$$

where $h = max\{h_i\}$. □