Part II

Generalities

Chapter 4

Finite Differences

Part I tackled a simple problem by simple methods. Along the way various general issues were revealed, such as discretisation, time-stepping, and large sparse linear equations. In Part II we look in greater detail at these general issues. There are three ways to turn continuous partial differential equations into a finite discrete problem. In this chapter we consider again Finite Differences, to be followed in subsequent chapters by Finite Elements and Spectral Methods.

In Finite Differencing, we hold the unknown functions at a discrete number of points on a grid. Normally the points are equispaced with a separation of Δx.

4.1 Higher orders

4.1.1 Central differencing

For equispaced points, we can use central differencing to obtain a second-order estimate for derivatives. We have already used expressions for the first and second derivatives

$$\begin{aligned} f_i' &= \frac{f_{i+1} - f_{i-1}}{2\Delta x} + O(\Delta x^2), \\ f_i'' &= \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta x} + O(\Delta x^2). \end{aligned}$$

We can use these to find higher order derivatives, e.g.

$$\begin{aligned} f_i''' &= \frac{f_{i+1}'' - f_{i-1}''}{2\Delta x} + O(\Delta x^2) \\ &= \frac{f_{i+2} - 2f_{i+1} + 2f_{i-1} - f_{i-2}}{2\Delta x^3} + O(\Delta x^2) \\ f_i'''' &= \frac{f_{i+1}'' - 2f_i'' + f_{i-1}''}{\Delta x} + O(\Delta x^2) \\ &= \frac{f_{i+2} - 4f_{i+1} + 6f_i - 4f_{i-1} + f_{i-2}}{\Delta x^4} + O(\Delta x^2), \end{aligned}$$

and so on. The coefficients in the even derivatives are the binomial coefficients in Pascal's triangle. The coefficients in the odd derivatives are the difference between the binomial coefficients with a positive and negative shift.

The local errors can be found using a Taylor series

$$\begin{aligned} f_{i+1} &= f(x = i\Delta x + \Delta x) \\ &= f_i + \Delta x f_i' + \tfrac{1}{2}\Delta x^2 f_i'' + \tfrac{1}{6}\Delta x^3 f_i''' + \tfrac{1}{24}\Delta x^4 f_i'''' + \ldots. \end{aligned}$$

Then for the difference in the first derivative above

$$f_{i+1} - f_{i-1} = 2\Delta x f_i' + \tfrac{1}{3}\Delta x^3 f_i''' + O(\Delta x^5).$$

Thus we have an expression for the local error

$$f_i' = \frac{f_{i+1} - f_{i-1}}{2\Delta x} - \tfrac{1}{6}\Delta x^2 f_i''' + O(\Delta x^4).$$

But we have above a second-order accurate expression for the third derivative f_i'''. Substituting this, we can obtain a fourth-order accurate expression for the first derivative,

$$f_i' = \frac{-\frac{1}{12}f_{i+2} - \frac{2}{3}f_{i+1} - \frac{2}{3}f_{i-1} + \frac{1}{12}f_{i-2}}{\Delta x} + O(\Delta x^4).$$

One should check these expression, that $f = 1, x, x^2, x^3, x^4$ all give the correct result $(0, 1, 0, 0, 0)$.

Similarly, we can obtain a fourth-order accurate expression for the second derivative

$$f_i'' = \frac{-\frac{1}{12}f_{i+2} + \frac{4}{3}f_{i+1} - \frac{5}{2}f_i + \frac{4}{3}f_{i-1} - \frac{1}{12}f_{i-2}}{\Delta x^2} + O(\Delta x^4).$$

4.1.2 One-sided differencing

At boundaries it is not possible to use central differencing, because one would need information about the function outside the domain. Hence there is a need to use information from points on just one side. Still with equispaced points we have

$$\begin{aligned} f_0' &= \frac{f_1 - f_0}{\Delta x} + O(\Delta x), \\ f_0'' &= \frac{f_2 - 2f_1 + f_0}{\Delta x} + O(\Delta x), \\ f_0''' &= \frac{f_3 - 3f_2 + 3f_1 - f_0}{\Delta x} + O(\Delta x) \end{aligned}$$

We can again find an expression for the local error using Taylor series. Thus for the one-sided first derivative above

$$f_1 - f_0 = \Delta x f_0' + \tfrac{1}{2}\Delta x^2 f_0'' + O(\Delta x^3).$$

Using the first-order expression above for the second derivative, we obtain the improved second-order accurate expression for the one-sided first derivative

$$f_0' = \frac{-\frac{1}{2}f_2 + 2f_1 - \frac{3}{2}f_0}{\Delta x} + O(\Delta x^2).$$

Similarly one can find

$$f_0'' = \frac{-f_3 + 4f_2 - 5f_1 + 2f_0}{\Delta x^2} + O(\Delta x^2).$$

These expressions were used in the driven-cavity problem when extracting information about derivatives on the boundary to second-order accuracy.

4.1.3 Non-equispaced points

To find an expression for the kth derivative $f^{(k)}(x_0)$ with accuracy $O(\Delta x^l)$ one fits a polynomial of degree $k+l$ through $k+l+1$ points $x_0+\Delta x_i$,

$$f(x_0+\Delta x_i) = a_0 + a_1\Delta x_i + a_2\Delta x_i^2 + \dots a_{k+l}\Delta x_i^{k+l}.$$

One solves for the polynomial coefficients a_j, typically using a computer-algebra package such as MAPLE. Finally the desired kth derivative is

$$f^{(k)}(x_0) = a_k\, k!.$$

Central differencing on equispaced points is to be preferred because it delivers one degree higher accuracy than the same number of points would give if they were not equispaced or the differencing was not central.

It is not normally worth going to higher accuracy that the lowest approximation or its first improvement. At higher accuracy, the expressions use information from a wide spread of points. It is known that fitting high order polynomials through fixed points soon generates polynomials which differ greatly from the function in between the fitting points. A much better approach would be to use Splines, see later.

4.2 Compact fourth-order Poisson solver

4.2.1 One-dimension version

First we consider

$$\frac{d^2\phi}{dx^2} = \rho.$$

Using fourth-order accurate central differencing, we have the discretised form

$$-\tfrac{1}{12}\phi_{i+2} + \tfrac{4}{3}\phi_{i+1} - \tfrac{5}{2}\phi_i + \tfrac{4}{3}\phi_{i-1} - \tfrac{1}{12}\phi_{i-2} = \Delta x^2\rho_i.$$

This wide numerical molecule gives problems next to the boundaries because it requests values of ϕ outside the boundary. The large width also tends to have large coefficients in the $O(\Delta x^4)$ error.

To find a more compact approach, we start by considering again the error in the second-order approximation to the second derivative

$$\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{\Delta x^2} = \phi_i'' + \tfrac{1}{6}\Delta x^2\phi_i'''' + O(\Delta x^4).$$

Now $\phi_i'' = \rho_i$ and so

$$\phi_i'''' = \rho_i'' = \frac{\rho_{i+1}-2\rho_i+\rho_{i+1}}{\Delta x^2} + O(\Delta x^2).$$

Hence

$$\frac{\phi_{i+1}-2\phi_i+\phi_{i-1}}{\Delta x^2} = \tfrac{1}{6}\rho_{i+1} + \tfrac{2}{3}\rho_i + \tfrac{1}{6}\rho_{i-1} + O(\Delta x^4).$$

This expression gives an algorithm for the solution of the Poisson problem in one dimension with a fourth-order accurate answer.

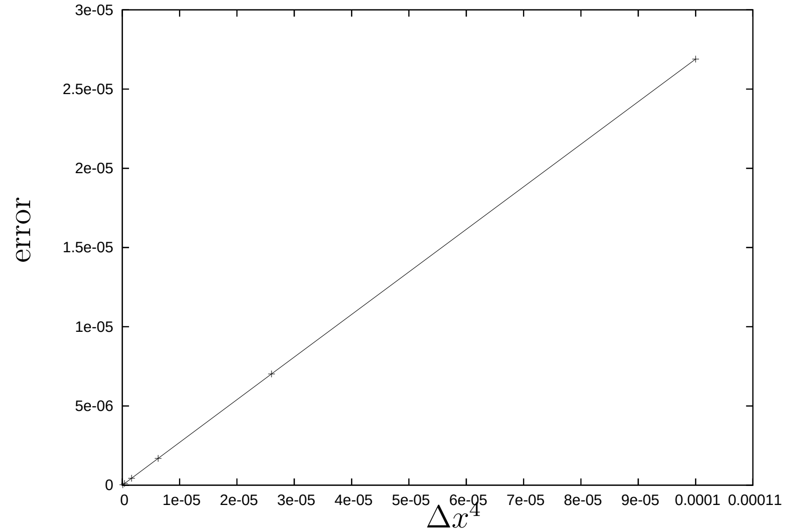

Figure 4.1: Test for fourth-order error, with $N = 10$, 14, 20, 40 and 56.

4.2.2 Two dimensions

We use the same idea now in two dimensions. Thus we use

$$\nabla^2 \rho = \nabla^2 \nabla^2 \phi = \frac{\partial^4 \phi}{\partial x^4} + 2 \frac{\partial^4 \phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \phi}{\partial y^4}.$$

There is a little complication that our standard numerical molecule for the Laplacian does not have an error of the above form, but

$$\begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix} \phi = \Delta x^2 \nabla^2 \phi + \tfrac{1}{12} \Delta x^4 \left(\frac{\partial^4 \phi}{\partial x^4} + \frac{\partial^4 \phi}{\partial y^4} \right).$$

Fortunately an alternative form of the Laplacian, one rotated by $\frac{1}{4}\pi$, has a different combination of fourth derivatives

$$\begin{pmatrix} \frac{1}{2} & & \frac{1}{2} \\ & -2 & \\ \frac{1}{2} & & \frac{1}{2} \end{pmatrix} \phi = \Delta x^2 \nabla^2 \phi + \tfrac{1}{12} \Delta x^4 \left(\frac{\partial^4 \phi}{\partial x^4} + 6 \frac{\partial^4 \phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \phi}{\partial y^4} \right).$$

Then taking $\frac{2}{3}$ of the first and $\frac{1}{3}$ of the second expression, we have an error which is the Laplacian of ρ. This gives the compact fourth-order accurate algorithm

$$\frac{1}{\Delta x^2} \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{2}{3} & -\frac{10}{3} & \frac{2}{3} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{pmatrix} \phi = \begin{pmatrix} 0 & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{2}{3} & \frac{1}{12} \\ 0 & \frac{1}{12} & 0 \end{pmatrix} \rho + O(\Delta x^4).$$

As in §2.4, this algorithm is tested by the same analytic solution

$$\rho = 2\pi^2 \sin \pi x \sin \pi y \quad \text{and} \quad \phi = -\sin \pi x \sin \pi y.$$

In figure 4.1 we see an error decreasing as $0.27\Delta x^4$. Thus at $N = 20$, this fourth-order algorithm gives an error of $2\,10^{-6}$ compared with the earlier second-order algorithm which gave an error of $2\,10^{-3}$ at this resolution.

The same trick of examining the leading error and using the governing equation applied twice can be used to derive the compact Crandall fourth-order algorithm for the diffusion equation $u_t = u_{xx}$:

$$u_i^{n+1} + \left(\frac{1}{12} - \frac{\Delta t}{2\Delta x^2}\right)\left(u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}\right) = u_i^n + \left(\frac{1}{12} + \frac{\Delta t}{2\Delta x^2}\right)\left(u_{i+1}^n - 2u_i^n + u_{i-1}^n\right).$$

4.3 Upwinding

The advection term $\mathbf{u}\cdot\boldsymbol{\nabla}\phi$ propagates information in the direction of $\mathbf{u}$. This is violated by the central difference approximation

$$u_i\frac{\phi_{i+1} - \phi_{i-1}}{2\Delta x},$$

where when $u_i > 0$ downstream information ϕ_{i+1} will influence the time evolution of ϕ_i.

One can correct this erroneous flow of information by evaluating the spatial derivative using only upstream data, i.e. using the earlier one-sided derivatives

$$u\frac{\partial \phi}{\partial x} = \begin{cases} u_i\dfrac{\phi_i - \phi_{i-1}}{\Delta x} & \text{if } u_i > 0, \\ u_i\dfrac{\phi_{i+1} - \phi_i}{\Delta x} & \text{if } u_i < 0, \end{cases}$$

and a similar expression for $v\partial\phi/\partial y$. These one-sided derivatives are only first-order accurate compared with the second-order accurate central differencing. One can of course use extra data points for a second-order one-sided derivative

$$u\frac{\partial \phi}{\partial x} = \begin{cases} u_i\dfrac{3\phi_i - 2\phi_{i-1} + \phi_{i-2}}{2\Delta x} & \text{if } u_i > 0 \\ u_i\dfrac{-\phi_{i+2} + 2\phi_{i+1} - 3\phi_i}{2\Delta x} & \text{if } u_i < 0. \end{cases}$$

this along with the similar expression for $v\partial\phi/\partial y$ has a rather wide numerical molecule. A more compact molecule which is not always strictly upwinding is for $u > 0$ and $v > 0$

$$\frac{u}{\Delta x}\begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ -\frac{1}{2} & 1 & \frac{1}{2} \end{pmatrix}\phi + \frac{v}{\Delta x}\begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 1 & 1 & 0 \\ -\frac{1}{2} & -1 & 0 \end{pmatrix}\phi.$$

4.4 Other grids

The geometry of a problem often dictates the use of non-Cartesian coordinates, such as polars. There are enormous advantages in only using orthogonal coordinate systems,

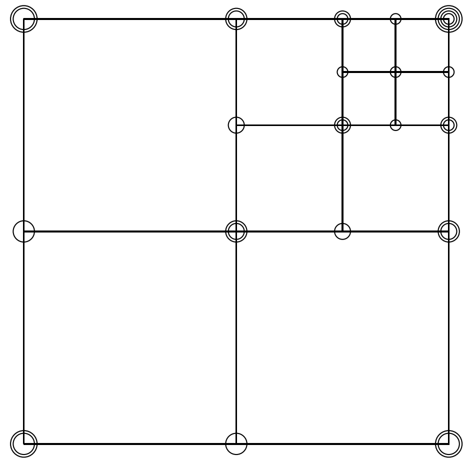

Figure 4.2: A series of grids with increased resolution in a corner.

because differentials such as the Laplacian become rather complicated by the difference between co- and contra-variant tensors.

Increased resolution of a small region with important activity can be obtained by using stretched grids, i.e. mappings $x(\xi)$ and/or $y(\eta)$. Note the independent stretches of the two coordinates, so that the coordinates remain orthogonal. It is much better to use central differencing on an equispaced grid in the mapped coordinates ξ and η with smooth mappings $x(\xi)$ and $y(\eta)$ than to use non-equispaced grids in x and y. The disadvantage of using the two independent stretchings of the two coordinates is that increased resolution will occur wherever either coordinate is stretched while one may well wish to have increased resolution only in a corner where both are stretched.

Methods do exist for using localised increased resolution, e.g. figure 4.2, but there are difficulties in successfully transferring information from the courser grids to the finer and vice versa.

One general problem of stretched grids is that the stability of time-stepping is controlled by the smallest grid block, i.e. to avoid the diffusive numerical instability one requires

$$\Delta t < \tfrac{1}{4} Re\, \Delta x_{\mathrm{min}}^2,$$

and the advection instability

$$\Delta t < (\Delta x/U)_{\mathrm{min}}\,.$$

These restriction become acute for polar coordinates

$$\Delta x_{\mathrm{min}} = r_{\mathrm{min}} \Delta\theta_{\mathrm{min}},$$

and $r_{min} = \Delta r$ if the origin is in the computational domain. When working on the surface of a sphere, these very small spatial grid separations near the two poles can be avoided by patching together six equal spherical-square grids which have all grid blocks of similar size.

In infinite domains, one sometimes needs to bring infinity nearer. This can be achieved with stretches such as

$$x = e^{\xi} \quad \text{or} \quad x = \frac{\xi}{1-\xi}.$$

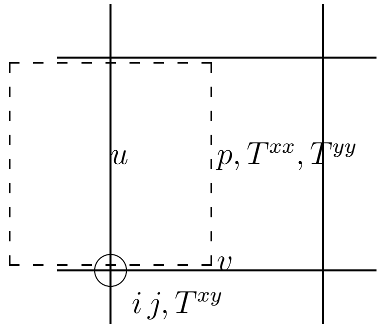

Figure 4.3: The staggered grid for $u_{i\,j+\frac{1}{2}}$, $v_{i+\frac{1}{2}\,j}$, $p_{i+\frac{1}{2}\,j+\frac{1}{2}}$ as before, and stresses $T^{xx}_{i+\frac{1}{2}\,j+\frac{1}{2}}$, $T^{yy}_{i+\frac{1}{2}\,j+\frac{1}{2}}$ and $T^{xy}_{i\,j}$; and in dash the Finite Volume centred on $i\,j+\frac{1}{2}$ for the u-momentum equation.

4.5 Conservative schemes

In §3.7 the algorithm gave slightly different values for the force on the top and the force on the sidewalls and bottom of the driven cavity. It is possible to devise an algorithm which conserves exactly the global momentum, so that the net force on all the walls vanishes. Conservative schemes are based on two ideas, a conservative formulation of governing equations, and the application of those equations to Finite Volumes of Fluid.

The Navier-Stokes equations can be recast into the conservative form

$$\frac{\partial}{\partial t}(\rho \mathbf{u}) + \boldsymbol{\nabla}\cdot\mathbf{T} = 0$$

with total momentum flux

$$\mathbf{T} = \rho\mathbf{u}\mathbf{u} + p\mathbf{I} - 2\mu\mathbf{E}$$

combining Reynolds stresses, isotropic pressure and viscous stresses.

On the staggered grid of §3.6, the diagonal components of the stress are stored with the pressure at $i+\frac{1}{2}\,j+\frac{1}{2}$, and the off-diagonal component at $i\,j$, see figure 4.3. For the x-component of momentum, the conservative form of the momentum equation is applied to the dashed volume marked in figure 4.3, stretching from $x = (i-\frac{1}{2})\Delta x$ to $(i+\frac{1}{2})\Delta x$ and from $y = j\Delta x$ to $(j+1)\Delta x$. The result is

$$\rho u^{n+1}_{i\,j+\frac{1}{2}} = \rho u^{n}_{i\,j+\frac{1}{2}} - \Delta t\left(\frac{T^{xx}_{i+\frac{1}{2}\,j+\frac{1}{2}} - T^{xx}_{i-\frac{1}{2}\,j+\frac{1}{2}}}{\Delta x} + \frac{T^{xy}_{i\,j+1} - T^{xy}_{i,j}}{\Delta x}\right).$$

When adding up over the internal points the changes in the momentum, all the fluxes at internal boundaries between two adjacent finite volumes precisely cancel leaving just fluxes at the outside. The sum of finite volumes for the u-momentum extends from the bottom to the top surface, while it falls one half-volume short at the two side walls. To achieve conservation of momentum for the entire volume it is necessary to set a boundary condition for the pressure on the wall as

$$p = T^{xx}_{\frac{1}{2}\,j+\frac{1}{2}} \quad \text{on} \quad x = 0, \quad y = (j+\tfrac{1}{2})\Delta x,$$

and similarly on the other walls.

The Reynolds stresses have to be evaluated by suitable averages over velocities held at other locations on the staggered grid. The viscous stresses can be evaluated without any averaging.

$$
\begin{aligned}
T^{xx}_{i+\frac{1}{2}\,j+\frac{1}{2}} &= \rho\left(\frac{u_{i+1\,j+\frac{1}{2}}+u_{i\,j+\frac{1}{2}}}{2}\right)^2 + p_{i+\frac{1}{2}\,j+\frac{1}{2}} - 2\mu\frac{u_{i+1\,j+\frac{1}{2}}-u_{i\,j+\frac{1}{2}}}{\Delta x} \\
T^{xy}_{i\,j} &= \rho\left(\frac{u_{i\,j+\frac{1}{2}}+u_{i\,j-\frac{1}{2}}}{2}\right)\left(\frac{v_{i+\frac{1}{2}\,j}+v_{i-\frac{1}{2}\,j}}{2}\right) \\
&\quad -\mu\left(\frac{u_{i\,j+\frac{1}{2}}-u_{i\,j-\frac{1}{2}}}{\Delta x}+\frac{v_{i+\frac{1}{2}\,j}-v_{i-\frac{1}{2}\,j}}{\Delta x}\right) \\
T^{yy}_{i+\frac{1}{2}\,j+\frac{1}{2}} &= \rho\left(\frac{v_{i+\frac{1}{2}\,j+1}+v_{i+\frac{1}{2}\,j}}{2}\right)^2 + p_{i+\frac{1}{2}\,j+\frac{1}{2}} - 2\mu\frac{v_{i+\frac{1}{2}\,j+1}-v_{i+\frac{1}{2}\,j}}{\Delta x}.
\end{aligned}
$$

The idea of a conservative form should be used in curvilinear coordinates, so that the form

$$\nabla^2\phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2}$$

is better numerical to the theoretical equivalent

$$\frac{\partial^2\phi}{\partial r^2} + \frac{2}{r}\frac{\partial\phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\phi}{\partial\theta^2},$$

as can be seen by the discretisation

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi}{\partial r}\right) \approx \frac{\left(r^2\frac{\partial\phi}{\partial r}\right)_{i+\frac{1}{2}} - \left(r^2\frac{\partial\phi}{\partial r}\right)_{i-\frac{1}{2}}}{r_i^2\Delta r} \quad \text{with} \quad \left(r^2\frac{\partial\phi}{\partial r}\right)_{i+\frac{1}{2}} \approx r^2_{i+\frac{1}{2}}\frac{\phi_{i+1}-\phi_i}{\Delta r}.$$

For two-phase flows, e.g. water and air, the so-called 'Volume of Fluid Method' or 'One-Fluid Method' is a conservative scheme treating the two phases as a single fluid with a density ρ and viscosity μ which varies according to whether the part of a volume is one phase or the other.

The advection term in the Navier-Stokes equation has several alternative forms for conserving different properties.

$$
\begin{aligned}
\mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{u} &= \boldsymbol{\nabla}\cdot\mathbf{u}\,\mathbf{u} && \text{conserves momentum} \\
&= \boldsymbol{\nabla}\tfrac{1}{2}u^2 - \mathbf{u}\wedge\boldsymbol{\omega} && \text{rotational form} \\
&= \tfrac{1}{2}\mathbf{u}\cdot\boldsymbol{\nabla}\mathbf{u} + \tfrac{1}{2}\boldsymbol{\nabla}\cdot\mathbf{u}\,\mathbf{u} && \text{conserves energy.}
\end{aligned}
$$

The last form is known as the 'skew-symmetric differential form'. If the scalar product with $\mathbf{u}$ is taken with the last form, one obtains

$$\left(u_i u_j(u_i^{j+1}-u_i^{j-1}) + u_i(u_j^{j+1}u_i^{j+1} - u_j^{j-1}u_i^{j-1})\right)/2\Delta x,$$

where the subscripts refer to components and the superscripts to location. On summing across the grid, the first and fourth terms cancel at internal points and the second and third terms cancel. Hence this form just rearranges energy without changing the total quantity. Conserving energy restricts the solution from blowing up everywhere (could still blow up in a region of decreasing size), whereas fixed finite momentum does not restrict the growth of large cancelling positive and negative momentum.

Chapter 5

Finite Elements

Finite Elements are good for engineering problems with complex geometries. One has 'only' to triangulate the domain and then it is routine to apply the governing equations using the method. Finite Elements were developed first for elliptic equations and are very good for them. They are reasonably good for parabolic equations but usually poor for hyperbolic equations. Finite Elements naturally yield conservative schemes, and are often more accurate than they deserve to be.

The disadvantages are that it is hard to generate the grid, difficult to write programs on unstructured grids, difficult to solve a Poisson problem efficiently on an unstructured grid, and difficult to present the answers on an unstructured grid.

5.1 The two ideas

The Finite Elements approach is based on two ideas. The first is to represent the unknown functions everywhere within each element, i.e. everywhere within the domain. This should be contrasted with the Finite Difference approach of knowing the unknowns only at grid points. The following two subsections are devoted to different possible representations of the unknown functions.

The second idea is a so-called *weak formulation* of the governing equations. This often comes down to a variational statement of the governing problem. This idea is explored in later subsections.

5.2 Representations in one dimension

We divide the one-dimensional domain $a < x < b$ into n segments with $a = x_0 < x_1 < \ldots < x_n = b$ to produce n elements $x_{i-1} < x < x_i$ for $i = 1, 2, \ldots, n$. Note the segments need not have the same length.

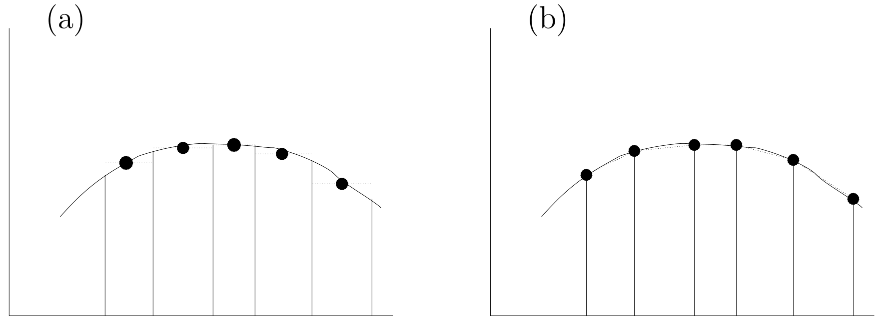

Figure 5.1: Finite Elements in one dimension with (a) constant elements and (b) continuous linear elements.

5.2.1 Constant elements

At the crudest level, we can represent the unknown function $f(x)$ as constant within each element

$$f(x) = f_i \quad \text{in} \quad x_{i-1} \le x < x_i,$$

as in figure 5.1a. Note that the f_i should not be thought of as values of the function at grid point x_i as in Finite Differences but rather as just a parameter associated with the representation of the function over the element.

5.2.2 Linear elements

Better than piecewise constant is a piecewise linear representation of the unknown function $f(x)$

$$f(x) = f_{i-1}\frac{x_i - x}{x_i - x_{i-1}} + f_i\frac{x - x_{i-1}}{x_i - x_{i-1}} \quad \text{in} \quad x_{i-1} \le x < x_i,$$

as in figure 5.1b. Again the f_i are parameters of the representation rather than spot values. If the same value of the parameter is used in the adjacent elements, the resulting representation would be continuous over the whole domain.

5.2.3 Quadratic elements

The next step in sophistication is obviously to represent the unknown function by a quadratic function over each element. Because the expressions now start to become complicated, it is a first worth mapping each element to a unit interval $0 \le \xi \le 1$ with

$$x(\xi) = x_{i-1} + (x_i - x_{i-1})\xi.$$

Then the representation by a quadratic is

$$f(x) = f_{i-1}(1-\xi)(1-2\xi) + f_{i-\frac{1}{2}}4\xi(1-\xi) + f_i\xi(2\xi-1) \quad \text{in} \quad x_{i-1} \le x < x_i$$

While the representation is smooth within each element, there is still a discontinuity in the derivative of the representation across the boundary from one element to the next.

5.2.4 Cubic elements

There are two choices of representation by a cubic over each element. The first

$$f(x) = f_{i-1}(1-3\xi)(1-\frac{3}{2}\xi)(1-\xi) + f_{i-\frac{2}{3}}\theta_2(\xi) + f_{i-\frac{1}{3}}\theta_1(\xi) + f_i\theta_0(\xi),$$

with obvious expressions for the various $\theta(\xi)$. While this cubic can better follow the unknown function within the element, it still has a discontinuity in the derivative of the representation across the boundary from one element to the next.

A second representation

$$\begin{aligned} f(x) &= f_{i-1}(1-\xi)^2(1+2\xi) + f'_{i-1}(1-\xi)^2\xi(x_i - x_{i-1}) \\ &\quad + f_i\xi^2(3-2\xi) + f'_i\xi^2(1-\xi)(x_i - x_{i-1}), \end{aligned}$$

in terms of the four parameters f_{i-1}, f'_{i-1}, f_i and f'_i, will be both continuous and have a continuous derivative across the boundary from one element to another so long as the parameters f_i and f'_i are the same for the two adjacent elements. The second derivative $f''(x)$ of the representation would be discontinuous at the boundary between elements.

5.2.5 Basis functions

In all the cases above, we can write the representation over the *whole domain* as the sum of a products of simple basis functions $\phi(x)$ multiplied by unknown parameters or 'amplitudes' f_i

$$f(x) = \sum f_i\phi_i(x) \quad \text{in} \quad a < x < b.$$

The basis functions are zero through most of the domain, being non-zero in one or two elements only. Note that the parameters now labelled as just f_i include the f'_i in the second cubic representation.

For the constant elements, the basis functions are

$$\phi_i(x) = \begin{cases} 1 & \text{in} \quad x_{i-1} \le x < x_i \\ 0 & \text{otherwise.} \end{cases}$$

For the linear elements, the basis function are

$$\phi_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{x_i - x_{i-1}} & \text{in} \quad x_{i-1} \le x \le x_i \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i} & \text{in} \quad x_i \le x \le x_{i+1} \\ 0 & \text{otherwise,} \end{cases}$$

with obvious modifications for the end elements. See figure 5.2.

For the second cubic elements with continuous derivative across boundaries, there are two types of basis functions, one $\phi_i(x)$ associated with the parameter f_i and the other

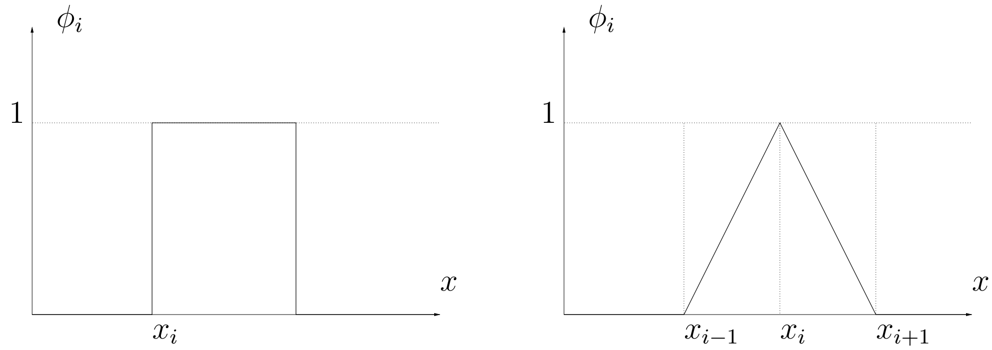

Figure 5.2: Basis functions for constant elements and linear elements.

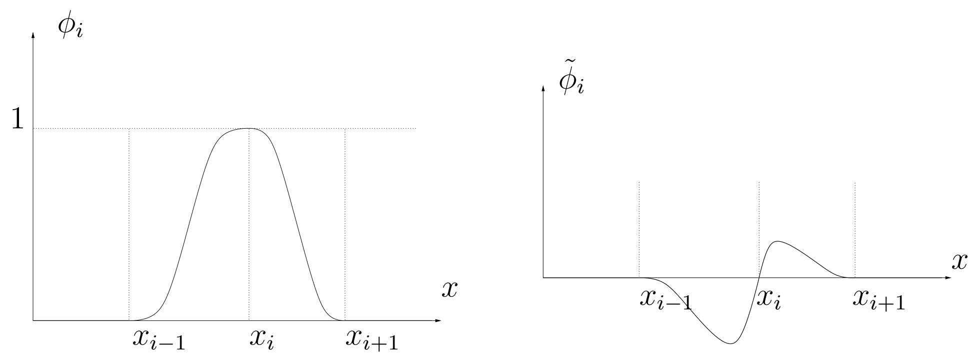

Figure 5.3: Basis for cubic elements with continuous first derivative.

$\tilde{\phi}_i(x)$ associated with f_i'.

$$\phi_i(x) = \begin{cases} (x_{i+1}-x)^2(x_{i+1}+2x-3x_i)/(x_{i+1}-x_i)^3 & \text{in} \quad x_i \le x < x_{i+1} \\ (x-x_{i-1})^2(3x_i-2x-x_{i-1})/(x_i-x_{i-1})^3 & \text{in} \quad x_{i-1} \le x < x_i \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{\phi}_i(x) = \begin{cases} (x-x_i)(x_{i+1}-x)^2/(x_{i+1}-x_i)^2 & \text{in} \quad x_i \le x < x_{i+1} \\ (x-x_i)(x-x_{i-1})^2/(x_i-x_{i-1})^2 & \text{in} \quad x_{i-1} \le x < x_i \\ 0 & \text{otherwise.} \end{cases}$$

See figure 5.3.

5.3 Representations in two dimensions

We start with representations over triangular elements with straight sides.

5.3.1 Constant elements

The simplest representation of an unknown function is by a different constant in each element

$$f(x) = f_i \quad \text{in each triangle } i.$$

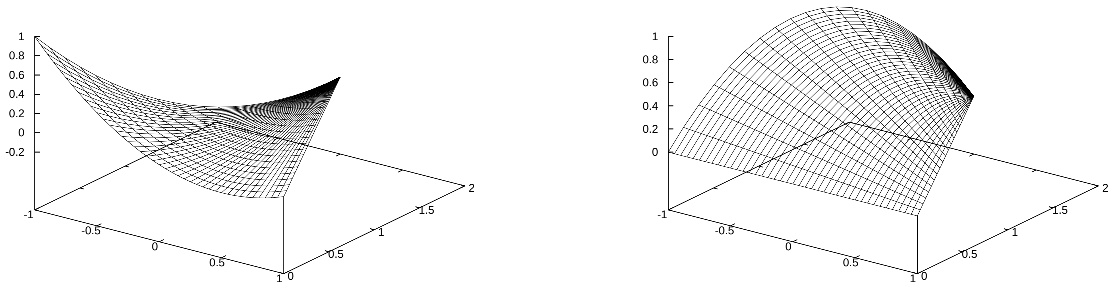

Figure 5.4: The quadratic elements in two dimensions, $\ell_{23}(\mathbf{x})(2\ell_{23}(\mathbf{x}) - 1)$ and $4\ell_{12}(\mathbf{x})\ell_{31}(\mathbf{x})$.

5.3.2 Linear elements

For linear elements, we need the linear interpolation functions $\ell_{12}(\mathbf{x})$, $\ell_{23}(\mathbf{x})$ and $\ell_{31}(\mathbf{x})$, where each function vanishes along two sides and is unity on the other vertex, i.e. $\ell_{12}(\mathbf{x}_1) = 0 = \ell_{12}(\mathbf{x}_2)$ and $\ell_{12}(\mathbf{x}_3) = 1$, where $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$ are the three vertices. Hence

$$\ell_{12}(x, y) = \frac{(x - x_1)(y_2 - y_1) - (x_2 - x_1)(y - y_1)}{(x_3 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_3 - y_1)}.$$

With these three linear interpolation functions, an unknown function can be represented linearly within one element as

$$f(\mathbf{x}) = f_1\ell_{23}(\mathbf{x}) + f_2\ell_{31}(\mathbf{x}) + f_3\ell_{12}(\mathbf{x}).$$

By making the vertex values f_i the same for elements sharing that vertex, the representation becomes continuous over the whole domain.

5.3.3 Quadratic elements

For a quadratic representation of the unknown functions, we can use products of the above linear interpolations functions. Thus

$$\begin{aligned} f(\mathbf{x}) &= f_1\ell_{23}(\mathbf{x})(2\ell_{23}(\mathbf{x}) - 1) + f_2\ell_{31}(\mathbf{x})(2\ell_{31}(\mathbf{x}) - 1) + f_3\ell_{12}(\mathbf{x})(2\ell_{12}(\mathbf{x}) - 1) \\ &\quad + f_{23}4\ell_{12}(\mathbf{x})\ell_{31}(\mathbf{x}) + f_{31}4\ell_{23}(\mathbf{x})\ell_{12}(\mathbf{x}) + f_{12}4\ell_{31}(\mathbf{x})\ell_{23}(\mathbf{x}). \end{aligned}$$

This representation takes the value f_1 at the vertex $\mathbf{x}_1$ and the value f_{23} at the midpoint along the edge from $\mathbf{x}_2$ to $\mathbf{x}_3$, see figure 5.4. By making the vertex values and midpoint values the same for vertices and midpoints shared by two elements, the representation becomes continuous and a continuous tangential derivative along the edges. The normal derivative to an edge is however discontinuous, and the derivative is quite unpleasant at a vertex.

5.3.4 Cubic elements

A general cubic in two dimensions has 10 degrees of freedom. With this flexibility, it is possible to construct a cubic representation which takes given values and given (two-dimensional) derivatives at each vertex of a triangle, leaving one degree of freedom unused. This latter could be taken as something like the value at the centre of the triangle (and vanishing with vanishing derivatives at all the vertices), which is called a 'bubble' function.

5.3.5 Basis functions

Adopting one of the above representations, or a further generalisation, over each element, we can write the representation over the whole domain in terms of amplitudes f_i and basis functions $\phi_i(\mathbf{x})$

$$f(\mathbf{x}) = \sum f_i \phi_i(\mathbf{x}).$$

Note again that the amplitudes will normally not be a value of the unknown function at any special point. Also the basis functions will be non-zero only in a small number of elements. For example for linear elements, the basis function associated with one vertex will be non-zero only in the triangles which share that particular vertex. As the linear basis function will vanish along the edges opposite the particular vertex, it will take the form of a several-sided pyramid. The local nature of the basis functions helps create sparse coupling matrices when the finite elements are applied to nonlinear partial differential equations.

5.3.6 Rectangles

For complex geometries, the domain is most easily divided into triangular elements. In simpler geometries, rectangular elements are sometimes used.

The simplest representation of an unknown functions is by different constants in each element.

It is not possible to make a linear interpolation between the values at the four corners of a rectangle. Instead a so-called *bilinear* form is used which is separately linear in the two coordinates. Suppose the xy-rectangle has been mapped to the unit square $0 \leq \xi \leq 1$, $0 \leq \eta \leq 1$. Then the bilinear representation would be

$$f(\mathbf{x}) = f_1 \xi \eta + f_2 (1-\xi)\eta + f_3 \xi (1-\eta) + f_4 (1-\xi)(1-\eta).$$

Note the only quadratic term is the product $\xi\eta$. The parameters f_i are the values of the representation in one of the four corners. By using the same values for all elements sharing a corner, one constructs a representation which is continuous over the whole domain.

One can similarly construct a bi-quadratic representation which is the sum of 9 terms, each the product of separate quadratics in the two coordinates. One possibility is to make each of the terms vanish at all but one of the corners or midpoints. Again using the same values of the parameters for elements sharing the same corners of midpoints, one constructs a representation which is continuous over the whole domain, and which has a continuous tangential derivative along each edge, although the normal derivatives are not continuous and the derivative is not nice at a corner.

The one- and two-dimensional representations can be generalised in an obvious way to three dimensions. One can also used curved elements by first mapping them onto straight-sided elements.

5.4 Variational statement of the Poisson problem

The Poisson problem arises in many branches of physics and is an illustration of more general self-adjoint elliptic problems.

$$\nabla^2 f = \rho \quad \text{in volume } V$$

$$\text{with boundary condition, say} \quad f = g \quad \text{on surface } S,$$

with $\rho(\mathbf{x})$ and $g(\mathbf{x})$ given. For this problem, we have a Rayleigh-Ritz variational formulation: out of all those functions $f(\mathbf{x})$ which satisfy the boundary conditions, the one which minimises

$$I(f) = \int_V \left(\frac{1}{2} |\nabla f|^2 + \rho f \right) dV$$

also satisfies the Poisson problem.

Into this variational statement we substitute our finite element representation

$$f(\mathbf{x}) = \sum f_i \phi_i(\mathbf{x})$$

Then

$$I(f) = \frac{1}{2} \sum_{i\,j} f_i f_j \int \nabla\phi_i \cdot \nabla\phi_j + \sum_i f_i \int \rho\phi_i.$$

We define the *global stiffness matrix*

$$K_{ij} = \int \nabla\phi_i \cdot \nabla\phi_j,$$

and *forcing*

$$r_i = \int \rho\phi_i.$$

Then minimising $I(f)$ over the amplitudes f_i, we obtain an equation for the unknown amplitudes

$$K_{ij} f_j + r_i = 0.$$

Note with the f_i determined by this equation, the representation $f(\mathbf{x}) = \sum f_i \phi_i(\mathbf{x})$ satisfies

$$-\int \nabla f \cdot \nabla\phi_j = \int \rho\phi_j \quad \text{for all } j,$$

i.e. the governing equation is satisfied in each of the finite directions accessible to the finite element representation. This is called a *week formulation* of the partial differential equation. It can be used, and indeed will shortly be used, with representations $f(\mathbf{x})$ which are continuous and have a derivative ∇f, so that the weak formulation can be considered, but which have no nice $\nabla^2 f$ (along the edges), so that the original partial differential equation has no simple meaning.

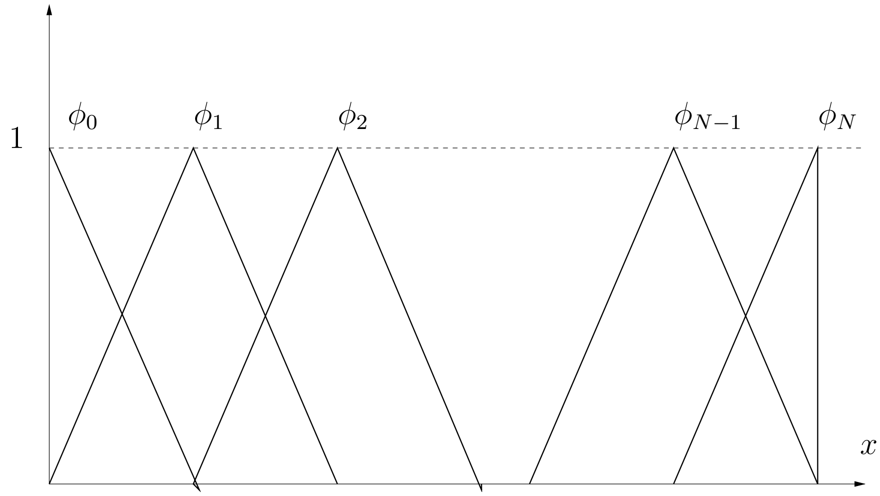

Figure 5.5: Linear elements in one dimension.

5.5 Details in one dimension

To see the Finite Element method working in practice, consider the one-dimensional Poisson problem for the unknown $f(x)$ in

$$\frac{d^2 f}{dx^2} = \rho \quad \text{in } a < x < b.$$

$$\text{with } f(a) = A \text{ and } f(b) = B,$$

where $\rho(x), A$ and B are given.

We divide the interval $[a, b]$ into N equal segments of length $h = (b - a)/N$. We use linear finite elements with basis functions as in figure 5.5. Hence the unknown $f(x)$ is represented

$$f(x) = A\phi_0(x) + B\phi_N(x) + \sum_{i=1}^{N-1} f_i \phi_i(x)$$

in which the boundary conditions have been built in. At interior points i,j=1,2,. . . ,N-1, the global stiffness matrix takes the values.

$$K_{ij} = \int \nabla\phi_i \cdot \nabla\phi_j = \begin{cases} 2/h & \text{if } i = j, \\ -1/h & \text{if } i = j \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

These results follow from the gradient being either zero or $\pm 1/h$, with agreement over two segments of length h when $i = j$, with opposite signs over the one common segment when $i = j \pm 1$ and with no overlap of the nonzero segments otherwise. If we take the given $\rho(x)$ to be piecewise constant, then the forcing r_i are given by

$$r_i = \int \rho(x)\phi_i = h\rho_i.$$

The equation governing the unknown amplitudes f_i then becomes

$$\frac{1}{h}\left(-f_{i-1}+2f_i-f_{i+1}\right)+h\rho_i=0 \quad \text{for } i=1,2,\ldots,N-1,$$

which is the same equation for the point values in the finite difference approach.

Note if the interval had been divided unequally into segments of length $h_{\frac{1}{2}}, h_{\frac{3}{2}}, \ldots, h_{N-\frac{1}{2}}$, then the Finite Element equation for the unknown amplitudes f_i would have been

$$\frac{1}{h_{i-\frac{1}{2}}}\left(-f_{i-1}+f_i\right)+\frac{1}{h_{i+\frac{1}{2}}}\left(f_i-f_{i+1}\right)+\frac{h_{i-\frac{1}{2}}+h_{I+\frac{1}{2}}}{2}\rho_i=0.$$

This form shows the finite element approach naturally produces a conservative scheme.

Note if the forcing integrals r_i had been evaluated more accurately, we would have obtained

$$r_i=\int\rho(x)\phi_i(x)=\rho_i+\frac{h^3}{12}\rho_i''+O(h^5).$$

The governing equation for these finite elements would then give an error of $O(h^4)$ in the representation on the boundaries of the elements, although the error at interior points would be $O(h^2)$.

5.6 Details in two dimensions

A more typical example of the Finite Element method working is the solution of the Poisson problem in two dimensions. We use linear basis functions on triangular elements, see figure 5.6. Consider the basis function $\phi_1(\mathbf{x})$ which vanishes on the 23-side and is unity at the 1-vertex. The magnitude of its gradient is one upon the altitude h_1 from the 23-side to the 1-vertex, i.e.

$$|\nabla\phi_1|=\frac{1}{h_1}.$$

Hence the contribution of this triangle to the diagonal element of the global stiffness matrix is

$$K_{11}=\int\nabla\phi_1\cdot\nabla\phi_1=\frac{A}{h_1^2},$$

where A is the area of the triangle. Now the angle between $\nabla\phi_1$ and $\nabla\phi_2$ is $\pi-\theta_3$, where θ_3 is the included angle at the 3-vertex. Hence the off-diagonal element of the stiffness matrix has a contribution from this triangle of

$$K_{12}=\int\nabla\phi_1\cdot\nabla\phi_2=-\frac{\cos\theta_3 A}{h_1h_2}.$$

Now by elementary geometry, the altitudes can be related to the lengths of the sides by

$$h_1=\ell_2\sin\theta_3 \quad\text{and}\quad h_2=\ell_1\sin\theta_3.$$

While the area can be expressed

$$A=\tfrac{1}{2}\ell_1\ell_2\sin\theta_3.$$

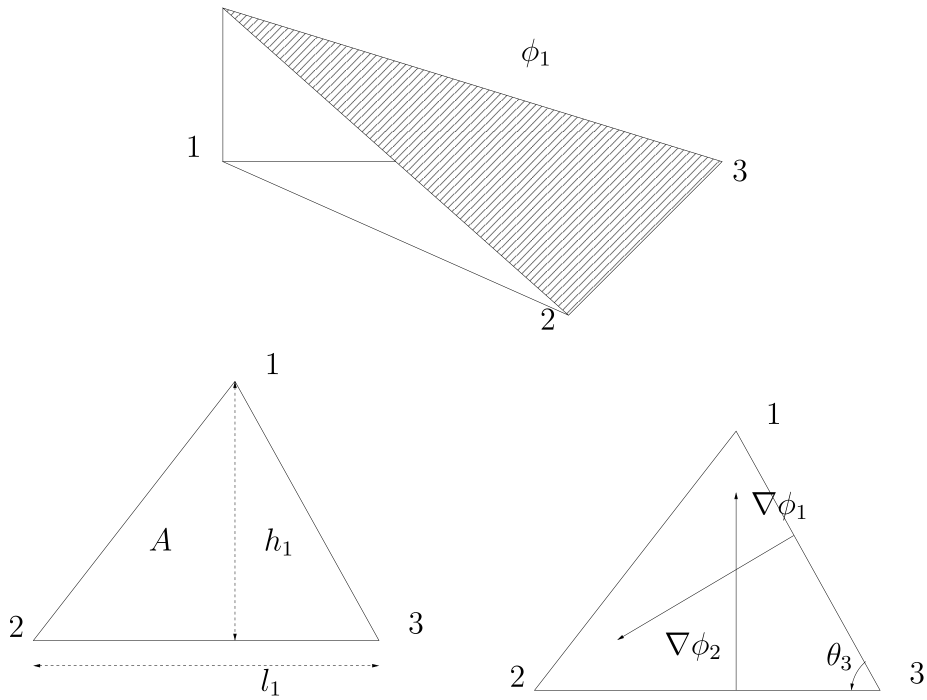

Figure 5.6: Linear element in two dimensions.

Hence the off-diagonal contribution simplifies to

$$K_{12} = \frac{\cos\theta_3 A}{h_1 h_2} = \tfrac{1}{2}\cot\theta_3.$$

Now the length of one side can be split into two parts

$$\ell_1 = h_1 \cot\theta_3 + h_1 \cot\theta_2.$$

Hence the diagonal contribution simplifies to

$$K_{11} = \frac{A}{h_1^2} = \tfrac{1}{2}\left(\cot\theta_3 + \cot\theta_2\right).$$

Note that the contributions from the triangle to the stiffness matrix sum to zero,

$$K_{11} + K_{12} + K_{13} = 0.$$

This follows from the sum of the basis functions is a linear function which is unity at each vertex, and so

$$\phi_1(\mathbf{x}) + \phi_2(\mathbf{x}) + \phi_3(\mathbf{x}) \equiv 1.$$

Being constant, the sum has zero gradient, i.e.

$$\nabla\phi_1 \cdot (\nabla\phi_1 + \nabla\phi_2 + \nabla\phi_3) \equiv 0.$$

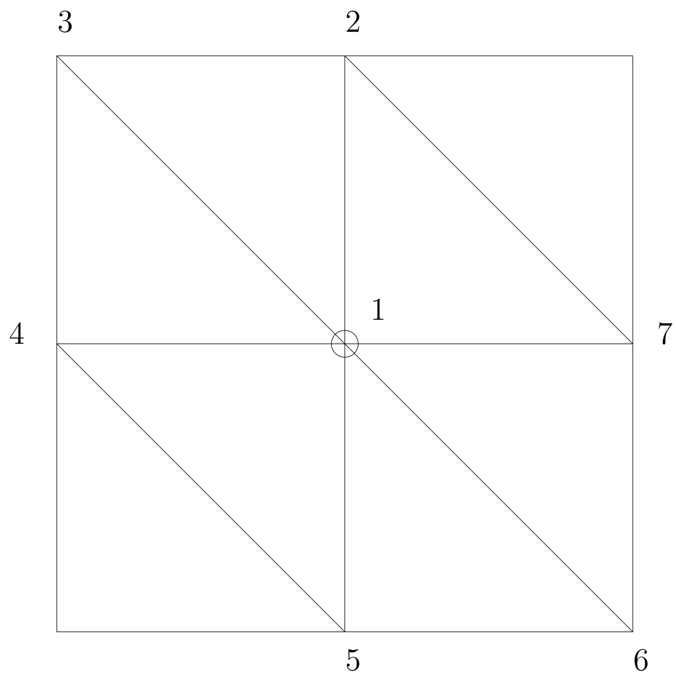

Figure 5.7: Linear element in two dimensions.

We now assemble the contributions from the different triangles that meet at one vertex. For simplicity, we consider a square Cartesian grid with the squares divided by a diagonal to form triangular elements, see figure 5.7.

For the 123-triangle, the contribution to K_{13} involves the other (not 1 or 3) angle, $\theta_2 = \frac{\pi}{2}$, whose cotangent vanishes, so the contribution to $K_{13} = 0$. Similarly the contribution to K_{12} involves $\theta_3 = \frac{\pi}{4}$, whose cotangent is unity, so $K_{12} = -\frac{1}{2}$. The contribution to the diagonal element is the negative sum of the off-diagonals, so $K_{11} = \frac{1}{2}$. Turning to the 172-triangle, we have contributions to the two off-diagonal elements equal, $K_{17} = K_{12} = -\frac{1}{2}\cot\frac{\pi}{4} = -\frac{1}{2}$. And by the diagonal element being the negative sum of the off-diagonal, the contribution to $K_{11} = 1$. Now all the triangles which meet at vertex 1 are similar to one or other of these triangles, so we can add up the contributions to find

$$K_{11} = 4, \quad K_{12} = K_{14} = K_{15} = K_{17} = -1, \quad K_{12} = K_{16} = 0.$$

We must now evaluate the forcing terms from ρ,

$$r_i = \int \rho\phi_i.$$

At the lowest approximation we take $\rho(\mathbf{x})$ to be constant ρ_i, then integrating the linearly varying $\phi(\mathbf{x})$ over just one triangle, we have a contribution from the triangle of $\frac{1}{3}A\rho_i$. Adding up the contributions from the 6 triangles at vertex 1, and using $A = \frac{1}{2}h^2$ where h is the size of the squares, we have

$$r_1 = h^2\rho_1.$$

Hence the Poisson problem in Finite Elements

$$K_{ij}f_j + r_i = 0$$

becomes on our special triangular grid

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} f + h^2\rho_i = 0,$$

which is identical to the Finite Difference equation when the amplitudes f_i are the grid values.

On more general, unstructured grids there is a programing challenge of maintaining useful lists. One needs a list of points P, which have coordinates (x_P, y_P) and an indicator of whether the point is interior or on the boundary. One needs a list of triangles T, which have vertices P_{1T}, P_{2T}, P_{3T}. This list can be searched to find in which triangle an arbitrary point lies. One needs the inverse list, i.e. for each vertex P which M triangles $T_{1P}, T_{2P}, \ldots, T_{MP}$ contain the vertex. An alternative to this list is a list of edges E joining points P_{1E} and P_{2E}. From either of these alternative lists, one can assemble the sparse stiffness matrix. There are other possible lists, whose usefulness depends on the particular problem being tackled, where the overheads of setting up and maintaining the lists has to be balanced by savings in CPU times and memory consumed.

5.7 Galerkin formulation

Many branches of physics have a variational formulation, in which the governing partial differential equations are the Euler-Lagrange equations of the minimisation of some expression for an Action. But some physics does not have a variational statement, e.g. the Navier-Stokes equations. In such cases , we use a so-called *weak formulation* of the problem, sometimes also called a *Galerkin formulation*.

Consider the governing equation written in the symbolic form

$$A(u) = f,$$

where A is a nonlinear partial differential equation for the unknown $u(\mathbf{x}, t)$. We use a Finite Element representation

$$u(\mathbf{x}, t) = \sum_i^N u_i(t)\phi_i(\mathbf{x}),$$

with a finite number N of amplitudes $u_i(t)$ and localised basis functions $\phi_i(\mathbf{x})$. Thus we can view our approximation to the exact solution to be a member of a finite vector space spanned by the basis functions. On this vector space, we define an inner product

$$\langle a, b\rangle = \int a(\mathbf{x})b(\mathbf{x})\, dV.$$

The Galerkin approximation is to require the residual in satisfying the equation, $A(u) - f$, to have no component in our solution space, i.e. to be orthogonal to all the basis functions

$$\langle A(u) - f, \phi_j\rangle = 0 \quad \text{all } j.$$

If the operator A contains derivatives of order higher than the first, then one can integrate by parts to transfer half the derivatives on u to operate instead on the ϕ_j. For example in §5.4, the second order derivative in the original Poisson equation for f was reduced to first order only on the ϕ in the global stiffness matrix K. Indeed the basis functions used in the two previous sections were piecewise linear, so their first derivative was square-integrable in K while their second derivatives existed only in the form of delta functions along the edges. This reduction in the differentiability requirements of the basis functions is the origin of the word *weak*: the Finite Element representation is required to satisfy *weaker* conditions than those necessary to satisfy the original governing equation. For those interested in the rigorous underpinning of the subject, one would need to prove (i) that a solution to the weak formulation of the Finite Element problem exists for any N, (ii) that the solutions converges to a limit as $N \to \infty$, and (iii) that the limit solution is nice and smooth, and finally (iv) that the limit solution satisfies the original governing equation.

Nearly always the basis functions can represent a constant function. With an appropriate normalisation, we would have

$$\sum \phi_j(\mathbf{x}) \equiv 1,$$

perhaps not including all the basis functions. Now the weak formulation is

$$\langle A(u) - f, \phi_j \rangle = 0 \quad \text{all } j.$$

Summing over the j, we therefore have

$$\langle A(u) - f, 1 \rangle = 0,$$

i.e.

$$\int A(u) = \int f.$$

Thus the Finite Element approach automatically satisfies a global conservations property.

5.8 Diffusion equation

5.8.1 Weak formulation

We consider the application of the weak formulation to the diffusion equation as a a step towards the more complicated Navier-Stokes equation

$$u_t = \nabla^2 u.$$

The weak formulation is

$$\langle u_t - \nabla^2 u, \phi_j \rangle = 0 \quad \text{all } j.$$

Integrating by parts,

$$\langle u_t, \phi_j \rangle = -\langle \nabla u, \nabla \phi_j \rangle.$$

Substituting in the Finite Element representation,

$$u(\mathbf{x}, t) = \sum_i u_i(t)\phi_i(\mathbf{x}),$$

we obtain

$$\sum_i \dot{u}_i(t)\langle \phi_i, \phi_j \rangle = -\sum_i u_i(t)\langle \nabla\phi_i, \nabla\phi_j \rangle,$$

i.e.

$$M_{ij}\dot{u}_j = -K_{ij}u_j,$$

with new

$$\text{'Mass' matrix} \quad M_{ij} = \langle \phi_i, \phi_j \rangle,$$

and the earlier

$$\text{'Stiffness' matrix} \quad K_{ij} = \langle \nabla\phi_i, \nabla\phi_j \rangle.$$

5.8.2 In one dimension

Using the linear elements on equal intervals h, we have

$$M_{ij} = \begin{cases} \frac{2}{3}h & i = j \\ \frac{1}{6}h & i = j \pm 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad K_{ij} = \begin{cases} 2/h & i = j \\ -1/h & i = j \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

Hence at an interior point we obtain the equation governing the evolution of the amplitudes

$$h\left(\tfrac{1}{2}\dot{u}_{i-1} + \tfrac{2}{3}\dot{u}_i + \tfrac{1}{2}\dot{u}_{i+1}\right) = \frac{1}{h}\left(u_{i-1} - 2u_i + u_{i+1}\right).$$

Note this leaves a linear algebra problem to find the $\dot{u}_i$. The tridiagonal matrix M can however be rapidly inverted.

The equation above is called a 'semi-discretised' form, a form in which the spatial variation has been discretised but which leaves open the method of time-integration. One could us a simple forward time stepping

$$u_i^{n+1} = u_i^n + \Delta t \dot{u}_i^n,$$

or one of a number of more refined schemes. However it is certainly very unwise to decretise the time stepping using Finite Elements in time.

5.8.3 In two dimensions

We again use the linear basis functions on triangular elements, see figure 5.6. The contributions from a triangle to the elements of the mass matrix are

$$M_{ij} = \begin{cases} \frac{1}{12}h^2 & i = j \\ \frac{1}{24}h^2 & i \neq j, \end{cases}$$

and the same stiffness matrix as in §5.6. Hence assembling the contributions from the different triangles in figure 5.7, we obtain

$$\tfrac{1}{2}h^2\left(\dot{u}_1 + \tfrac{1}{6}(\dot{u}_2 + \dot{u}_3 + \dot{u}_4 + \dot{u}_5 + \dot{u}_6 + \dot{u}_7)\right) = u_2 + u_4 + u_5 + u_7 - 4u_1.$$

Again there is a linear algebra problem to find $\dot{u}_i$.

5.9 Navier-Stokes equation

5.9.1 Weak formulation

We use a Finite Element representation for the velocity $\mathbf{u}$ and pressure p,

$$\begin{aligned}\mathbf{u}(\mathbf{x},t) &= \sum_i \mathbf{u}_i(t)\phi_i(\mathbf{x}),\\ p(\mathbf{x},t) &= \sum_i p_i(t)\psi_i(\mathbf{x}),\end{aligned}$$

with unknown amplitudes $\mathbf{u}_i$ and p and basis functions ϕ_i and ψ_i. Note that one normally uses different basis functions for the velocity and pressure, for example the pressure with one less derivative than the velocity. The weak formulation of the Navier-Stokes equation is then

$$\left\langle \rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}\right) + \nabla p - \mu\nabla^2\mathbf{u}, \phi_j \right\rangle = 0 \quad \text{all } \phi_j,$$

and the incompressibility constraint

$$\langle \nabla\cdot\mathbf{u}, \psi_j\rangle \quad \text{all } \psi_j.$$

Integrating by parts the viscous term, we obtain the system of equations governing the amplitudes

$$\rho\left(M_{ij}\dot{\mathbf{u}}_j + Q_{ijk}\mathbf{u}_j\mathbf{u}_k\right) = -B_{ji}p_j - \mu K_{ij}\mathbf{u}_j,$$

and

$$-B_{ij}\mathbf{u}_j = 0,$$

with the earlier mass M and stiffness K matrices and two new coupling matrices

$$Q_{ijk} = \langle \phi_i\nabla\phi_j, \phi_k\rangle \quad \text{and} \quad B_{ij} = \langle \nabla\psi_i, \phi_j\rangle = -\langle \psi_i, \nabla\phi_j\rangle.$$

5.9.2 Time integration

The time marching of the above semi-discretised form can be made using one's favourite Finite Difference (in time) method. e.g. the simplest explicit forward step

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n + \Delta t\dot{\mathbf{u}}_i^n.$$

The incompressibility can be taken into account by a projection split step method, as in §**??**, i.e.

$$\begin{aligned}\mathbf{u}^* &= \mathbf{u}_i^n + \Delta t\left(\dot{\mathbf{u}}_i^n \quad \text{without the } p \text{ term}\right),\\ \mathbf{u}^{n+1} &= \mathbf{u}^* + \Delta t\left(\dot{\mathbf{u}}_i^n \quad \text{with just the } p \text{ term}\right),\end{aligned}$$

with p chosen so the incompressibility is satisfied at the end of the step

$$Bu^{n+1} = 0.$$

A pressure update method is better than this pressure projection method, and there are second order $O(\Delta t^2)$ alternatives.

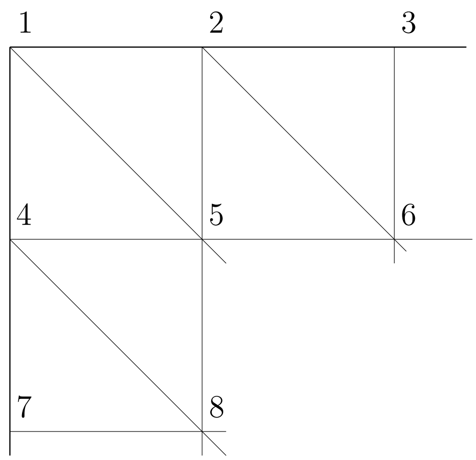

Figure 5.8: Pressure locking

5.9.3 Pressure problem – locking

Consider using triangular elements with velocity linear and pressure constant over each element. The pressure basis functions $\psi_j(\mathbf{x})$ will be unity in triangle Δ_j and zero in all the other triangles. The weak formulation of the incompressibility constraint,

$$\langle \nabla \cdot \mathbf{u}, \psi_j \rangle = 0 \quad \text{all } j,$$

then gives

$$\oint_{\Delta_j} u_n = 0,$$

i.e. no net volume flux out of triangle Δ_j.

We now examine the consequences of the condition of no net volume flux out of each triangle on the grid in figure 5.8, of squares of side h with diagonals added to form triangular elements. The no-slip boundary condition of both components of velocity vanishing is applied to the boundary vertices 1,2,3,4 and 7. Let (u_5, v_5) be the unknown amplitude of the velocity at vertex 5. Consider triangle 145. With the velocity varying linearly along the edge 45, the flux in across edge 45 is $\frac{1}{2}hv_5$. The flux out across edge 15 is however $\frac{1}{2}h(u_5 + v_5)$. Hence the zero net mass flux constraint gives $u_5 = 0$. Now consider the triangle 125. The net flux into this triangle is $\frac{1}{2}hv_5$, whose vanishing gives $v_5 = 0$. Now that both components of velocity vanish at vertex 5, we can use the same argument to deduce that both components must vanish at vertices 6 and 8. We can continue this process right across the grid, deducing that the velocity must vanish everywhere. This deduction follows from the constraint of no net mass flux into each triangle and the no-slip boundary condition. The momentum equation has yet to be considered.

This paradoxical behaviour becomes easier to understand if we look at the whole domain rather than one triangle. For one triangle we seem to have 1 unknown pressure amplitude and 6 unknown velocity component amplitudes (2 components at 3 vertices), i.e. many more degree of freedom in the velocity that the pressure. Now consider the small grid in figure 5.9 of 4 squares by 3 squares. There are 24 triangles, each with a pressure

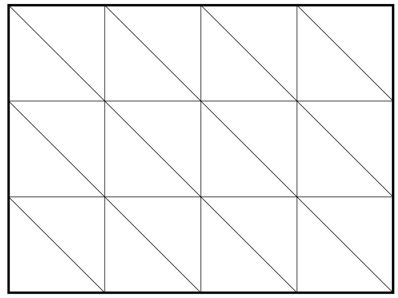

Figure 5.9: Locking due to many triangles and few interior points.

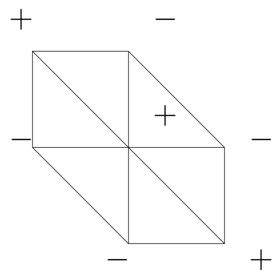

Figure 5.10: A spurious pressure mode

amplitude. There are however only 6 interior vertices with therefore only 12 unknown velocity component amplitudes. The problem thus comes from each vertex being shared by 6 triangles. To redress the imbalance, one can introduce so-called 'bubble' functions for both velocity components. These basis functions vanish at every vertex and are unity at the centre of the triangle, and are hence cubic in nature. This introduces 2 extra unknown velocity component amplitudes to each triangle, which means there are now more degrees of freedom for the momentum equation than the incompressibility constraint. Typically of practice is to use triangular elements with either linear pressure with linear velocity plus a cubic velocity bubble or slightly better linear pressure with quadratic velocity plus a cubic velocity bubble.

5.9.4 Pressure problem – spurious modes

As in §**??** on Finite Differences, Finite Elements can have spurious pressure modes which do not contribute to the momentum equation. In fact, they are nearly unavoidable in Finite Elements which does not have an equivalent to Finite Difference's staggered grid. Consider pressure linear over the triangles in figure 5.9. If the pressure is ± 1 with alternating signs at adjacent vertices along the horizontal and vertical, then around one vertex there is a pressure distribution is as in figure 5.10. By symmetry this distribution of pressure has no pressure gradient at the vertex, and hence does not contribute to the momentum equation.

The problem spurious pressure modes in Finite Elements reduces to the coupling matrix B having eigensolutions, i.e. non-trivial pressures p_j such that

$$B_{ji} p_j = 0.$$

One would like to find basis functions for the velocity and pressure which produced a coupling matrix B which had no eigensolutions. This requirement is also called the Babuška-Brezzi condition.

If a set of well behaved basis functions cannot be used, an alternative method of suppressing the spurious pressure modes is to modify the physics, replacing the incompressibility equation by

$$\nabla \cdot \mathbf{u} = \beta h^2 \nabla^2 p,$$

where h is the size of the elements and β is a constant, whose optimal value is thought to be $\beta = 0.025$. The idea behind this trick is that for physically sensible long scale pressure variations the corrections is $O(h^2)$ and decreasing as the spatial resolution is refined. On the other hand for the non-physical spurious pressure modes which vary over a distance h the Laplacian of pressure is large and acts to reduce the mode. The weak formulation of the modified incompressibility condition is

$$B_{ij} \mathbf{u}_j + \beta h^2 K'_{ij} p_j = 0,$$

with stiffness for the pressure basis functions

$$K'_{ij} = \langle \nabla \psi_i, \nabla \psi_j \rangle.$$

This equation has the desirable feature of bringing in the pressure whose role is to help achieve incompressibility but which does not normally occur in the constraint.