CHAPTER 11

Finite element methods

1. Preliminaries

For a number of applications the restrictions imposed by finite difference or spectral methods with respect to the computational grid are too severe. This is especially the case in structural engineering where the elasticity equations are solved in domains of complicated geometry such as the interior of an automobile engine. A review of the finite difference and spectral methods would show that the reason relatively simple grids are required is that the differential form of the equation is used. Finite volume methods had no such restriction since they used an integral formulation, and indeed complicated geometries may be treated by finite volume methods. Another class of methods which are based upon an integral formulation are the finite element and closely related boundary element methods. We shall concentrate on finite element methods for now.

The basic idea behind the finite element methods is to employ a piecewise local approximation $\tilde{q}$ of the unknown function q that satisfies some PDE of interest. The piecewise local approximation is defined over some general discretization of the domain of definition of q denoted by Ω. Instead of directly using the piecewise local approximation in the PDE we employ a weighted residual formulation. There arises the significant question of how to best relate the integral formulation to the PDE of interest. Once the discretization, piecewise local approximation and integral formulation are determined a system of equations is obtained whose solution gives the complete approximation to the problem of interest. We shall look at each of these components in detail.

1.1. Spatial discretizations. A domain Ω may be discretized into simple elements in very many ways. Nonetheless only a few are typically used in practice. General affine geometry furnishes some guidance for general discretization techniques. We know for instance that any $d-$dimensional domain may be expressed as a reunion of simplicia

$$\Omega = \cup_k S_k \ . \tag{1.1}$$

Simplicia are the simplest continuum geometric entitites one can construct in a space of dimension d. For 1D spaces the simplicia are line segments. In 2D they are triangles and in 3D they are tetrahedra. The measure of each of these elements is easily determined by the formulas:

(1) line segment in 1D of nodes $\{x_1, x_2\}$

$$l = x_2 - x_1 = \begin{vmatrix} 1 & 1 \\ x_1 & x_2 \end{vmatrix} \tag{1.2}$$

FIGURE 1. Example of the discretization into triangles of the domain between a circle and a NACA-0012 airfoil.

(2) triangle in 2D with nodes $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$

$$A = \frac{1}{2}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \tag{1.3}$$

(3) tetrahedron in 3D with nodes $\{(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)\}$

$$V = \frac{1}{6}\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_3 & z_3 & z_4 \end{vmatrix} \tag{1.4}$$

In the above formulas the element measures are given with sign, the sign corresponding to orientation of the nodes. We understand that the positive value is to be taken whenever a true geometric measure (length, area, volume) is required. Simplicia have many attractive theoretical properties, in particular there exists a definition of what an optimal discretization is for a number of PDE's of interest, especially elliptic PDE's such as the Poisson equation. Fig. 1 shows an example of such a discretization.

Another widely used discretization is into generalized polyhedra having $2d$ sides, i.e. line segments in 1D, quadrilaterals in 2D, hexahedra in 3D. These have the advantage of enabling easier organization of programs since there is a natural ordering of the indices identifying each element. Thus discretizations which use these types of elements give rise to *structured computational grids*, similar to those encountered in finite difference methods whereas discretizations using simplicia lead to *unstructured computational grids*.

1.2. Piecewise interpolations. Once a discretization scheme for the geometric domain has been established the next step is to define a local approximation of q over the element E. Typically the approximation is an interpolation based upon values Q_j defined somewhere within the element E, but this is not obligatory and other approximations (spectral elements, Chebyshev elements) may be used. The position where the values Q_j are to be defined must be established. A simple choice is the element nodes but again this is not obligatory and the values may be positioned at other points within E. Finally an interpolation scheme must be established such as polynomial interpolation. Let us give some typical examples:

1.2.1. *Linear elements in 1D.* The element E has two nodes $\{x_1, x_2\}$, $x_2 > x_1$. Values representing $q(x)$ are defined at the nodes $\{Q_1, Q_2\}$. These define a linear polynomial approximation valid over E

$$\tilde{q}(x) = \frac{(x - x_1)Q_2 + (x_2 - x)Q_1}{x_2 - x_1} = N_1(x)Q_1 + N_2(x)Q_2 \tag{1.5}$$

The functions $N_1(x)$, $N_2(x)$ have properties reminiscent of the Dirac delta

$$N_1(x_1) = 1,\ N_1(x_2) = 0 \tag{1.6}$$

$$N_2(x_1) = 0,\ N_2(x_2) = 1 \tag{1.7}$$

FIGURE 2. Linear 1D form functions.

FIGURE 3. Quadratic element form functions in 1D.

and are called *form functions*. The particular ones used here are called the 1D linear form functions and are depicted in Fig. (2)

1.2.2. *Quadratic elements in 1D.* The element E has three nodes $\{x_1, x_2, x_2\}$ and the local approximation is

$$\tilde{q}(x) = N_1(x)Q_1 + N_2(x)Q_2 + N_3(x)Q_3 \tag{1.8}$$

with the form functions

$$N_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} \tag{1.9}$$

$$N_2(x) = \frac{(x-x_3)(x-x_1)}{(x_2-x_3)(x_2-x_1)} \tag{1.10}$$

$$N_3(x) = \frac{(x-x_1)(x-x_3)}{(x_3-x_1)(x_3-x_2)} \tag{1.11}$$

1.2.3. *Linear elements on triangles in 2D.* The element E has 3 nodes of coordinates $\{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$ at which the values Q_1, Q_2, Q_3 are defined. The local approximation of q is given by

$$\tilde{q}(x, y) = N_1(x, y)Q_1 + N_2(x, y)Q_2 + N_3(x, y)Q_3 \tag{1.12}$$

with the form functions

$$N_1(x,y) = \frac{1}{2A}\begin{vmatrix} 1 & 1 & 1 \\ x & x_2 & x_3 \\ y & y_2 & y_3 \end{vmatrix} = \frac{1}{2A}\left(xy_2 - yx_2 - xy_3 + yx_3 + x_2y_3 - x_3y_2\right) \tag{1.13}$$

$$N_2(x,y) = \frac{1}{2A}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x & x_3 \\ y_1 & y & y_3 \end{vmatrix} = \frac{1}{2A}\left(yx_1 - xy_1 + xy_3 - yx_3 - x_1y_3 + y_1x_3\right) \tag{1.14}$$

$$N_3(x,y) = \frac{1}{2A}\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x \\ y_1 & y_2 & y \end{vmatrix} = \frac{1}{2A}\left(xy_1 - yx_1 - xy_2 + yx_2 + x_1y_2 - x_2y_1\right) \tag{1.15}$$

Notice how the properties of simplicia enable the form functions to be easily determined.

1.2.4. *Linear along each direction elements on quadrilaterals in 2D.* The element E has 4 nodes $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4\}$. It is convenient to introduce a local coordinate system (ξ, η) so that the nodes correspond to the local coordinates $(\pm 1, \pm 1)$. The local approximation is then given in the local coordinates

by

$$\tilde{q}(\xi,\eta) = \sum_{k=1}^{4} N_k(\xi,\eta) Q_k \tag{1.16}$$

with the form functions

$$N_1(\xi,\eta) = \frac{1}{4}(1+\xi)(1+\eta) \tag{1.17}$$

$$N_2(\xi,\eta) = \frac{1}{4}(1-\xi)(1+\eta) \tag{1.18}$$

$$N_3(\xi,\eta) = \frac{1}{4}(1-\xi)(1-\eta) \tag{1.19}$$

$$N_4(\xi,\eta) = \frac{1}{4}(1+\xi)(1-\eta) \tag{1.20}$$

The local transformation of coordinates can also be written in terms of the form functions

$$x(\xi,\eta) = \sum_{k=1}^{4} N_k(\xi,\eta) x_k,\ y(\xi,\eta) = \sum_{k=1}^{4} N_k(\xi,\eta) y_k \tag{1.21}$$

2. Variational derivation of weighted residual formulations

We now turn to the problem of how to obtain a measure of the error introduced in approximating the exact solution q to the PDE of interest with its piecewise approximation $\tilde{q}$. Some techniques were presented in the general presentation of weighted residual methods carried out in Chapter 2. For a wide class of problems of interest, especially elliptic problems there exist alternative formulations that lead to more efficient numerical algorithms. These are based upon variational and functional analysis and we shall consider the basics of the theory here.

2.1. Variational calculus. Consider the problem of determining the extremum of the integral

$$I(q) = \int_a^b f(x, q, q')dx \tag{2.1}$$

over all functions $q : \mathbb{R} \to \mathbb{R}$ that belong to some class, for example piecewise continuous functions and that satisfy the boundary conditions $q(x = a) = q_a$, $q(x = b) = q_b$. $I(q)$ is called a functional in that it associates a scalar value to each element from a space of functions. We can consider small perturbations of the function q that we denote by $\delta\dot{q}$ The perturbations maintain the boundary conditions, i.e.

$$\delta q(x = a) = 0,\ \delta q(x = b) = 0\ . \tag{2.2}$$

The change in I is

$$\delta I = I(q + \delta q) - I(q) = \int_a^b f(x, q + \delta q, q' + \delta q')dx - \int_a^b f(x, q, q')dx\ . \tag{2.3}$$

We shall consider q, q' as independent variables in f and carry out series expansions to obtain

$$\delta I = \int_a^b \left[\left(\frac{\partial f}{\partial q} \right) \delta q + \left(\frac{\partial f}{\partial q'} \right) \delta q' \right] dx\ . \tag{2.4}$$

We can interchange the δ and d/dx operators in the second term and then integrate by parts

$$\int_a^b \left(\frac{\partial f}{\partial q'}\right) \delta q' dx = \int_a^b \left(\frac{\partial f}{\partial q'}\right) \delta \frac{dq}{dx} dx = \int_a^b \left(\frac{\partial f}{\partial q'}\right) \frac{d}{dx} (\delta q)\, dx = \tag{2.5}$$

$$= \left(\frac{\partial f}{\partial q'}\right) (\delta q)\Big|_{x=a}^{x=b} - \int_a^b \frac{d}{dx}\left(\frac{\partial f}{\partial q'}\right) (\delta q)\, dx \tag{2.6}$$

Applying the boundary conditions and then replacing the above result in (2.4) leads to

$$\delta I = \int_a^b \left[\left(\frac{\partial f}{\partial q}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial q'}\right)\right] \delta q dx\ . \tag{2.7}$$

For I to be at an extremum δI must maintain the same sign under any perturbation of the extremum. This is only possible if the factor multiplying δq in the above integral is zero everywhere. If it were not then δq_1 would give some value δI_1 and $-\delta q_1$ would lead to the opposite value $-\delta I_1$ and I would not be at an extremum. We therefore have

$$\left(\frac{\partial f}{\partial q}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial q'}\right) = 0 \tag{2.8}$$

as the condition for I to be at an extremum. This is known as the *Euler variational principle*. At the extremum we obviously have $\delta I = 0$.

The importance of the Euler variational principle for numerical solution of PDE's rests upon the link it furnishes between an integral formulation $I(q)$ and a differential equation (2.8). We can write down specific forms of f that lead to PDE's of great practical interest. For example replacing

$$f(x, q, q') = \frac{1}{2}\left(\frac{dq}{dx}\right)^2 - gq \tag{2.9}$$

in (2.8) leads to the differential equation

$$q'' = g \tag{2.10}$$

with the boundary conditions $q(x = a) = q_a$, $q(x = b) = q_b$. This is the standard 2 point boundary problem for a second order ODE. Recall that this can be solved by either direct discretization leading to the linear system of equations

$$Q_{j-1} - 2Q_j + Q_{j+1} = h^2 g_j,\ j = 1, \ldots, N-1 \tag{2.11}$$

or by using a shooting method combined with an initial value solve in which we seek $z = q'(x = a)$ that leads to $q(x = b; z) = q_b$. The variational formulation above suggests a third approach. Instead of directly solving the ODE we can seek q that minimizes $I(q)$ with f given by (2.9). This is extremely useful in constructing finite element approximations as we will see below.

Other important expressions of the Euler variational principle can be derived for various situations. Let us consider the ones most often encountered.

(1) Functional of two functions in 1D. The functional is

$$I(p, q) = \int_a^b f(x, p, p', q, q')dx \tag{2.12}$$

and the Euler variational principle leads to

$$\left(\frac{\partial f}{\partial p}\right) + \left(\frac{\partial f}{\partial q}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial p'}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial q'}\right) = 0 \tag{2.13}$$

(2) Functional of a 2D function.

$$I(q) = \int_c^d \int_a^b f(x, y, q, q_x, q_y) dx dy \tag{2.14}$$

$$\left(\frac{\partial f}{\partial q}\right) - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial q_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial q_y}\right) = 0 \tag{2.15}$$

(3) Functional involving second order derivatives in 1D.

$$I(p, q) = \int_a^b f(x, p, p', p'') dx \tag{2.16}$$

$$\left(\frac{\partial f}{\partial q}\right) - \frac{d}{dx}\left(\frac{\partial f}{\partial q'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial f}{\partial q''}\right) = 0 \tag{2.17}$$

2.2. Ritz methods. In the Ritz formulation of the finite element method we seek a piecewise approximation that minimizes the functional associated with the PDE of interest. The piecewise local approximation can be expressed as

$$\tilde{q}(x) = \sum_e \sum_k Q_k^e N_k^e(x) \tag{2.18}$$

where the e sum is over all elements and the k sum is over all nodes within an element. The unknowns of the problem are the nodal values Q_k^e. The form functions $N_k^e(x)$ correspond to some chosen approximation scheme. Let f be associated with the PDE we are interested in solving. The problem reduces to finding $\{Q_k^e\}$ that minimizes

$$I(\tilde{q}) = \int_a^b f(x, \tilde{q}, \tilde{q}_x) dx \ . \tag{2.19}$$

This can be solved by finding the solution to the system of equations

$$\frac{\partial}{\partial Q_k^e} I(\tilde{q}) = 0 \tag{2.20}$$

with e going over all elements and k over all element nodes.

Note that the entire procedure rests upon the ability to determine a function f that corresponds to a PDE of practical interest. In many situations we have physical guidance that such a variational principle formulation exists. The basic underpinning is furnished by analytical mechanics and the physical principle of least action which finds various expressions in different disciplines. The principle of least action asserts that of all the generalized trajectories $(p, q) = \{q_k(t), p_k(t) \mid k = 1, \ldots, 3N\}$ of a system of N particles, the one actually followed minizes the action S

$$S = \int_{t_0}^{t_1} \mathcal{L}(t, p, q) dt \tag{2.21}$$

with $\mathcal{L}$ being the Lagrangean of the system. Here q_k denote generalized coordinates and p_k generalized momenta. Though not always immediately apparent this leads to other expressions typically called *minimum energy functionals*. These can be

written for systems with no dissipative effects. Here are some examples of functions f linked to important PDE's:

(1) Poisson equation in 2D

$$f = \frac{1}{2}\left(q_x^2 + q_y^2\right) - gq \tag{2.22}$$

for which (2.15) gives

$$q_{xx} + q_{yy} = g \tag{2.23}$$

(2) Poisson equation in 3D

$$f = \frac{1}{2}\left(q_x^2 + q_y^2 + q_z^2\right) - gq \tag{2.24}$$

for which the Euler variational principle

$$\left(\frac{\partial f}{\partial q}\right) - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial q_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial q_y}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial q_z}\right) = 0 \tag{2.25}$$

gives

$$q_{xx} + q_{yy} + q_{zz} = g \tag{2.26}$$

2.3. Galerkin methods. The Ritz formulation typically leads to a system of equations which has nice numerical properties. However there are many systems for which a variational formulation is not possible typically because the system has dissipative behavior. In such situations we can again use an integral reformulation of the PDE of interest based upon the concept of a weak solution already introduced in the study of hyperbolic problems. Suppose we're looking for a solution to the problem

$$\mathcal{A}q = g \tag{2.27}$$

with $\mathcal{A}$ some differential operator. A function q that directly satisfies (2.27) is called a *classical solution*. Consider now some space of test functions v and a scalar product defined for the functions q and v. From (2.27) we can derive

$$(\mathcal{A}q, v) = (g, v) \tag{2.28}$$

where $(\cdot, \cdot)$ denotes the scalar product, e.g.

$$(u, v) = \int_a^b u(x)v(x)dx \ . \tag{2.29}$$

In (2.28) we can apply integration by parts to obtain

$$(q, \mathcal{A}^* v) = (g, v) \tag{2.30}$$

where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$. This typically enables us to avoid differentiating functions q that might be discontinuous. We can now use (2.30) to determine the unknown coefficients of a finite element approximation

$$\tilde{q}(x) = \sum_e \sum_k Q_k^e N_k^e(x) \tag{2.31}$$

by requiring

$$\sum_e \sum_k Q_k^e \left(N_k^e(x), \mathcal{A}^* v\right) = (g, v) \ . \tag{2.32}$$

The only piece missing is how we choose the test functions v. In a Galerkin formulation these are chosen to be the form functions themselves leading to

$$\sum_e \sum_k Q_k^e \left(N_k^e(x), \mathcal{A}^* N_j^e(x)\right) = \left(g, N_j^e(x)\right) , \tag{2.33}$$

thus defining a linear system

$$\mathbf{A}Q = b \tag{2.34}$$

$$A_{jk} = \left(N_k^e(x), \mathcal{A}^* N_j^e(x)\right) \ . \tag{2.35}$$

2.4. A detailed example. Let us now carry out the steps involved in solving a Poisson equation in 2D using a Ritz formulation and quadrilateral elements. The mathematical statement of the problem is

$$\begin{cases} q_{xx} + q_{yy} = g & (x,y) \in \Omega \\ q = b & (x,y) \in \partial\Omega \end{cases} \tag{2.36}$$

with the domain $\Omega = [a,b] \times [c,d]$ and $\partial\Omega$ denoting the boundary of Ω on which Dirichlet conditions are given. The element form functions are given by (1.17)-(1.20) and the function f is given by (2.22). The function $I(\tilde{q})$ is

$$I(\tilde{q}) = \int_c^d \int_a^b f(x,y,\tilde{q},\tilde{q}_x,\tilde{q}_y)dxdy = \int_c^d \int_a^b \left[\frac{1}{2}\left(\tilde{q}_x^2 + \tilde{q}_y^2\right) - g\tilde{q}\right] dxdy \ . \tag{2.37}$$

The finite element approximation is determined by the chosen form functions and the nodal values $\{Q_k^e\}$. The extremum of $I(\tilde{q})$ is attained when

$$\frac{\partial}{\partial Q_k^e} I(\tilde{q}) = 0 \tag{2.38}$$

which leads to

$$\int_c^d \int_a^b \left[\left(\tilde{q}_x \frac{\partial \tilde{q}_x}{\partial Q_k^e} + \tilde{q}_y \frac{\partial \tilde{q}_y}{\partial Q_k^e}\right) - g\frac{\partial \tilde{q}}{\partial Q_k^e}\right] dxdy = 0 \tag{2.39}$$

Note that

$$\frac{\partial \tilde{q}}{\partial Q_k^e} = N_k^e, \ \frac{\partial \tilde{q}_x}{\partial Q_k^e} = \frac{\partial N_k^e}{\partial x}, \ \frac{\partial \tilde{q}_y}{\partial Q_k^e} = \frac{\partial N_k^e}{\partial y} \tag{2.40}$$

so these derivatives no longer contain the unknowns $\{Q_k^e\}$. We thus obtain

$$\sum_e \sum_j \left[\iint \left(\frac{\partial N_j^e}{\partial x}\frac{\partial N_k^e}{\partial x} + \frac{\partial N_j^e}{\partial y}\frac{\partial N_k^e}{\partial y}\right) dxdy\right] Q_k^e = \sum_e \iint g N_k^e \ dxdy \tag{2.41}$$

with k going over all the element nodes. The sum over the elements is typically known as an assembly operation, leading to the computation of the matrix elements

$$A_{jk} = \sum_e \iint \left(\frac{\partial N_j^e}{\partial x}\frac{\partial N_k^e}{\partial x} + \frac{\partial N_j^e}{\partial y}\frac{\partial N_k^e}{\partial y}\right) dxdy \tag{2.42}$$

known as the system *stiffness matrix*. We can easily compute the elements of this matrix. Analytical computation is possible as in

$$\frac{\partial N_k^e}{\partial x} = \frac{\partial N_k^e}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial N_k^e}{\partial \eta}\frac{\partial \eta}{\partial x} \tag{2.43}$$

$$\frac{\partial \xi}{\partial x} = \frac{\frac{D(\xi,y)}{D(\xi,\eta)}}{\frac{D(x,y)}{D(\xi,\eta)}} = \frac{1}{J}\begin{vmatrix} 1 & 0 \\ y_\xi & y_\eta \end{vmatrix} = \frac{y_\eta}{J} \tag{2.44}$$

$$\frac{\partial \eta}{\partial x} = \frac{\frac{D(\eta,y)}{D(\xi,\eta)}}{\frac{D(x,y)}{D(\xi,\eta)}} = \frac{1}{J}\begin{vmatrix} 0 & 1 \\ y_\xi & y_\eta \end{vmatrix} = -\frac{y_\xi}{J} \tag{2.45}$$

$$J = \begin{vmatrix} x_\xi & x_\eta \\ y_\xi & y_\eta \end{vmatrix} = x_\xi y_\eta - x_\eta y_\xi \tag{2.46}$$

The $x(\xi,\eta)$ and $y(\xi,\eta)$ dependencies are given by (1.21) so we obtain

$$\frac{\partial x}{\partial \xi} = \sum_{k=1}^{4} \frac{\partial N_k}{\partial \xi} x_k, \;\; \frac{\partial y}{\partial \xi} = \sum_{k=1}^{4} \frac{\partial N_k}{\partial \xi} y_k \tag{2.47}$$

$$\frac{\partial x}{\partial \eta} = \sum_{k=1}^{4} \frac{\partial N_k}{\partial \eta} x_k, \;\; \frac{\partial y}{\partial \eta} = \sum_{k=1}^{4} \frac{\partial N_k}{\partial \eta} y_k \tag{2.48}$$

But analytical evaluations are not really required in this case. We can recognize that the integrand in (2.42) is quadratic in (x,y) and that a 4-point Gauss-Legendre quadrature leads to an exact evaluation.