Chapter 9

The Finite Element Method for 2D elliptic PDEs

The procedure of the finite element method to solve 2D problems is the same as that for 1D problems, as the flow chart below demonstrates.

$$
\begin{aligned}
&\text{PDE} \longrightarrow \text{Integration by parts} \longrightarrow \text{weak form in } V : a(u,v) = L(v)\\
&\text{or } \min_{v\in V} F(v) \longrightarrow V_h \text{ (finite dimensional space and basis functions)}\\
&\longrightarrow a(u_h, v_h) = L(v_h) \longrightarrow u_h \text{ and error analysis.}
\end{aligned}
$$

9.1 The second Green's theorem and integration by parts in 2D

Let us first recall the 2D version of the well known divergence theorem in Cartesian coordinates.

Theorem 9.1. *If* $\mathbf{F} \in H^1(\Omega) \times H^1(\Omega)$ *is a vector in 2D, then*

$$\iint_\Omega \nabla\cdot\mathbf{F}\, dx\, dy = \int_{\partial\Omega} \mathbf{F}\cdot\mathbf{n}\, ds\,, \tag{9.1}$$

where $\mathbf{n}$ *is the unit normal direction pointing outward at the boundary* $\partial\Omega$ *with line element* ds*, and* ∇ *is the gradient operator once again is* $\nabla = [\frac{\partial}{\partial x}\,,\ \frac{\partial}{\partial y}]^T$.

The second Green's theorem is a corollary of the divergence theorem if we set $\mathbf{F} = v\,\nabla u = \left[v\dfrac{\partial u}{\partial x}, v\dfrac{\partial u}{\partial y}\right]^T$. Thus since

$$
\begin{aligned}
\nabla\cdot\mathbf{F} &= \frac{\partial}{\partial x}\left(v\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(v\frac{\partial u}{\partial y}\right)\\
&= \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + v\frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} + v\frac{\partial^2 u}{\partial y^2}\\
&= \nabla u \cdot \nabla v + v\,\Delta u,
\end{aligned}
$$

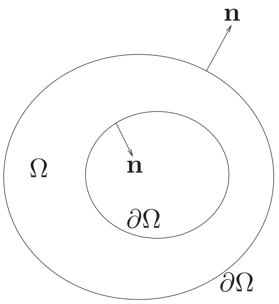

Figure 9.1. *A diagram of a two dimensional domain Ω, its boundary ∂Ω and its unit normal direction.*

where $\Delta u = \nabla\cdot\nabla u = u_{xx} + u_{yy}$, we obtain

$$\begin{aligned}\iint_\Omega \nabla\cdot\mathbf{F}\,dxdy &= \iint_\Omega (\nabla u\cdot\nabla v + v\,\Delta u)\,dxdy\\ &= \int_{\partial\Omega} \mathbf{F}\cdot\mathbf{n}\,ds\\ &= \int_{\partial\Omega} v\,\nabla u\cdot\mathbf{n}\,ds = \int_{\partial\Omega} v\,\frac{\partial u}{\partial n}\,ds\,,\end{aligned}$$

where $\mathbf{n} = (n_x, n_y)$ $(n_x^2 + n_y^2 = 1)$ is the unit normal direction, and $\frac{\partial u}{\partial n} = \nabla u\cdot\mathbf{n} = n_x\frac{\partial u}{\partial x} + n_y\frac{\partial u}{\partial y}$, the normal derivative derivative of u, see Fig. 9.1 for an illustration. This result immediately yields the formula for integration by parts in 2D.

Theorem 9.2. *If $u(x,y)\in H^2(\Omega)$ and $v(x,y)\in H^1(\Omega)$ where Ω is a bounded domain, then*

$$\iint_\Omega v\,\Delta u\,dxdy = \int_{\partial\Omega} v\frac{\partial u}{\partial n}\,ds - \iint_\Omega \nabla u\cdot\nabla v\,dxdy\,. \tag{9.2}$$

Note: the normal derivative $\partial u/\partial n$ is sometimes written more concisely as u_n.

Some important elliptic PDEs in 2D Cartesian coordinates are:

$$\begin{aligned} u_{xx} + u_{yy} &= 0, && \text{Laplace equation,}\\ -u_{xx} - u_{yy} &= f(x,y), && \text{Poisson equation,}\\ -u_{xx} - u_{yy} + \lambda u &= f, && \text{generalized Helmholtz equation,}\\ u_{xxxx} + 2u_{xxyy} + u_{yyyy} &= 0, && \text{Bi-harmonic equation.}\end{aligned}$$

When $\lambda > 0$, the generalized Helmholtz equation is easier to solve than when $\lambda < 0$. Incidentally, the expressions involved in these PDEs may also be abbreviated using the gradient operator ∇, *e.g.*, $u_{xx} + u_{yy} = \nabla\cdot\nabla u = \Delta u$ as mentioned before. We also recall that a general linear second order elliptic PDE has the form

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

with discriminant $b^2 - ac < 0$. A second order self-adjoint elliptic partial differential equation has the form

$$-\nabla \cdot (p(x,y)\nabla u) + q(x,y)u = f(x,y)\,. \tag{9.3}$$

9.1.1 Boundary conditions

In 2D, the domain boundary $\partial\Omega$ is one or several curves. We consider the following various linear boundary conditions.

- Dirichlet boundary condition on the entire boundary, *i.e.*, $u(x,y)|_{\partial\Omega} = u_0(x,y)$ is given.
- Neumann boundary condition on the entire boundary, *i.e.*, $\partial u/\partial n|_{\partial\Omega} = g(x,y)$ is given.

  In this case, the solution to a Poisson equation may not be unique or even exist, depending upon whether a compatibility condition is satisfied. Integrating the Poisson equation over the domain, we have

  $$\begin{aligned}\iint_\Omega f dxdy &= -\iint_\Omega \Delta u\, dxdy = -\iint_\Omega \nabla\cdot\nabla u\, dxdy \\ &= -\int_{\partial\Omega} u_n\, ds = -\int_{\partial\Omega} g(x,y)\, ds\,,\end{aligned} \tag{9.4}$$

  which is the compatibility condition to be satisfied for the solution to exist. If a solution does exist, it is not unique as it is determined within an arbitrary constant.
- Mixed boundary condition on the entire boundary, *i.e.*,

  $$\alpha(x,y)u(x,y) + \beta(x,y)\frac{\partial u}{\partial n} = \gamma(x,y)$$

  is given, where $\alpha(x,y)$, $\beta(x,y)$, and $\gamma(x,y)$ are known functions.
- Dirichlet, Neumann, and Mixed boundary conditions on some parts of the boundary.

9.2 Weak form of second order self-adjoint elliptic PDEs

Now we derive the weak form of the self-adjoint PDE (9.3) with a homogeneous Dirichlet boundary condition on part of the boundary $\partial\Omega_D$, $u|_{\partial\Omega_D} = 0$ and a homogeneous Neumann boundary condition on the rest of boundary $\partial\Omega_N = \partial\Omega - \partial\Omega_D$, $\frac{\partial u}{\partial n}|_{\partial\Omega_N} = 0$. Multiplying the equation (9.3) by a test function $v(x,y) \in H^1(\Omega)$, we have

$$\iint_\Omega \Big\{ -\nabla\cdot(p(x,y)\nabla u) + q(x,y)\,u \Big\}\, v\, dxdy = \iint_\Omega fv\, dxdy\,;$$

and on using the formula for integration by parts the left-hand side becomes

$$\iint_\Omega \Big(p\nabla u\cdot\nabla v + quv \Big)\, dxdy - \int_{\partial\Omega} pvu_n\, ds\,,$$

so the weak form is

$$\iint_{\Omega} (p\nabla u \cdot \nabla v + quv)\, dxdy = \iint_{\Omega} fvdxdy + \int_{\partial\Omega_N} pg(x,y)v(x,y)\, ds \qquad \forall v(x,y) \in H^1(\Omega)\,. \tag{9.5}$$

Here $\partial\Omega_N$ is the part of boundary where a Neumann boundary condition is applied; and the solution space resides in

$$V = \Big\{ v(x,y)\ ,\ v(x,y) = 0\ ,\ (x,y) \in \partial\Omega_D\ ,\ v(x,y) \in H^1(\Omega) \Big\}\,, \tag{9.6}$$

where $\partial\Omega_D$ is the part of boundary where a Dirichlet boundary condition is applied.

9.2.1 Verification of conditions of the Lax-Milgram Lemma

The bilinear form for (9.3) is

$$a(u,v) = \iint_{\Omega} (p\nabla u \cdot \nabla v + quv)\, dxdy\,, \tag{9.7}$$

and the linear form is

$$L(v) = \iint_{\Omega} fv\, dxdy \tag{9.8}$$

for a Dirichlet BC on the entire boundary. As before, we assume that

$$0 < p_{min} \le p(x,y) \le p_{max}\ ,\ 0 \le q(x) \le q_{max}\ ,\ p \in C(\Omega)\ ,\ q \in C(\Omega)\,.$$

We need the Poincaré inequality to prove the V-elliptic condition.

Theorem 9.3. *If $v(x,y) \in H_0^1(\Omega)$, $\Omega \subset R^2$,* i.e., *$v(x,y) \in H^1(\Omega)$ and vanishes at the boundary $\partial\Omega$ (can be relaxed to a point on the boundary), then*

$$\iint_{\Omega} v^2 dxdy \le C \iint_{\Omega} |\nabla v|^2\, dxdy\,, \tag{9.9}$$

where C is a constant.

Now we are ready to check the conditions of the Lax-Milgram Lemma.

1. It is obvious that $a(u,v) = a(v,u)$.
2. It is easy to see that

$$\begin{aligned} |a(u,v)| &\le \max\{p_{max}, q_{max}\} \left| \iint_{\Omega} (|\nabla u \cdot \nabla v| + |uv|)\, dxdy \right| \\ &= \max\{p_{max}, q_{max}\}\, |(|u|,|v|)_1| \\ &\le \max\{p_{max}, q_{max}\}\, \|u\|_1 \|v\|_1\,, \end{aligned}$$

so $a(u,v)$ is a continuous and bounded bilinear operator.

3. From the Poincaré inequality

$$\begin{aligned} |a(v,v)| &= \left| \iint_\Omega p\left(|\nabla v|^2 + qv^2 \right) dxdy \right| \\ &\geq p_{min} \iint_\Omega |\nabla v|^2 \, dxdy \\ &= \frac{1}{2} p_{min} \iint_\Omega |\nabla v|^2 \, dxdy + \frac{1}{2} p_{min} \iint_\Omega |\nabla v|^2 \, dxdy \\ &\geq \frac{1}{2} p_{min} \iint_\Omega |\nabla v|^2 \, dxdy + \frac{p_{min}}{2C} \iint_\Omega |v|^2 \, dxdy \\ &\geq \frac{1}{2} p_{min} \min \left\{ 1, \ \frac{1}{C} \right\} \|v\|_1^2 \, , \end{aligned}$$

therefore $a(u,v)$ is V-elliptic.

4. Finally, we show that $L(v)$ is continuous:

$$|L(v)| = |(f,v)_0| \leq \|f\|_0 \|v\|_0 \leq \|f\|_0 \|v\|_1 \, .$$

Consequently, the solutions to the weak form and the minimization form are unique and bounded in $H_0^1(\Omega)$.

9.3 Triangulation and basis functions

The general procedure of the finite element method is the same for any dimension, and the Galerkin finite element method involves the following main steps.

- Generate a triangulation over the domain. Usually the triangulation is composed of either triangles or rectangles. There are a number of mesh generation software packages available, *e.g.*, the Matlab PDE toolbox from Mathworks, Triangle from Carnegie Mellon University, *etc*. Some are available through the Internet.
- Construct basis functions over the triangulation. We mainly consider the conforming finite element method in this book.
- Assemble the stiffness matrix and the load vector element by element, using either the Galerkin finite method (the weak form) or the Ritz finite method (the minimization form).
- Solve the system of equations.
- Do the error analysis.

In Fig. 9.2, we show a diagram of simple mesh generation process. The circular domain is approximated by a polygon with five vertices (selected points on the boundary). We then connect the five vertices to get initial five triangles (solid line) to obtain an initial coarse mesh. We can refine the mesh using the so called middle point rule by connecting all the middle points of all triangles in the initial mesh to obtain a finer mesh (solid and dashed lines).

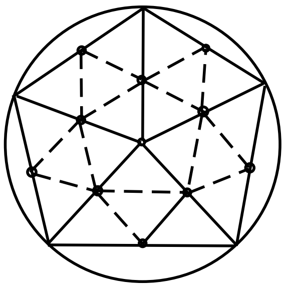

Figure 9.2. *A diagram of a simple generation process and the middle point rule.*

9.3.1 Triangulation and mesh parameters

Given a general domain, we can approximate the domain by a polygon and then generate a triangulation over the polygon, and we can refine the triangulation if necessary. A simple approach is the mid-point rule by connecting all the middle points of three sides of existing triangles to get a refined mesh.

A triangulation usually has the mesh parameters

$$\begin{aligned}
\Omega_p :&\quad \text{polygonal region} = K_1 \cup K_2 \cup K_3 \cdots \cup K_{nelem}\,,\\
K_j :&\quad \text{are non-overlapping triangles, } j = 1, 2, \cdots, nelem\,,\\
N_i :&\quad \text{are nodal points, } i = 1, 2, \cdots, nnode\,,\\
h_j :&\quad \text{the longest side of } K_j\,,\\
\rho_j :&\quad \text{the diameter of the circle inscribed in } K_j \text{ (encircle)},\\
h :&\quad \text{the largest of all } h_j, \quad h = \max\{h_j\}\,,\\
\rho :&\quad \text{the smallest of all } \rho_j, \quad \rho = \min\{\rho_j\}\,,
\end{aligned}$$

with

$$1 \ge \frac{\rho_j}{h_j} \ge \beta > 0\,,$$

where the constant β is a measurement of the triangulation quality, see Fig. 9.4 for an illustration of such a ρ's and h's. The larger the β, the better the quality of the triangulation. Given a triangulation, a node is also the vertex of all adjacent triangles. We do not discuss hanging nodes here.

9.3.2 The FE space of piecewise linear functions over a triangulation

For linear second order elliptic PDEs, we know that the solution space is in the $H^1(\Omega)$. Unlike the 1D case, an element $v(x, y)$ in $H^1(\Omega)$ may not be continuous under the Sobolev embedding theorem. However, in practice most solutions are indeed continuous, especially for second order PDEs with certain regularities. Thus, we still look for a solution in the continuous function space $C^0(\Omega)$. Let us first consider how to construct piecewise linear functions over a triangulation with the Dirichlet BC

$$u(x,y)|_{\partial\Omega} = 0\,.$$

Given a triangulation, we define

$$\begin{aligned} V_h \quad = \quad & \Big\{ v(x,y) \text{ is continuous in } \Omega \text{ and piecewise linear over each } K_j\,, \\ & v(x,y)|_{\partial\Omega} = 0 \,\Big\}\,. \end{aligned} \tag{9.10}$$

We need to determine the dimension of this space and construct a set of basis functions. On each triangle, a linear function has the form

$$v_h(x,y) = \alpha + \beta x + \gamma y\,, \tag{9.11}$$

where α, β and γ are constants (three free parameters). Let

$$P_k = \{\, p(x,y)\,, \text{ a polynomial of degree of } k\,\}\,. \tag{9.12}$$

We have the following theorem.

Theorem 9.4.

1. *A linear function $p_1(x,y) = \alpha+\beta x+\gamma y$ defined on a triangle is uniquely determined by its values at the three vertices.*
2. *If $p_1(x,y) \in P_1$ and $p_2(x,y) \in P_1$ are such that $p_1(A) = p_2(A)$ and $p_1(B) = p_2(B)$, where A and B are two points in the xy-plane, then $p_1(x,y) \equiv p_2(x,y)$, $\forall (x,y) \in I_{AB}$, where I_{AB} is the line segment between A and B.*

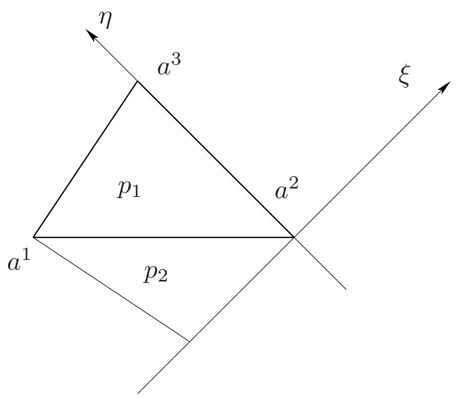

Figure 9.3. *A diagram of a triangle with three vertices a^1, a^2, and a^3; an adjacent triangle with a common side; and the local coordinate system in which a^2 is the origin and a^2a^3 is the η axis.*

Proof: Assume the vertices of the triangle are (x_i, y_i), $i = 1, 2, 3$. The linear function takes the value v_i at the vertices, *i.e.*,

$$p(x_i, y_i) = v_i \,,$$

so we have the three equations

$$\begin{aligned} \alpha + \beta x_1 + \gamma y_1 &= v_1 \,, \\ \alpha + \beta x_2 + \gamma y_2 &= v_2 \,, \\ \alpha + \beta x_3 + \gamma y_3 &= v_3 \,. \end{aligned}$$

The determinant of this linear algebraic system is

$$det \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} = \pm 2 \text{ area of the triangle} \neq 0 \;\text{ since }\; \frac{\rho_j}{h_j} \geq \beta > 0 \,, \tag{9.13}$$

hence the linear system of equations has a unique solution.

Now let us prove the second part of the theorem. Suppose that the equation of the line segment is

$$l_1 x + l_2 y + l_3 = 0 \,, \quad l_1^2 + l_2^2 \neq 0 \,.$$

We can solve for x or for y:

$$\begin{aligned} x &= -\frac{l_2 y + l_3}{l_1} \quad \text{if} \quad l_1 \neq 0 \,, \\ \text{or} \quad y &= -\frac{l_1 x + l_3}{l_2} \quad \text{if} \quad l_2 \neq 0 \,. \end{aligned}$$

Without loss of generality, let us assume $l_2 \neq 0$ such that

$$\begin{aligned} p_1(x, y) &= \alpha + \beta x + \gamma y \\ &= \alpha + \beta x - \frac{l_1 x + l_3}{l_2} \gamma \\ &= \left(\alpha - \frac{l_3}{l_2}\gamma\right) + \left(\beta - \frac{l_1}{l_2}\gamma\right) x \\ &= \alpha_1 + \beta_1 x \,. \end{aligned}$$

Similarly, we have

$$p_2(x, y) = \bar{\alpha}_1 + \bar{\beta}_1 x \,.$$

Since $p_1(A) = p_2(A)$ and $p_1(B) = p_2(B)$,

$$\begin{aligned} \alpha_1 + \beta_1 x_1 &= p(A) \,, & \bar{\alpha}_1 + \bar{\beta}_1 x_1 &= p(A) \,, \\ \alpha_1 + \beta_1 x_2 &= p(B) \,, & \bar{\alpha}_1 + \bar{\beta}_1 x_2 &= p(B) \,, \end{aligned}$$

where both of the linear system of algebraic equations have the same coefficient matrix

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}$$

that is non-singular since $x_1 \neq x_2$ (because points A and B are distinct). Thus we conclude that $\alpha_1 = \bar{\alpha}_1$ and $\beta_1 = \bar{\beta}_1$, so the two linear functions have the same expression along the line segment, *i.e.*, they are identical along the line segment.

Corollary 9.5. *A piecewise linear function in $C^0(\Omega) \cap H^1(\Omega)$ over a triangulation (a set of non-overlapping triangles) is uniquely determined by its values at the vertices.*

Theorem 9.6. *The dimension of the finite dimensional space composed of piecewise linear functions in $C^0(\Omega) \cap H^1(\Omega)$ over a triangulation for (9.3) is the number of interior nodal points plus the number of nodal points on the boundary where the natural BC are imposed (Neumann and mixed boundary conditions).*

Example 9.1. *Given the triangulation shown in Fig. 9.4, a piecewise continuous function $v_h(x,y)$ is determined by its values on the vertices of all triangles, more precisely, $v_h(x,y)$ is determined from*

$$\begin{aligned}
&(0,0,v(N_1))\,, \quad (x,y) \in K_1\,, && (0,v(N_2),v(N_1))\,, \quad (x,y) \in K_2\,,\\
&(0,0,v(N_2))\,, \quad (x,y) \in K_3\,, && (0,0,v(N_2))\,, \quad (x,y) \in K_4\,,\\
&(0,v(N_3),v(N_2))\,, \quad (x,y) \in K_5\,, && (0,0,v(N_3))\,, \quad (x,y) \in K_6\,,\\
&(0,v(N_1),v(N_3))\,, \quad (x,y) \in K_7\,, && (v(N_1),v(N_2),v(N_3))\,, \quad (x,y) \in K_8\,.
\end{aligned}$$

Note that although three values of the vertices are the same, like the values for K_3 and K_4, the geometries are different, hence, the functions will likely have different expressions on different triangles.

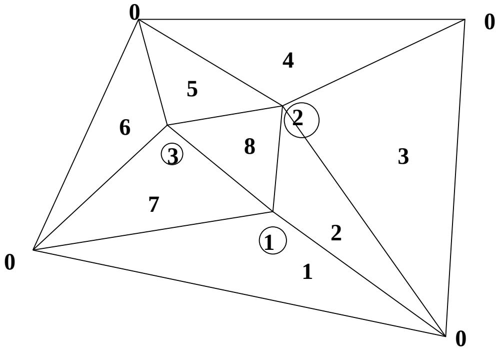

Figure 9.4. *A diagram of a simple triangulation with a homogeneous boundary condition.*

9.3.3 Global basis functions

A global basis function of the piecewise linear functions in $C^0(\Omega) \cap H^1(\Omega)$ can be defined as

$$\phi_i(N_j) = \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{9.14}$$

where N_j are nodal points. The shape (mesh plot) of $\phi_i(N_j)$ looks like a "tent" without a door; and its support of $\phi_i(N_j)$ is the union of the triangles surrounding the node N_i, *cf.*, Fig. 9.5, where Fig. 9.5 (a) is the mesh plot of the global basis function, and Fig. 9.5 (b) is the plot of a triangulation and the contour plot of the global basis function centered at a node. The basis function is piecewise linear and it is supported only in the surrounding triangles.

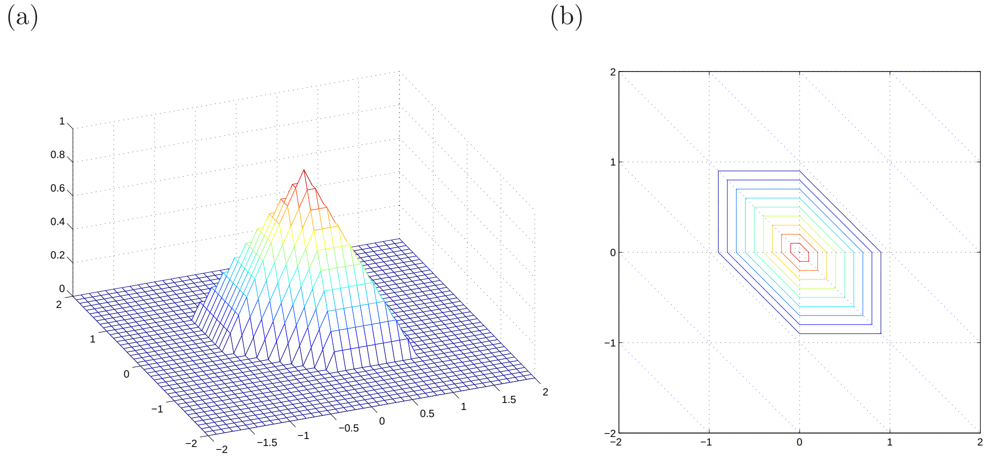

Figure 9.5. *A global basis function ϕ_j. (a) the mesh plot of the global function; (b) the triangulation and the contour plot of the global basis function.*

It is almost impossible to give a closed form of a global basis function except for some very special geometries (*cf.*, the example in the next section). However, it is much easier to write down the shape function.

Example 9.2. *Let us consider a Poisson equation and a uniform mesh, as an example to demonstrate the piecewise linear basis functions and the finite element method:*

$$\begin{aligned} -(u_{xx} + u_{yy}) &= f(x, y)\,, \quad (x, y) \in (a,\, b) \times (c,\, d)\,, \\ u(x, y)|_{\partial\Omega} &= 0\,. \end{aligned}$$

We know how to use the standard central finite difference scheme with the five point stencil to solve the Poisson equation. With some manipulations, the linear system of equations

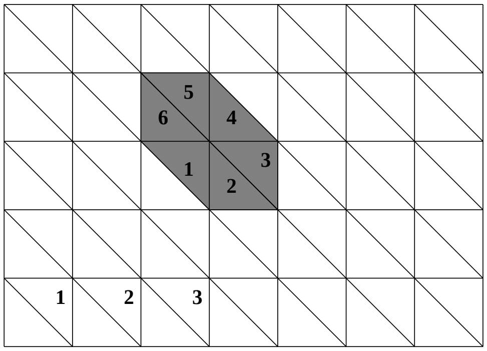

Figure 9.6. *A uniform triangulation defined on a rectangular domain.*

on using the finite element method with a uniform triangulation (*cf.*, Fig. 9.6) proves to be the same as that obtained from the finite difference method.

Given a uniform triangulation as shown in Fig. 9.6, if we use row-wise natural ordering for the nodal points

$$(x_i, y_j)\,, \quad x_i = ih\,, \quad y_j = jh\,, \quad h = \frac{1}{n}\,, \quad i = 1, 2, \cdots, m-1\,, \;\; j = 1, 2, \cdots, n-1\,,$$

then the global basis function defined at $(x_i, y_j) = (ih, jh)$ are

$$\phi_{j(n-1)+i} = \begin{cases} \dfrac{x-(i-1)h+y-(j-1)h}{h} - 1 & \text{Region 1} \\ \dfrac{y-(j-1)h}{h} & \text{Region 2} \\ \dfrac{h-(x-ih)}{h} & \text{Region 3} \\ 1 - \dfrac{x-ih+y-jh}{h} & \text{Region 4} \\ \dfrac{h-(y-jh)}{h} & \text{Region 5} \\ \dfrac{x-(i-1)h}{h} & \text{Region 6} \\ 0 & \text{otherwise}\,. \end{cases}$$

If $m = n = 3$, there are 9 interior nodal points such that the stiffness matrix is a 9×9

matrix:

$$A = \begin{bmatrix} * & * & 0 & * & 0 & 0 & 0 & 0 & 0 \\ * & * & * & o & * & 0 & 0 & 0 & 0 \\ 0 & * & * & 0 & o & * & 0 & 0 & 0 \\ * & o & 0 & * & * & 0 & * & 0 & 0 \\ 0 & * & o & * & * & * & o & * & 0 \\ 0 & 0 & * & 0 & * & * & 0 & o & * \\ 0 & 0 & 0 & * & o & 0 & * & * & 0 \\ 0 & 0 & 0 & 0 & * & o & * & * & * \\ 0 & 0 & 0 & 0 & 0 & * & 0 & * & * \end{bmatrix},$$

where '$*$' stands for the nonzero entries and 'o' happens to be zero for Poisson equations. Generally, the stiffness matrix is block tri-diagonal:

$$A = \begin{bmatrix} B & -I & 0 & & & \\ -I & B & -I & & & \\ & & \cdots & \cdots & & \\ & & & \cdots & \cdots & \\ & & & -I & B & -I \\ & & & & -I & B \end{bmatrix}, \text{ where } B = \begin{bmatrix} 4 & -1 & 0 & & & \\ -1 & 4 & -1 & & & \\ & & \cdots & \cdots & & \\ & & & \cdots & \cdots & \\ & & & -1 & 4 & -1 \\ & & & & -1 & 4 \end{bmatrix}$$

and I is the identity matrix. The component of the load vector F_i can be approximated as

$$\iint_D f(x,y)\phi_i dxdy \simeq f_{ij} \iint_D \phi_i \, dxdy = h^2 f_{ij} \,,$$

so after dividing by h^2 we get the same system of equations as in the finite difference scheme, namely,

$$-\frac{U_{i-1,j} + U_{i+1,j} + U_{i,j-1} + U_{i,j+1} - 4U_{ij}}{h^2} = f_{ij} \,,$$

with the same ordering.

9.3.4 The interpolation function and error analysis

We know that the finite element solution u_h is the best solution in terms of the energy norm in the finite dimensional space V_h, *i.e.*, $\|u - u_h\|_a \le \|u - v_h\|_a$, assuming that u is the solution to the weak form. However, this does not give a quantitative estimate for the finite element solution, and we may wish to have a more precise error estimate in terms of the solution information and the mesh size h. This can be done through the interpolation function, for which an error estimate is often available from the approximation theory. Note that the solution information appears as part of the error constants in the error estimates, even though the solution is unknown. We will use the mesh parameters defined on page 224 in the discussion here.

Definition 9.7. *Given a triangulation of T_h, let $K \in T_h$ be a triangle with vertices a^i, $i = 1, 2, 3$. The interpolation function for a function $v(x, y)$ on the triangle is defined as*

$$v_I(x,y) = \sum_{i=1}^{3} v(a^i)\phi_i(x,y), \qquad (x,y) \in K\,, \tag{9.15}$$

where $\phi_i(x, y)$ is the piecewise linear function that satisfies $\phi_i(a^j) = \delta_i^j$ (with δ_i^j being the Kronecker delta). A global interpolation function is defined as

$$v_I(x,y) = \sum_{i=1}^{nnode} v(a^i)\phi_i(x,y), \qquad (x,y) \in T_h\,, \tag{9.16}$$

where a^i's are all nodal points and $\phi_i(x, y)$ is the global basis function centered at a^i.

Theorem 9.8. *If $v(x, y) \in C^2(K)$, then we have an error estimate for the interpolation function on a triangle K,*

$$\|v - v_I\|_\infty \le 2h^2 \max_{|\alpha|=2} \|D^\alpha v\|_\infty\,, \tag{9.17}$$

where h is the longest side. Furthermore, we have

$$\max_{|\alpha|=1} \|D^\alpha (v - v_I)\|_\infty \le \frac{8h^2}{\rho} \max_{|\alpha|=2} \|D^\alpha v\|_\infty\,. \tag{9.18}$$

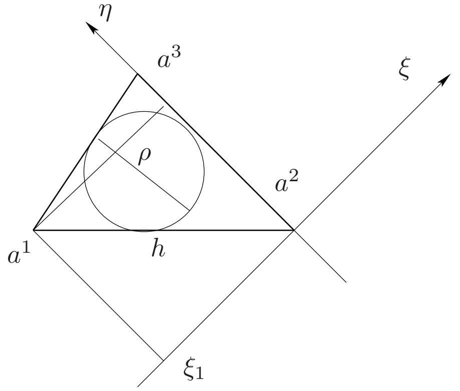

Figure 9.7. *A diagram used to prove Theorem 9.8.*

Proof: From the definition of the interpolation function and the Taylor expansion of $v(a^i)$ at (x, y), we have

$$
\begin{aligned}
v_I(x,y) &= \sum_{i=1}^{3} v(a^i)\phi_i(x,y) \\
&= \sum_{i=1}^{3} \phi_i(x,y) \left(v(x,y) + \frac{\partial v}{\partial x}(x,y)(x_i - x) + \frac{\partial v}{\partial y}(x,y)(y_i - y) + \right. \\
&\quad \left. \frac{1}{2}\frac{\partial^2 v}{\partial x^2}(\xi,\eta)(x_i-x)^2 + \frac{\partial^2 v}{\partial x \partial y}(\xi,\eta)(x_i - x)(y_i - y) + \frac{1}{2}\frac{\partial^2 v}{\partial y^2}(\xi,\eta)(y_i-y)^2 \right) \\
&= \sum_{i=1}^{3} \phi_i(x,y)v(x,y) + \sum_{i=1}^{3} \phi_i(x,y) \left(\frac{\partial v}{\partial x}(x,y)(x_i - x) + \frac{\partial v}{\partial y}(x,y)(y_i - y) \right) \\
&\qquad + R(x,y)\,,
\end{aligned}
$$

where (ξ, η) is a point in the triangle K. It is easy to show that

$$
|R(x,y)| \le 2h^2 \max_{|\alpha|=2} \|D^\alpha v\|_\infty \sum_{i=1}^{3} |\phi_i(x,y)| = 2h^2 \max_{|\alpha|=2} \|D^\alpha v\|_\infty\,,
$$

since $\phi(x, y) \ge 0$ and $\sum_{i=1}^{3} \phi_i(x, y) = 1$. If we take $v(x, y) = 1$, which is a linear function, then $\partial v/\partial x = \partial v/\partial y = 0$ and $\max_{|\alpha|=2} \|D^\alpha v\|_\infty = 0$. The interpolation is simply the function itself, since it uniquely determined by the values at the vertices of T, hence

$$
v_I(x,y) = v(x,y) = \sum_{i=1}^{3} v(a^i)\phi_i(x,y) = \sum_{i=1}^{3} \phi_i(x,y) = 1\,. \tag{9.19}
$$

If we take $v(x, y) = d_1x+d_2y$, which is also a linear function, then $\partial v/\partial x = d_1$, $\partial v/\partial y = d_2$, and $\max_{|\alpha|=2} \|D^\alpha v\|_\infty = 0$. The interpolation is again simply the function itself, since it uniquely determined by the values at the vertices of K. Thus from the previous Taylor expansion and the identity $\sum_{i=1}^{3} \phi_i(x, y) = 1$, we have

$$
v_I(x,y) = v(x,y) = v(x,y) + \sum_{i=1}^{3} \phi_i(x,y)\left(d_1(x_i - x) + d_2(y_i - y)\right) = v(x,y)\,, \tag{9.20}
$$

hence $\sum_{i=1}^{3} \phi_i(x, y) (d_1(x_i - x) + d_2(y_i - y)) = 0$ for any d_1 and d_2, *i.e.*, the linear part in the expansion is the interpolation function. Consequently, for a general function $v(x, y) \in C^2(K)$ we have

$$
v_I(x,y) = v(x,y) + R(x,y)\,, \qquad \|v - v_I\|_\infty \le 2h^2 \max_{|\alpha|=2} \|D^\alpha v\|_\infty\,,
$$

which completes the proof of the first part of the theorem.

To prove the second part concerning the error estimate for the gradient, choose a point (x_0, y_0) inside the triangle K and apply the Taylor expansion at (x_0, y_0) to get

$$
\begin{aligned}
v(x,y) &= v(x_0,y_0) + \frac{\partial v}{\partial x}(x_0,y_0)(x - x_0) + \frac{\partial v}{\partial y}(x_0,y_0)(y - y_0) + R_2(x,y), \\
&= p_1(x,y) + R_2(x,y), \qquad |R_2(x,y)| \le 2h^2 \max_{|\alpha|=2} \|D^\alpha v\|_\infty\,.
\end{aligned}
$$

Rewriting the interpolation function $v_I(x,y)$ as

$$v_I(x,y) = v(x_0,y_0) + \frac{\partial v}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial v}{\partial y}(x_0,y_0)(y-y_0) + R_1(x,y),$$

where $R_1(x,y)$ is a linear function of x and y, we have

$$v_I(a^i) = p_1(a^i) + R_1(a^i),\ i = 1,2,3\,,$$

from the definition above. On the other hand, $v_I(x,y)$ is the interpolation function, such that also

$$v_I(a^i) = v(a^i) = p_1(a^i) + R_2(a^i), \qquad i = 1,2,3\,.$$

Since $p_1(a^i) + R_1(a^i) = p_1(a^i) + R_2(a^i)$, it follows that $R_1(a^i) = R_2(a^i)$, *i.e.*, $R_1(x,y)$ is the interpolation function of $R_2(x,y)$ in the triangle K, and we have

$$R_1(x,y) = \sum_{i=1}^{3} R_2(a^i)\phi_i(x,y)\,.$$

With this equality and on differentiating

$$v_I(x,y) = v(x_0,y_0) + \frac{\partial v}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial v}{\partial y}(x_0,y_0)(y-y_0) + R_1(x,y)$$

with respect to x, we get

$$\frac{\partial v_I}{\partial x}(x,y) = \frac{\partial v}{\partial x}(x_0,y_0) + \frac{\partial R_1}{\partial x}(x,y) = \frac{\partial v}{\partial x}(x_0,y_0) + \sum_{i=1}^{3} R_2(a^i)\frac{\partial \phi_i}{\partial x}(x,y)\,.$$

Applying the Taylor expansion for $\partial v(x,y)/\partial x$ at (x_0,y_0) gives

$$\frac{\partial v}{\partial x}(x,y) = \frac{\partial v}{\partial x}(x_0,y_0) + \frac{\partial^2 v}{\partial x^2}(\bar{x},\bar{y})(x-x_0) + \frac{\partial^2 v}{\partial x \partial y}(\bar{x},\bar{y})(y-y_0)\,,$$

where $(\bar{x},\bar{y})$ is a point in the triangle K. From the last two equalities, we obtain

$$\begin{aligned}\left|\frac{\partial v}{\partial x} - \frac{\partial v_I}{\partial x}\right| &= \left|\frac{\partial^2 v}{\partial x^2}(\bar{x},\bar{y})(x-x_0) + \frac{\partial^2 v}{\partial x \partial y}(\bar{x},\bar{y})(y-y_0) - \sum_{i=1}^{3} R_2(a^i)\frac{\partial \phi_i}{\partial x}\right| \\ &\le \max_{|\alpha|=2} \|D^\alpha v\|_\infty \left(2h + 2h^2 \sum_{i=1}^{3}\left|\frac{\partial \phi_i}{\partial x}\right|\right).\end{aligned}$$

It remains to prove that $|\partial\phi_i/\partial x| \le 1/\rho$, $i = 1,2,3$. We take $i = 1$ as an illustration, and use a shift and rotation coordinate transform such that a^2a^3 is the η axis and a^2 is the origin (*cf.* Fig. 9.7):

$$\begin{aligned}\xi &= (x-x_2)\cos\theta + (y-y_2)\sin\theta\,,\\ \eta &= -(x-x_2)\sin\theta + (y-y_2)\cos\theta\,.\end{aligned}$$

Then $\phi_1(x,y) = \phi_1(\xi,\eta) = C\xi = \xi/\xi_1$, where ξ_1 is the ξ coordinate in the (ξ,η) coordinate system, such that

$$\left|\frac{\partial\phi_1}{\partial x}\right| = \left|\frac{\partial\phi_1}{\partial\xi}\cos\theta - \frac{\partial\phi_1}{\partial\eta}\sin\theta\right| \le \left|\frac{1}{\xi_1}\cos\theta\right| \le \frac{1}{|\xi_1|} \le \frac{1}{\rho}\,.$$

The same estimate applies to $\partial\phi_i/\partial x$, $i = 2, 3$, so finally we have

$$\left|\frac{\partial v}{\partial x} - \frac{\partial v_I}{\partial x}\right| \le \max_{|\alpha|=2}\|D^\alpha v\|_\infty\left(2h + \frac{6h^2}{\rho}\right) \le \frac{8h^2}{\rho}\max_{|\alpha|=2}\|D^\alpha v\|_\infty\,,$$

from the fact that $\rho \le h$. Similarly, we may obtain the same error estimate for $\partial v_I/\partial y$. □

Corollary 9.9. *Given a triangulation of T_h, we have the following error estimates for the interpolation function:*

$$\|v - v_I\|_{L^2(T_h)} \le C_1 h^2\|v\|_{H^2(T_h)}\,, \qquad \|v - v_I\|_{H^1(T_h)} \le C_2 h\|v\|_{H^2(T_h)}\,, \tag{9.21}$$

where C_1 and C_2 are constants.

9.3.5 Error estimates of the FE solution

Let us now recall the 2D Sturm-Liouville problem in a bounded domain Ω:

$$-\nabla\cdot(p(x,y)\nabla u(x,y)) + q(x,y)u(x,y) = f(x,y)\,, \quad (x,y)\in\Omega\,,$$

$$u(x,y)_{\partial\Omega} = u_0(x,y)\,,$$

where $u_0(x,y)$ is a given function, *i.e.*, a Dirichlet BC is prescribed. If we assume that $p, q \in C(\Omega)$, $p(x,y) \ge p_0 > 0$, $q(x,y) \ge 0$, $f \in L^2(\Omega)$ and the boundary $\partial\Omega$ is smooth (in C^1), then we know that the weak form has a unique solution and the energy norm $\|v\|_a$ is equivalent to the H^1 norm $\|v\|_1$. Furthermore, we know that the solution $u(x,y) \in H^2(\Omega)$. Given a triangulation T_h with a polygonal approximation to the outer boundary $\partial\Omega$, let V_h be the piecewise linear function space over the triangulation T_h, and u_h be the finite element solution. With those assumptions, we have the following theorem for the error estimates.

Theorem 9.10.

$$\|u - u_h\|_a \le C_1 h\|u\|_{H^2(T_h)}\,, \qquad \|u - u_h\|_{H^1(T_h)} \le C_2 h\|u\|_{H^2(T_h)}\,, \tag{9.22}$$

$$\|u - u_h\|_{L^2(T_h)} \le C_3 h^2\|u\|_{H^2(T_h)}\,, \qquad \|u - u_h\|_\infty \le C_4 h^2\|u\|_{H^2(T_h)}\,, \tag{9.23}$$

where C_i are constants.

Sketch of the proof. Since the finite element solution is the best solution in the energy norm, we have

$$\|u - u_h\|_a \le \|u - u_I\|_a \le \bar{C}_1\|u - u_I\|_{H^1(T_h)} \le \bar{C}_1\bar{C}_2 h\|u\|_{H^2(T_h)}\,,$$

because the energy norm is equivalent to the H^1 norm. Furthermore, because of the equivalence we get the estimate for the H^1 norm as well. The error estimates for the L^2 and L^∞ norm are not trivial in 2D, and the reader may care to consult other advanced textbooks on finite element methods.

9.4 Transforms, shape functions, and quadrature formulas

Any triangle with nonzero area can be transformed to the right-isosceles master triangle, or standard triangle $\triangle$, *cf.* the right diagram in Fig. 9.8. There are three nonzero basis functions over this standard triangle $\triangle$, namely,

$$\psi_1(\xi,\eta) = 1 - \xi - \eta\,, \tag{9.24}$$

$$\psi_2(\xi,\eta) = \xi\,, \tag{9.25}$$

$$\psi_3(\xi,\eta) = \eta\,. \tag{9.26}$$

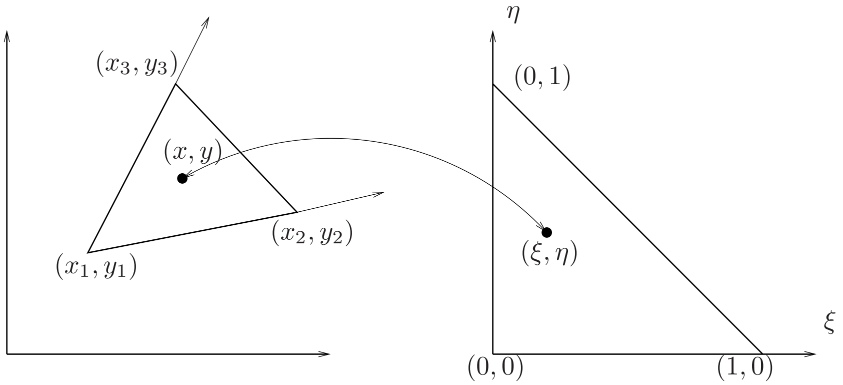

Figure 9.8. *The linear transform from an arbitrary triangle to the standard triangle (master element) and the inverse map.*

The linear transform from a triangle with vertices (x_1,y_1), (x_2,y_2) and (x_3,y_3) arranged in the counter-clockwise direction to the master triangle $\triangle$ is

$$x = \sum_{j=1}^{3} x_j\psi_j(\xi,\eta)\,, \qquad y = \sum_{j=1}^{3} y_j\psi_j(\xi,\eta)\,, \tag{9.27}$$

or

$$\xi = \frac{1}{2A_e}\Big((y_3-y_1)(x-x_1) - (x_3-x_1)(y-y_1)\Big)\,, \tag{9.28}$$

$$\eta = \frac{1}{2A_e}\Big(-(y_2-y_1)(x-x_1) + (x_2-x_1)(y-y_1)\Big)\,, \tag{9.29}$$

where A_e is the area of the triangle that can be calculated using the formula in (9.13).

9.4.1 Quadrature formulas

In the assembling process, we need to evaluate the double integrals

$$\begin{aligned}
\iint_{\Omega_e} q(x,y)\phi_i(x,y)\phi_j(x,y)\,dxdy &= \iint_{\triangle} q(\xi,\eta)\,\psi_i(\xi,\eta)\psi_j(\xi,\eta)\left|\frac{\partial(x,y)}{(\partial\xi,\eta)}\right| d\xi d\eta\,,\\
\iint_{\Omega_e} f(x,y)\phi_j(x,y)\,dxdy &= \iint_{\triangle} f(\xi,\eta)\,\psi_j(\xi,\eta)\left|\frac{\partial(x,y)}{(\partial\xi,\eta)}\right| d\xi d\eta\,,\\
\iint_{\Omega_e} p(x,y)\nabla\phi_i\cdot\nabla\phi_j\,dxdy &= \iint_{\triangle} p(\xi,\eta)\,\nabla_{(x,y)}\psi_i\cdot\nabla_{(x,y)}\psi_j\left|\frac{(\partial(x,y)}{\partial\xi,\eta)}\right| d\xi d\eta
\end{aligned}$$

in which, for example, $q(\xi,\eta)$ should really be $q(x(\xi,\eta),y(\xi,\eta)) = \bar{q}(\xi,\eta)$ and so on. For simplification of the notations, we omit the bar symbol.

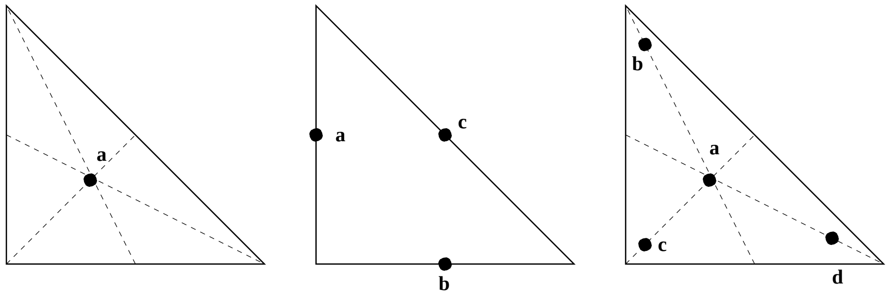

Figure 9.9. *A diagram of the quadrature formulas in 2D with one, three and four quadrature points, respectively.*

A quadrature formula has the form

$$\iint_{S_\triangle} g(\xi,\eta)d\xi d\eta = \sum_{k=1}^{L} w_k\, g(\xi_k,\eta_k)\,, \tag{9.30}$$

where $S_\triangle$ is the standard right triangle and L is the number of points involved in the quadrature. Below we list some commonly used quadrature formulas in 2D using one, three and four points. The geometry of the points are illustrated in Fig. 9.9, and the coordinates of the points and the weights are given in Table 9.1. It is noted that only the three-point quadrature formula is closed, since the three points are on the boundary of the triangle, and the other quadrature formulas are open.

9.5 Some implementation details

The procedure is essentially the same as in the 1D case, but some details are slightly different.

Table 9.1. *Quadrature points and weights corresponding to the geometry in Fig. 9.9.*

L	Points	(ξ_k, η_k)	w_k
1	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\frac{1}{2}$
3	a	$\left(0, \frac{1}{2}\right)$	$\frac{1}{6}$
	b	$\left(\frac{1}{2}, 0\right)$	$\frac{1}{6}$
	c	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$\frac{1}{6}$
4	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$-\frac{27}{96}$
	b	$\left(\frac{2}{15}, \frac{11}{15}\right)$	$\frac{25}{96}$
	c	$\left(\frac{2}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$
	d	$\left(\frac{11}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$

9.5.1 Description of a triangulation

A triangulation is determined by its elements and nodal points. We use the following notation:

- Nodal points: N_i, $(x_1, y_1), (x_2, y_2), \cdots, (x_{nnode}, y_{nnode})$, *i.e.*, we assume there are *nnode* nodal points.
- Elements: K_i, $K_1, K_2, \cdots, K_{nelem}$, *i.e.*, we assume there are *nelem* elements.
- A 2D array *nodes* is used to describe the relation between the nodal points and the elements: $nodes(3, nelem)$. The first index is the index of nodal point in an element, usually in the counter-clockwise direction, and the second index is the index of the element.

Example 9.3. *Below we show the relation between the index of the nodal points and elements, and its relations,* cf. *also Fig. 9.10.*

$$
\begin{aligned}
nodes(1,1) &= 5\,, \quad (x_5, y_5) = (0, h)\,,\\
nodes(2,1) &= 1\,, \quad (x_1, y_1) = (0, 0)\,,\\
nodes(3,1) &= 6\,, \quad (x_6, y_6) = (h, h)\,,\\
\\
nodes(1,10) &= 7\,, \quad (x_7, y_7) = (2h, h)\,,\\
nodes(2,10) &= 11\,, \quad (x_{11}, y_{11}) = (2h, 2h)\,,\\
nodes(3,10) &= 6\,, \quad (x_6, y_6) = (h, h)\,.
\end{aligned}
$$

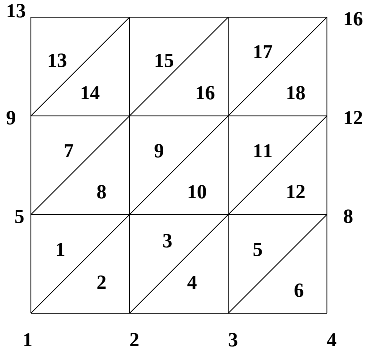

Figure 9.10. *A simple triangulation with the row-wise natural ordering.*

9.5.2 Outline of the FE algorithm using the piecewise linear basis functions

The main assembling process is the following loop.

```
for nel = 1:nelem
   i1 = nodes(1,nel);           % (x(i1),y(i1)), get nodal points
   i2 = nodes(2,nel);           % (x(i2),y(i2))
   i3 = nodes(3,nel);           % (x(i3),y(i3))
     ..............
```

- Computing the local stiffness matrix and the load vector.

```
ef=zeros(3,1);
ek = zeros(3,3);
for l=1:nq              % nq is the number of quadrature points.
  [xi_x(l),eta_y(l)] = getint,     % Get a quadrature point.
  [psi,dpsi] = shape(xi_x(l),eta_y(l));
  [x_l,y_l]  = transform,  % Get (x,y) from (\xi_x(l), \eta_y(l))
  [xk,xq,xf] = getmat(x_l,y_l);    % Get the material
```

```
                           %coefficients at the quadrature point.
  for i= 1:3
     ef(i) = ef(i) + psi(i)*xf*w(l)*J;   % J is the Jacobian
     for j=1:3
        ek(i,j)=ek(i,j)+ (T + xq*psi(i)*psi(j) )*J    % see below
     end
  end
end
```

Note that *psi* has three values corresponding to three nonzero basis functions; *dpsi* is a 3×2 matrix which contains the partial derivatives $\partial \psi_i / \partial \xi$ and $\partial \psi_i / \partial \eta$. The evaluation of T is

$$\iint_{\Omega_e} p(x,y)\, \nabla \phi_i \cdot \nabla \phi_j \, dx\, dy = \iint_{\Omega_e} p(\xi,\eta) \left(\frac{\partial \psi_i}{\partial x} \frac{\partial \psi_j}{\partial x} + \frac{\partial \psi_i}{\partial y} \frac{\partial \psi_j}{\partial y} \right) |J| \; d\xi \, d\eta \,,$$

where $J = \frac{\partial(x,y)}{\partial(\xi,\eta)}$ is the Jacobian of the transform. We need to calculate $\partial \psi_i / \partial x$ and $\partial \psi_i / \partial y$ in terms of ξ and η. Notice that

$$\frac{\partial \psi_i}{\partial x} = \frac{\partial \psi_i}{\partial \xi} \frac{\partial \xi}{\partial x} + \frac{\partial \psi_i}{\partial \eta} \frac{\partial \eta}{\partial x} \,,$$

$$\frac{\partial \psi_i}{\partial y} = \frac{\partial \psi_i}{\partial \xi} \frac{\partial \xi}{\partial y} + \frac{\partial \psi_i}{\partial \eta} \frac{\partial \eta}{\partial y} \,.$$

Since we we know that

$$\xi = \frac{1}{2A_e} \Big((y_3 - y_1)(x - x_1) - (x_3 - x_1)(y - y_1) \Big) \,,$$

$$\eta = \frac{1}{2A_e} \Big(-(y_2 - y_1)(x - x_1) + (x_2 - x_1)(y - y_1) \Big) \,,$$

we obtain those partial derivatives below,

$$\frac{\partial \xi}{\partial x} = \frac{1}{2A_e}(y_3 - y_1) \,, \quad \frac{\partial \xi}{\partial y} = -\frac{1}{2A_e}(x_3 - x_1) \,,$$

$$\frac{\partial \eta}{\partial x} = -\frac{1}{2A_e}(y_2 - y_1) \,, \quad \frac{\partial \eta}{\partial y} = \frac{1}{2A_e}(x_2 - x_1) \,.$$

- Add to the global stiffness matrix and the load vector.

```
for i= 1:3
   ig = nodes(i,nel);
   gf(ig) = gf(ig) + ef(i);
   for j=1:3
     jg  = nodes(j,nel);
     gk(ig,jg) = gk(ig,jg) + ek(i,j);
   end
end
```

- Solve the system of equations $\mathrm{gk}\, U = \mathrm{gf}$.
  - Direct method, *e.g.*, Gaussian elimination.

  - Sparse matrix technique, *e.g.*, $A = sparse(M, M)$.
  - Iterative method plus preconditioning, *e.g.*, Jacobi, Gauss-Seidel, SOR(ω), conjugate gradient methods, *etc.*

- Error analysis.
  - Construct interpolation functions.
  - Error estimates for interpolation functions.
  - Finite element solution is the best approximation in the finite element space in the energy norm.

9.6 Simplification of the FE method for Poisson equations

With constant coefficients, there is a closed form for the local stiffness matrix, in terms of the coordinates of the nodal points; so the finite element algorithm can be simplified. We now introduce the simplified finite element algorithm. A good reference is [35]: *An introduction to the finite element method with applications to non-linear problems* by R.E. White, John Wiley & Sons.

Let us consider the Poisson equation below

$$\begin{aligned} -\Delta u &= f(x,y),\ (x,y) \in \Omega\,, \\ u(x,y) &= g(x,y)\,,\ (x,y) \in \partial\Omega_1\,, \\ \frac{\partial u}{\partial n} &= 0\,,\ (x,y) \in \partial\Omega_2\,, \end{aligned}$$

where Ω is an arbitrary but bounded domain. We can use Matlab PDE Tool-box to generate a triangulation for the domain Ω.

The weak form is

$$\iint_\Omega \nabla u \cdot \nabla v \, dx\, dy = \iint_\Omega f v \, dx\, dy\,.$$

With the piecewise linear basis functions defined on a triangulation on Ω, we can derive analytic expressions for the basis functions and the entries of the local stiffness matrix.

Theorem 9.11. *Consider a triangle determined by* (x_1, y_1), (x_2, y_2) *and* (x_3, y_3). *Let*

$$a_i = x_j y_m - x_m y_j\,, \tag{9.31}$$

$$b_i = y_j - y_m\,, \tag{9.32}$$

$$c_i = x_m - x_j\,, \tag{9.33}$$

where i, j, m *is a positive permutation of* 1, 2, 3, e.g., $i = 1$, $j = 2$ *and* $m = 3$*;* $i = 2$, $j = 3$ *and* $m = 1$*; and* $i = 3$, $j = 1$ *and* $m = 2$. *Then the corresponding three nonzero basis functions are*

$$\psi_i(x,y) = \frac{a_i + b_i\, x + c_i\, y}{2\Delta}, \quad i = 1, 2, 3\,, \tag{9.34}$$

where $\psi_i(x_i,y_i)=1$, $\psi_i(x_j,y_j)=0$ *if* $i\neq j$, *and*

$$\Delta=\frac{1}{2}\,det\begin{bmatrix}1 & x_1 & y_1\\ 1 & x_2 & y_2\\ 1 & x_3 & y_3\end{bmatrix}=\pm\ \textit{area of the triangle.} \tag{9.35}$$

We prove the theorem for $\psi_1(x,y)$. Substitute a_1, b_1, and c_1 in terms of x_i and y_i in the definition of ψ_1, we have,

$$\begin{aligned}\psi_1(x,y)&=\frac{a_1+b_1x+c_1y}{2\Delta}\,,\\ &=\frac{(x_2y_3-x_3y_2)+(y_2-y_3)x+(x_3-x_2)y}{2\Delta}\,,\end{aligned}$$

so
$$\begin{aligned}\psi_1(x_2,y_2)&=\frac{(x_2y_3-x_3y_2)+(y_2-y_3)x_2+(x_3-x_2)y_2}{2\Delta}=0\,,\\ \psi_1(x_3,y_3)&=\frac{(x_2y_3-x_3y_2)+(y_2-y_3)x_3+(x_3-x_2)y_3}{2\Delta}=0\,,\\ \psi_1(x_1,y_1)&=\frac{(x_2y_3-x_3y_2)+(y_2-y_3)x_1+(x_3-x_2)y_1}{2\Delta}=\frac{2\Delta}{2\Delta}=1\,.\end{aligned}$$

We can prove the same feature for ψ_2 and ψ_3.

We also have the following theorem, which is essential for the simplified finite element method.

Theorem 9.12. *With the same notations as in Theorem 9.11, we have*

$$\iint_{\Omega_e}(\psi_1)^m(\psi_2)^n(\psi_3)^l\,dxdy=\frac{m!\,n!\,l!}{(m+n+l+2)\,!}2\Delta\,, \tag{9.36}$$

$$\iint_{\Omega_e}\nabla\psi_i\cdot\nabla\psi_j\,dxdy=\frac{b_ib_j+c_ic_j}{4\Delta}\,,$$

$$F_1^e=\iint_{\Omega_e}\psi_1f(x,y)\,dxdy\simeq f_1\frac{\Delta}{6}+f_2\frac{\Delta}{12}+f_3\frac{\Delta}{12}\,,$$

$$F_2^e=\iint_{\Omega_e}\psi_2f(x,y)\,dxdy\simeq f_1\frac{\Delta}{12}+f_2\frac{\Delta}{6}+f_3\frac{\Delta}{12}\,,$$

$$F_3^e=\iint_{\Omega_e}\psi_3f(x,y)\,dxdy\simeq f_1\frac{\Delta}{12}+f_2\frac{\Delta}{12}+f_3\frac{\Delta}{6}\,,$$

where $f_i=f(x_i,y_i)$.

The proof is straightforward since we have the analytic form for ψ_i. We approximate $f(x,y)$ using

$$f(x,y)\simeq f_1\psi_1+f_2\psi_2+f_3\psi_3, \tag{9.37}$$

and therefore

$$\begin{aligned} F_1^e &\simeq \iint_{\Omega_e} \psi_1 f(x,y)\,dxdy \\ &= f_1 \iint_{\Omega_e} \psi_1^2 dxdy + f_2 \iint_{\Omega_e} \psi_1\psi_2\,dxdy + f_3 \iint_{\Omega_e} \psi_1\psi_3\,dxdy\,. \end{aligned} \tag{9.38}$$

Note that the integrals in the last expression can be obtained from the formula (9.36). There is a negligible error from approximating $f(x,y)$ compared with the error from the finite element approximation when we seek approximate solution only in V_h space instead of $H^1(\Omega)$ space. Similarly we can get approximation F_2^e and F_3^e.

9.6.1 A pseudo-code of the simplified FE method

Assume that we have a triangulation, *e,g.*, a triangulation generated from Matlab by saving the mesh. Then we have

$p(1,1),\ p(1,2),\ \cdots,\ p(1,nnode)$ as x coordinates of the nodal points,
$p(2,1),\ p(2,2),\ \cdots,\ p(2,nnode)$ as y coordinates of the nodal points;

and the array t (the nodes in our earlier notation)

$t(1,1),\ t(1,2),\ \cdots,\ t(1,nele)$ as the index of the first node of an element,
$t(2,1),\ t(2,2),\ \cdots,\ t(2,nele)$ as the index of the second node of the element,
$t(3,1),\ t(3,2),\ \cdots,\ t(3,nele)$ as the index of the third node of the element;

and the array e to describe the nodal points on the boundary

$e(1,1),\ e(1,2),\ \cdots,\ e(1,nbc)$ as the index of the beginning node of a boundary edge,
$e(2,1),\ e(2,2),\ \cdots,\ e(2,nbc)$ as the index of the end node of the boundary edge.

A Matlab code for the simplified finite element method is listed below.

```
% Set-up: assume we have a triangulation p,e,t from Matlab PDE tool box
% already.

    [ijunk,nelem] = size(t);
    [ijunk,nnode] = size(p);

    for i=1:nelem
      nodes(1,i)=t(1,i);
      nodes(2,i)=t(2,i);
      nodes(3,i)=t(3,i);
    end

    gk=zeros(nnode,nnode);
    gf = zeros(nnode,1);
```

```
for nel = 1:nelem,   % Begin to assemble by element.

  for j=1:3,           % The coordinates of the nodes in the
    jj = nodes(j,nel);    % element.
    xx(j) = p(1,jj);
    yy(j) = p(2,jj);
  end

for nel = 1:nelem,   % Begin to assemble by element.

  for j=1:3,           % The coordinates of the nodes in the
    jj = nodes(j,nel);    % element.
    xx(j) = p(1,jj);
    yy(j) = p(2,jj);
  end

  for i=1:3,
    j = i+1 - fix((i+1)/3)*3;
    if j == 0
       j = 3;
    end
    m = i+2 - fix((i+2)/3)*3;
    if m  == 0
       m = 3;
    end

    a(i) = xx(j)*yy(m) - xx(m)*yy(j);
    b(i) = yy(j) - yy(m);
    c(i) = xx(m) - xx(j);
  end

  delta = ( c(3)*b(2) - c(2)*b(3) )/2.0;   % Area.

  for ir = 1:3,
    ii = nodes(ir,nel);
    for ic=1:3,
      ak = (b(ir)*b(ic) + c(ir)*c(ic))/(4*delta);
      jj = nodes(ic,nel);
      gk(ii,jj) = gk(ii,jj) + ak;
    end
      j = ir+1 - fix((ir+1)/3)*3;
         if j == 0
            j = 3;
         end
```

```
          m = ir+2 - fix((ir+2)/3)*3;
             if m == 0
                m = 3;
             end
          gf(ii) = gf(ii)+( f(xx(ir),yy(ir))*2.0 + f(xx(j),yy(j)) ...
                              + f(xx(m),yy(m)) )*delta/12.0;
      end

    end                  % End assembling by element.

%------------------------------------------------------
% Now deal with the Dirichlet BC

    [ijunk,npres] = size(e);
    for i=1:npres,
       xb = p(1,e(1,i));  yb=p(2,e(1,i));
       g1(i) = uexact(xb,yb);
    end

    for i=1:npres,
      nod = e(1,i);
      for k=1:nnode,
         gf(k) = gf(k) - gk(k,nod)*g1(i);
         gk(nod,k) = 0;
         gk(k,nod) = 0;
      end
         gk(nod,nod) = 1;
         gf(nod) = g1(i);
    end

    u=gk\gf;            % Solve the linear system.
    pdemesh(p,e,t,u)    % Plot the solution.

% End.
```

Example 9.4. *We test the simplified finite element method to solve a Poisson equation using the following example:*

- *Domain: Unit square with a hole,* cf. *Fig. 9.11.*
- *Exact solution:* $u(x,y) = x^2 + y^2$, *for* $f(x,y) = -4$.
- *BC: Dirichlet condition on the whole boundary.*
- *Use Matlab PDE Tool-box to generate initial mesh and then* export it*.*

Fig. 9.11 shows the domain and the mesh generated by the Matlab PDE Tool-box. The left plot in Fig. 9.12 is the mesh plot for the finite element solution, and the right plot is the error plot (the magnitude of the error is $O(h^2)$).

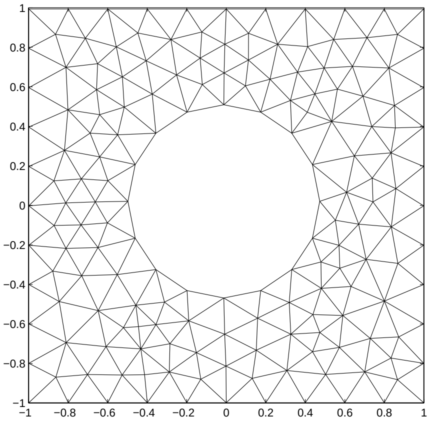

Figure 9.11. *A mesh generated from Matlab.*

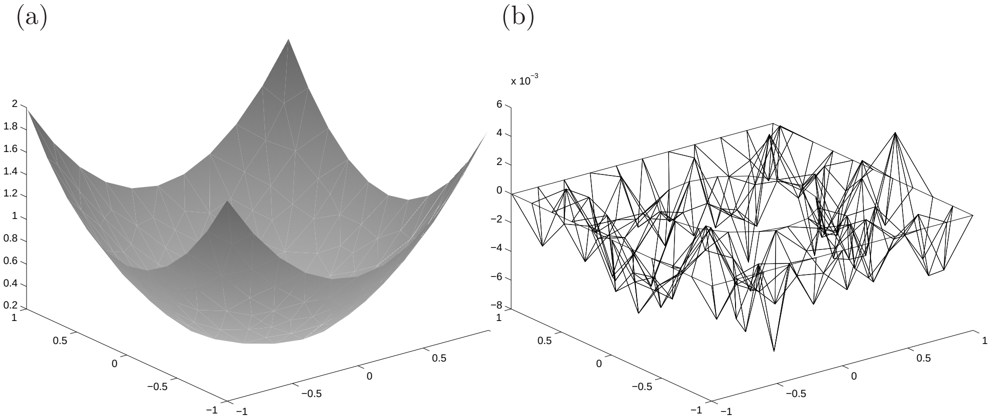

Figure 9.12. *(a) A plot of the finite element solution when $f(x, y) = -4$; (b) The corresponding error plot.*

9.7 Some FE spaces in $H^1(\Omega)$ and $H^2(\Omega)$

Given a triangulation (triangles, rectangles, quadrilaterals, *etc.*), let us construct different finite element spaces with finite dimensions. There are several reasons to do so, including:

- better accuracy of the finite element solution, with piecewise higher order polynomial basis functions; and
- to allow for higher order derivatives in higher order PDEs, *e.g.*, in solving the biharmonic equation in H^2 space.

As previously mentioned, we consider conforming piecewise polynomial finite element spaces. A set of polynomials of degree k is denoted by

$$P_k = \left\{ v(x,y) \,, \quad v(x,y) = \sum_{i,j=0}^{i+j\le k} a_{ij}\, x^i x^j \right\},$$

in the xy-plane. Below we list some examples,

$$\begin{aligned} P_1 &= \{ \, v(x,y), \quad v(x,y) = a_{00} + a_{10}x + a_{01}y \, \} \,, \\ P_2 &= \left\{ \, v(x,y), \quad v(x,y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 \, \right\} , \\ P_3 &= P_2 + \left\{ a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3 \right\} , \\ & \quad \cdots\cdots\cdots\cdots \,. \end{aligned}$$

Degree of freedom of P_k. For any fixed x^i, the possible y^j terms of in a $p_k(x,y) \in P_k$ are y^0, y^1, $\cdots$, y^{k-i}, *i.e.*, j ranges from 0 to $k-i$. Thus there are $k-i+1$ parameters for a given x^i, and the total degree of freedom is

$$\begin{aligned} \sum_{i=0}^{k} (k-i+1) &= \sum_{i=0}^{k} (k+1) - \sum_{i=0}^{k} i \\ &= (k+1)^2 - \frac{k(k+1)}{2} = \frac{(k+1)(k+2)}{2} \,. \end{aligned}$$

Some degrees of freedom for different k's are:

- 3 when $k = 1$, the linear function space P_1;
- 6 when $k = 2$, the quadratic function space P_2;
- 10 when $k = 3$, the cubic function space P_3;
- 15 when $k = 4$, the fourth order polynomials space P_4; and
- 21 when $k = 5$, the fifth order polynomials space P_5.

Regularity requirements: Generally, we cannot conclude that $v(x,y) \in C^0$ if $v(x,y) \in H^1$. However, if V_h is a finite dimensional space of piecewise polynomials, then that is indeed true. Similarly, if $v(x,y) \in H^2$ and $v(x,y)|_{K_i} \in P_k$, $\forall K_i \in T_h$, then $v(x,y) \in C^1$. The regularity requirements are important for the construction of finite element spaces.

As is quite well known, there are two ways to improve the accuracy. One way is to decrease the mesh size h, and the other is to use high order polynomial spaces P_k. If we use a P_k space on a given triangulation T_h for a linear second order elliptic PDE, the error estimates for the finite element solution u_h are

$$\|u - u_h\|_{H^1(\Omega)} \le C_1 h^k \|u\|_{H^{k+1}(\Omega)}, \; \|u - u_h\|_{L^2(\Omega)} \le C_2 h^{k+1} \|u\|_{H^{k+1}(\Omega)} \,. \qquad (9.39)$$

9.7.1 A piecewise quadratic function space

The degree of the freedom of a quadratic function on a triangle is six, so we may add thre auxiliary middle points along the three sides of the triangle.

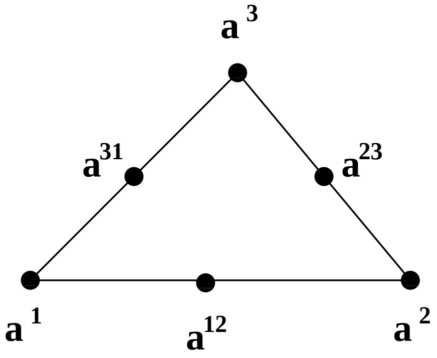

Figure 9.13. *A diagram of six points in a triangle to determine a quadratic function.*

Theorem 9.13. *Consider a triangle $K = (a^1, a^2, a^3)$, as shown in Fig. 9.13. A function $v(x, y) \in P_2(K)$ is uniquely determined by its values at*

$$v(a^i)\,,\ i = 1, 2, 3,\ \textit{and the three middle points}\ v(a^{12}),\ v(a^{23}),\ v(a^{31})\,.$$

As there are six parameters and six conditions, we expect to be able to determine the quadratic function uniquely. Highlights of the proof are as follows.

- We just need to prove the homogeneous case $v(a^i) = 0$, $v(a^{ij}) = 0$, since the right-hand side does not affect the existence and uniqueness.
- We can represent a quadratic function as a product of two linear functions, *i.e.*, $v(\mathbf{x}) = \psi_1(\mathbf{x})\omega(\mathbf{x}) = \psi_1(\mathbf{x})\psi_2(\mathbf{x})\omega_0$, with $\psi_i(\mathbf{x})$ denoting the local linear basis function such that $\psi_i(a^i) = 1$ and $\psi_i(a^j) = 0$ if $i \neq j$. Note that here we use $\mathbf{x} = (x, y)$ notation for convenience.
- It is easier to introduce a coordinate axis aligned with one of the three sides.

Proof: We introduce the new coordinates (*cf.* Fig. 9.7)

$$\xi = (x - x_2)\cos\alpha + (y - y_2)\sin\alpha\,,$$
$$\eta = -(x - x_2)\sin\alpha + (y - y_2)\cos\alpha\,,$$

such that a^2 is the origin and a^2a^3 is the η- axis. Then $v(x, y)$ can be written as

$$v(x, y) = v(x(\xi, \eta), y(\xi, \eta)) = \bar{v}(\xi, \eta) = \bar{a}_{00} + \bar{a}_{10}\xi + \bar{a}_{01}\eta + \bar{a}_{20}\xi^2 + \bar{a}_{11}\xi\eta + \bar{a}_{02}\eta^2\,.$$

Furthermore, under the new coordinates, we have

$$\psi_1(\xi, \eta) = \sigma + \beta\xi + \gamma\eta = \beta\xi\,, \quad \beta \neq 0,$$

since $\psi_1(a^2) = \psi_1(a^3) = 0$. Along the η-axis ($\xi = 0$), $\bar{v}(\xi, \eta)$ has the following form

$$\bar{v}(0, \eta) = \bar{a}_{00} + \bar{a}_{01}\eta + \bar{a}_{02}\eta^2.$$

Since $\bar v(a^2) = \bar v(a^3) = \bar v(a^{23}) = 0$, we get $\bar a_{00} = 0$, $\bar a_{01} = 0$ and $\bar a_{02} = 0$, therefore,

$$\begin{aligned} \bar v(\xi,\eta) &= \bar a_{10}\xi + \bar a_{11}\xi\eta + \bar a_{20}\xi^2 = \xi\left(\bar a_{10} + \bar a_{11}\eta + \bar a_{20}\xi\right) \\ &= \beta\xi\left(\frac{\bar a_{10}}{\beta} + \frac{\bar a_{20}}{\beta}\xi + \frac{\bar a_{11}}{\beta}\eta\right) \\ &= \psi_1(\xi,\eta)\omega(\xi,\eta)\,. \end{aligned}$$

Similarly, along the edge a^1a^3, we have

$$\begin{aligned} v(a^{13}) &= \psi_1(a^{13})\,\omega(a^{13}) = \frac{1}{2}\,\omega(a^{13}) = 0, \\ v(a^1) &= \psi_1(a^1)\omega(a^1) = \omega(a^1) = 0, \end{aligned}$$

i.e.,

$$\omega(a^{13}) = 0, \quad \omega(a^1) = 0\,.$$

By similar arguments, we conclude that

$$\omega(x,y) = \psi_2(x,y)\,\omega_0\,,$$

and hence

$$v(x,y) = \psi_1(x,y)\psi_2(x,y)\omega_0\,.$$

Using the zero value of v at a^{12}, we have

$$v(a^{12}) = \psi_1(a^{12})\,\psi_2(a^{12})\,\omega_0 = \frac{1}{2}\frac{1}{2}\,\omega_0 = 0\,,$$

so we must have $\omega_0 = 0$ and hence $v(x,y) \equiv 0$.

Continuity along the edges

Along each edge, a quadratic function $v(x,y)$ can be written as a quadratic function of one variable. For example, if the edge is represented as

$$y = ax + b \quad \text{or} \quad x = ay + b\,,$$

then

$$v(x,y) = v(x, ax+b) \quad \text{or} \quad v(x,y) = v(ay+b, y).$$

Thus the piecewise quadratic functions defined on two triangles with a common side are identical on the entire side if they have the same values at the two end points and at the mid-point of the side.

Representing quadratic basis functions using linear functions

To define quadratic basis functions with minimum compact support, we can determine the six nonzero functions using the values at three vertices and the mid-points $\mathbf{v} = (v(a^1), v(a^2), v(a^3), v(a^{12}), v(a^{23}), v(a^{13})) \in \mathbf{R}^6$. We can either take $\mathbf{v} = \mathbf{e}_i \in \mathbf{R}^6$, $i = 1, 2, \cdots, 6$ respectively, or determine a quadratic function on the triangle using the linear basis functions as stated in the following theorem.

Theorem 9.14. *A quadratic function on a triangle can be represented by*

$$\begin{aligned} v(x,y) \;\; = \;\; & \sum_{i=1}^{3} v(a^i)\phi_i(x,y)\left(2\phi_i(x,y)-1\right) \\ & \quad + \sum_{i,j=1,i<j}^{3} 4\, v(a^{ij})\,\phi_i(x,y)\,\phi_j(x,y)\,, \end{aligned} \tag{9.40}$$

where $\phi_i(x,y)$, $i = 1, 2, 3$, is one of the three linear basis function centered at one of the vertices a^i.

Proof: It is easy to verify the vertices if we substitute a^j into the right-hand side of the expression above,

$$v(a^j)\phi_j(a^j)\left(2\phi_j(a^j)-1\right) = v(a^j)\,,$$

since $\phi_i(a^j) = 0$ if $i \neq j$. We take one mid-point to verify the theorem. On substituting a^{12} into the left expression, we have

$$\begin{aligned} & v(a^1)\phi_1(a^{12})\left(2\phi_1(a^{12})-1\right) + v(a^2)\phi_2(a^{12})\left(2\phi_2(a^{12})-1\right) \\ & \quad + v(a^3)\phi_3(a^{12})\left(2\phi_3(a^{12})-1\right) + 4v(a^{12})\phi_1(a^{12})\phi_2(a^{12}) \\ & \quad + 4v(a^{13})\phi_1(a^{12})\phi_3(a^{12}) + 4v(a^{23})\phi_2(a^{12})\phi_3(a^{12}) \\ & \quad\; = v(a^{12})\,, \end{aligned}$$

since $2\phi_1(a^{12}) - 1 = 2 \times \frac{1}{2} - 1 = 0$, $2\phi_2(a^{12}) - 1 = 2 \times \frac{1}{2} - 1 = 0$, $\phi_3(a^{12}) = 0$ and $4\phi_1(a^{12})\phi_2(a^{12}) = 4 \times \frac{1}{2} \times \frac{1}{2} = 1$. Note that the local stiffness matrix is 6×6 when quadratic basis functions are used.

We have included a Matlab code of the finite element method using the quadratic finite element space over a uniform triangular mesh for solving a Poisson equation with a homogeneous (zero) Dirichlet boundary condition.

9.7.2 A cubic basis functions in $H^1 \cap C^0$

There are several ways to construct cubic basis functions in $H^1 \cap C^0$ over a triangulation, but a key consideration is to keep the continuity of the basis functions along the edges of neighboring triangles. We recall that the degree of freedom of a cubic function in 2D is ten, and one way is to add two auxiliary points along each side and one auxiliary point inside the triangle. thus together with the three vertices, we have ten points on a triangle to match the degree of the freedom (*cf.* the left diagram in Fig. 9.14). Existence and uniqueness conditions for such a cubic function are stated in the following theorem.

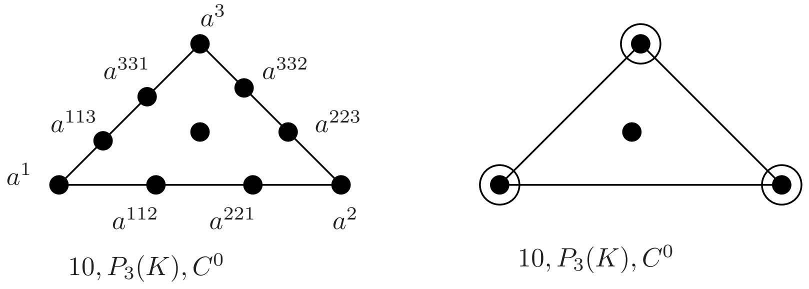

Figure 9.14. *A diagram of the freedom used to determine two different cubic basis functions in $H^1 \cap C^0$. We use the following notation: ● for function values; ○ for values of the first derivatives.*

Theorem 9.15. *A cubic function $v \in P_3(K)$ is uniquely determined by the values of*

$$v(a^i), \quad v(a^{iij}),\ i,j = 1,2,3,\ i \neq j \quad \text{and} \quad v(a^{123})\,, \tag{9.41}$$

where

$$a^{123} = \frac{1}{3}\left(a^1 + a^2 + a^3\right)\,, \quad a^{iij} = \frac{1}{3}\left(2a^i + a^j\right),\ i,j = 1,2,3,\ i \neq j\,. \tag{9.42}$$

Sketch of the proof: Similar to the quadratic case, we just need to prove that the cubic function is identically zero if $v(a^i) = v(a^{iij}) = v(a^{123}) = 0$. Again using the local coordinates where one of the sides of the triangle T is on an axis, we can write

$$v(\mathbf{x}) = C\phi_1(\mathbf{x})\phi_2(\mathbf{x})\phi_3(\mathbf{x})\,,$$

where C is a constant. Since $v(a^{123}) = C\phi_1(a^{123})\phi_2(a^{123})\phi_3(a^{123}) = 0$, we conclude that $C = 0$ since $\phi_i(a^{123}) \neq 0$, $i = 1,2,3$; and hence $v(\mathbf{x}) \equiv 0$.

With reference to the continuity along the common side of two adjacent triangles, we note that the polynomial of two variables again becomes a polynomial of one variable there, since we can substitute either x for y, or y for x from the line equations $l_0 + l_{10}x + l_{01}y = 0$. Furthermore, a cubic function of one variable is uniquely determined by the values of four distinct points.

There is another choice of cubic basis functions, using the first order derivatives at the vertices, *cf.* the right diagram in Fig. 9.14. This alternative is stated in the following theorem.

Theorem 9.16. *A cubic function $v \in P_3(K)$ is uniquely determined by the values of*

$$v(a^i), \quad \frac{\partial v}{\partial x_j}(a^i),\ i = 1,2,3,\ j = 1,2 \text{ and } i \neq j, \quad v(a^{123})\,, \tag{9.43}$$

where $\partial v/\partial x_j(a^i)$ represents $\partial v/\partial x(a^i)$ when $j = 1$ and $\partial v/\partial y(a^i)$ when $j = 2$, at the nodal point a^i.

At each vertex of the triangle, there are three degrees of freedom, namely, the function value and two first order partial derivatives; so in total there are nine degrees of freedom. An additional degree of freedom is the value at the centroid of the triangle. For the proof of the continuity, we note that on a common side of two adjacent triangles a cubic polynomial of one variable is uniquely determined by its function values at two distinct points plus the first order derivatives in the Hermite interpolation theory. The first order derivative is the tangential derivative along the common side defined as $\partial v/\partial t = \partial v/\partial x\, t_1 + \partial v/\partial y\, t_2$, where $\mathbf{t} = (t_1, t_2)$ such that $t_1^2 + t_2^2 = 1$ is the unit direction of the common side.

9.7.3 Basis functions in $H^2 \cap C^1$

To solve fourth order PDEs such as a 2D biharmonic equation

$$\Delta\,(u_{xx} + u_{yy}) = u_{xxxx} + 2u_{xxyy} + u_{yyyy} = 0\,, \tag{9.44}$$

using the finite element method, we need to construct basis functions in $H^2(\Omega) \cap C^1(\Omega)$. Since second order partial derivatives are involved in the weak form, we need to use polynomials with degree more than three. On a triangle, if the function values and partial derivatives up to second order are specified at the three vertices, the degree of freedom would be at least 18. The closest polynomial would be of degree five, as a polynomial $v(\mathbf{x}) \in P_5$ has degree of freedom 21, *cf.* the left diagram in Fig. 9.15.

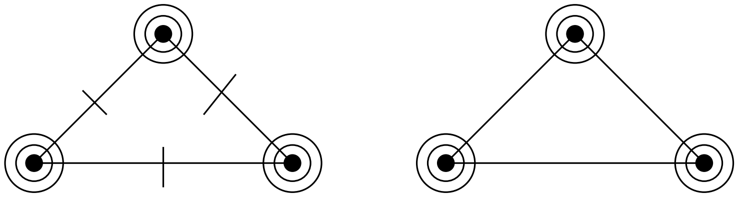

Figure 9.15. *A diagram of the freedom used to determine two different fifth order polynomial basis functions in $H^2 \cap C^1$. Left diagram, we specify $D^\alpha v(a^i)$, $0 \le \alpha \le 2$ at each vertex ($3 \times 6 = 18$) plus three normal derivatives $\partial v/\partial n(a^{ij})$ at the mid-point of the three edges. Right diagram, we can specify three independent constrains to reduce the degree of freedom, for example, $\partial v/\partial n(a^{ij}) = 0$ at the mid-point of the three edges.*

Theorem 9.17. *A quintic function $v(x, y) \in P_5(K)$ is uniquely determined by the values of*

$$D^\alpha v(a^i)\,,\ i = 1, 2, 3,\ |\alpha| \le 2, \quad \frac{\partial v}{\partial n}(a^{ij})\,,\ i, j = 1, 2, 3,\ i < j\,, \tag{9.45}$$

where $\partial v/\partial n(a^i) = n_1 \partial v/\partial x(a^i) + n_2 \partial v/\partial y(a^i)$ *represents the normal derivative of* $v(x)$ *at* a^i *and* $n = (n_1, n_2)$ $(n_1^2 + n_2^2 = 1)$ *is the outward unit normal at the boundary of the triangle.*

Sketch of the proof: We just need to show that $v(\mathbf{x}) = 0$ if $D^\alpha v(a^i) = 0$, $i = 1, 2, 3$, $|\alpha| \le 2$ and $\partial v/\partial n(a^{ij}) = 0$, $i, j = 1, 2, 3$, $i < j$. A fifth order polynomial $v(s)$ of one variable s is uniquely determined by the values of v and its derivatives $v'(s)$ and $v''(s)$ at two distinct points, so along a^2a^3, $v(\mathbf{x})$ must be zero for the given homogeneous conditions. We note that $\frac{\partial v}{\partial n}(\mathbf{x})$ is a fourth order polynomial of one variable along a^2a^3. Since all of the first and second order partial derivatives are zero at a^2 and a^3,

$$\frac{\partial v}{\partial n}(a^i) = 0\,, \qquad \frac{\partial}{\partial t}\left(\frac{\partial v}{\partial n}\right)(a^i) = 0,\ i = 2, 3\,,$$

and $\frac{\partial v}{\partial n}(a^{23}) = 0$. Here again, $\frac{\partial}{\partial t}$ is the tangential directional derivative. From the five conditions, we have $\frac{\partial v}{\partial n}(\mathbf{x}) = 0$ along a^2a^3, so we can factor $\phi_1^2(\mathbf{x})$ out of $v(\mathbf{x})$ to get

$$v(\mathbf{x}) = \phi_1^2(\mathbf{x})\, p_3(\mathbf{x})\,, \tag{9.46}$$

where $p_3(\mathbf{x}) \in P_3$. Similarly, we can factor out $\phi_2^2(\mathbf{x})$ and $\phi_3^2(\mathbf{x})$ to get

$$v(\mathbf{x}) = \phi_1^2(\mathbf{x})\, \phi_2^2(\mathbf{x})\, \phi_3^2(\mathbf{x})\, C\,, \tag{9.47}$$

where C is a constant. Consequently $C = 0$, otherwise $v(\mathbf{x})$ would be a polynomial of degree six, which contradicts that $v(\mathbf{x}) \in P_5$.

The continuity condition along a common side of two adjacent triangles in C^1 has two parts, namely, both the function and the normal derivative must be continuous. Along a common side of two adjacent triangles, a fifth order polynomial of $v(x, y)$ is actually a fifth order polynomial of one variable $v(s)$, which can be uniquely determined by the values $v(s)$, $v'(s)$ and $v''(s)$ at two distinct points. Thus the two fifth order polynomials on two adjacent triangles are identical along the common side if they have the same values of $v(s)$, $v'(s)$ and $v''(s)$ at the two shared vertices. Similarly, for the normal derivative along a common side of two adjacent triangles, we have a fourth order polynomial of one variable $\partial v/\partial n(s)$. The polynomials can be uniquely determined by the values $\partial v/\partial n(s)$ and $(d/ds)\,(\partial v/\partial n)\,(s)$ at two distinct points plus the value of a $\partial v/\partial n(s)$ at the mid-point. Thus the continuity of the normal derivative is also guaranteed.

An alternative approach is to replace the values of $\frac{\partial v}{\partial n}(a^{ij})$ at the three mid-points of the three sides by imposing another three conditions. For example, assuming that along a^2a^3 the normal derivative of the fifth order polynomial has the form

$$\frac{\partial v}{\partial n} = \widetilde{a_{00}} + \widetilde{a_{10}}\eta + \widetilde{a_{20}}\eta^2 + \widetilde{a_{30}}\eta^3 + \widetilde{a_{40}}\eta^4\,,$$

we can impose $\widetilde{a_{40}} = 0$. In other words, along the side of a^2a^3 the normal derivative of $\partial v/\partial n$ becomes a cubic polynomial of one variable. The continuity can again be guaranteed by the Hermite interpolation theory. Using this approach, the degree of the freedom is reduced to 18 from the original 21, *cf.* the right diagram in Fig. 9.15 for an illustration.

9.7.4 Finite element spaces on rectangular meshes

While triangular meshes are intensively used, particularly for arbitrary domains, meshes using rectangles are also popular for rectangular regions. Bilinear functions are often used as basis functions. Let us first consider a bilinear function space in $H^1 \cap C^0$. A bilinear function space over a rectangle K in 2D, as illustrated Fig. 9.16, is defined as

$$Q_1(K) = \Big\{ \, v(x,y), \quad v(x,y) = a_{00} + a_{10}x + a_{01}y + a_{11}xy \Big\}, \tag{9.48}$$

where $v(x,y)$ is linear in both x and y. The degree of the freedom of a bilinear function in $Q_1(K)$ is four.

Theorem 9.18. *A bilinear function $v(x,y) \in Q_1(K)$ is uniquely determined by its values at four corners.*

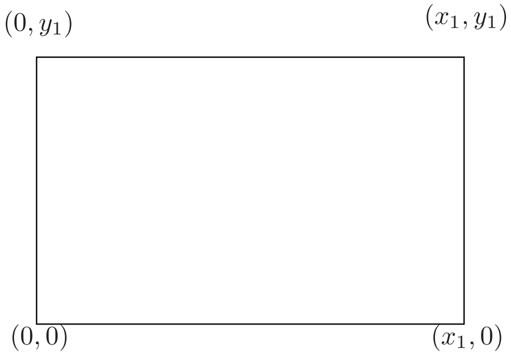

Figure 9.16. *A standard rectangle on which four bilinear basis functions can be defined.*

Proof: without loss of the generality, assume that the rectangle is determined by the four corners a^i: $(0,\ 0)$, $(x_1,\ 0)$, $(x_1,\ y_1)$ and $(0,\ y_1)$. The coefficient matrix of the linear system of algebraic equations that determines the coefficients a_{ij}, $i,j = 0,1$ is

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & x_1 & 0 & 0 \\ 1 & 0 & y_1 & 0 \\ 1 & x_1 & y_1 & x_1 y_1 \end{pmatrix},$$

with determinant $det(A) = x_1^2 y_1^2 \neq 0$ since $x_1 y_1 \neq 0$. Indeed, we have analytic expressions for the four nonzero basis functions over the rectangle, namely,

$$\phi_1(x,y) = 1 - \frac{x}{x_1} - \frac{y}{y_1} + \frac{xy}{x_1 y_1} \, , \tag{9.49}$$

$$\phi_2(x,y) = \frac{x}{x_1} - \frac{xy}{x_1 y_1} \, , \tag{9.50}$$

$$\phi_3(x,y) = \frac{xy}{x_1 y_1} \, , \tag{9.51}$$

$$\phi_4(x,y) = \frac{y}{y_1} - \frac{xy}{x_1 y_1} \, . \tag{9.52}$$

On each side of the rectangle, $v(x,y)$ is a linear function of one variable (either x or y), and uniquely determined by the values at the two corners. Thus any two basis functions along one common side of two adjacent rectangles are identical if they have the save values at the two corners, although it is hard to match the continuity condition if quadrilaterals are used instead of rectangles or cubic boxes.

A bi-quadratic function space over a rectangle is defined by

$$\begin{aligned} Q_2(K) \quad = \quad & \Big\{ v(x,y), \quad v(x,y) = a_{00} + a_{10}x + a_{01}y + a_{11}xy \\ & + a_{20}x^2 + a_{20}y^2 + a_{21}x^2y + a_{12}xy^2 + a_{22}x^2y^2 \Big\} \, . \end{aligned} \tag{9.53}$$

The degree of the freedom is nine. To construct basis functions in $H^1 \cap C^0$, as for the quadratic functions over triangles we can add four auxiliary points at the mid-points of the four sides plus a point, often the center of the in the rectangle.

In general, a bilinear function space of order k over a rectangle is defined by

$$Q_k(K) \quad = \quad \left\{ v(x,y), \quad v(x,y) = \sum_{i,j=0, i\le k, j\le k} a_{ij} x^i y^j \right\} . \tag{9.54}$$

In Fig. 9.17, we show two diagrams of finite element spaces defined on the rectangles and their degree of freedom. The first one is the bi-quadratic $Q_2(K)$ finite element in $H^1 \cap C^0$ whose degree of the freedom is nine and can be determined by the values at the marked points. The right diagram is the bi-cubic $Q_3(K)$ finite element in $H^2 \cap C^1$ whose degree of the freedom is sixteen and can be determined by the values at the marked points. The right diagram is the bi-cubic $Q_3(K)$ in $H^2 \cap C^1$ whose degree of the freedom is 16. The bi-cubic polynomial is the lowest bi-polynomial in $H^2 \cap C^1$ space. The bi-cubic function can be determined by its values, its partial derivatives $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$, and its the mixed partial derivative $\frac{\partial^2}{\partial x \partial y}$ at four vertices.

9.7.5 Some finite element spaces in 3D

In three dimensions, most commonly used meshes are tetrahedrons and cubics. In Fig. 9.18, we show two diagrams of finite element spaces defined on the tetrahedrons and their degree of freedom. The first one is the linear $T_1(K)$ finite element in $H^1 \cap C^0$ whose degree of the freedom is four and can be determined by the values at the four vertices. The right diagram is the quadratic $T_2(K)$ finite element in $H^1 \cap C^0$ whose degree of the freedom is

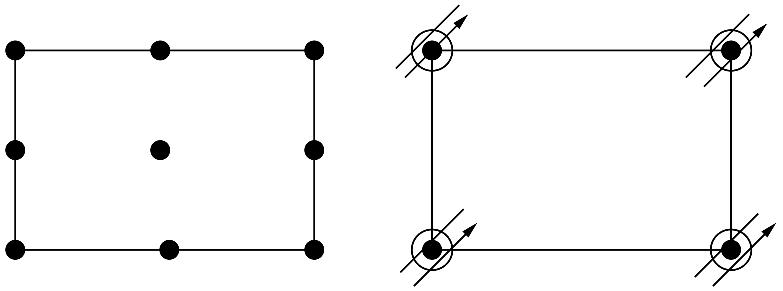

Figure 9.17. *Left diagram: $Q_2(K)$ (bi-quadratic) in $H^1 \cap C^0$ whose degree of the freedom is 9 which can be uniquely determined by the values at the marked points; Right diagram: $Q_3(K)$ (bi-cubic) in $H^2 \cap C^1$ whose degree of the freedom is 16, which can be determined by its values, first order partial derivatives marked as $/$, and mixed derivative marked as $\nearrow$, at the four corners.*

ten and can be determined by the values at the four vertices and the mid points of the six edges.

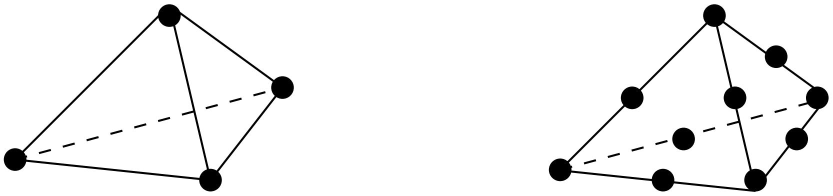

Figure 9.18. *Finite element spaces in 3D. Left diagram: $T_1(K)$ (linear) in $H^1 \cap C^0$ whose degree of the freedom is 4; Right diagram: $T_2(K)$ (quadratic) in $H^1 \cap C^0$ whose degree of the freedom is 10.*

9.7.6 *Non-conforming finite element spaces

For high order partial differential equations, such as biharmonic equations ($\Delta^2 u = f$, where Δ is the Laplacian operator $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$), in two or three dimensions, or systems of partial differential equations with certain constraints, such as divergence free condition, it is difficult to construct and verify conforming finite element spaces. Even if it possible, the degree of polynomial of the basis functions is relative high, for example, we need fifth polynomials for biharmonic equations in two space dimensions, which may lead to Gibb's oscillations near the edges. Other type of applications include non-fitted meshes or interface conditions for which it is difficult or impossible to construct finite elements that meet the conforming constraints. To overcome these difficulties, various approaches have been developed such as non-conforming finite element methods, discontinuous and weak Galerkin finite element methods. Here we mentioned some non-conforming finite element spaces that are developed in the framework of Gelerkin finite element methods.

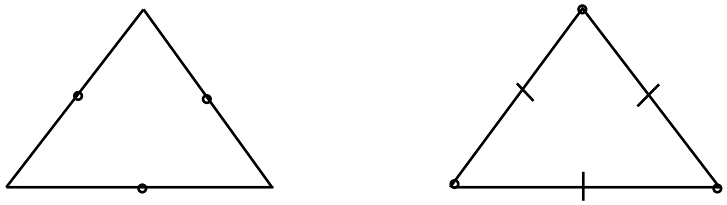

Figure 9.19. *Diagram of non-conforming finite element spaces. Left: Crouzeix-Raviart (C-R) linear nonconforming element that is determined by the values at the middle points of the edges. Right: Morley quadratic nonconforming element that is determined by the values at the vertices and the normal derivatives at the middle points of the edges.*

For triangle meshes, a non-conforming P_1 finite element space called Cronzeix-Raviart (C-R) finite element space is defined as a set of linear functions over all triangle that is continuous at the mid-points of all the edges. The basis functions can be determined by taking either unity at one middle point and zeros at other middle points of a triangle, see Fig. 9.19 (a) for an illustration. The theoretical analysis can be found in [3, 30], for example. A non-conforming Q_1 finite element space on rectangles called the Wilson element that is defined in a similar way but with the basis $\{1, x, y, xy\}$ of degree four. A rotated non-conforming Q_1 is defined in the similar way but using $\left\{1, x, y, x^2 - y^2\right\}$ as the basis. Note that, for the conforming bi-quadratic finite element space, those basis are equivalent, but it is not true for non-conforming finite element spaces anymore.

For bi-harmonic equations, a non-conforming finite element space defined on triangle meshes called the Morley finite element [30] has been developed. A Morley finite element on a triangle is defined as a quadratics functions that are determined by the values at the three vertices, and the normal derivative at the middle points of the three edges, see Fig. 9.19 (b) for an illustration. An alternative definition is to use the line integrals along the edges instead of the values at the middle points.

9.7.7 * The immersed finite element method (IFEM) for discontinuous coefficients

Following the idea of the immersed finite element method (IFEM) for one dimensional problems, we explain the IFEM for two dimensional interface problems when the coefficient $p(x, y)$ has a discontinuity across a closed smooth interface Γ. The interface Γ can be expressed as a parametric form $(X(s), Y(s)) \in C^2$, where s is a parameter, say the arc-length. The interface cuts through the domain Ω into two sub-domains Ω^+ and Ω^-, see the diagram in Fig. 9.20 (a).

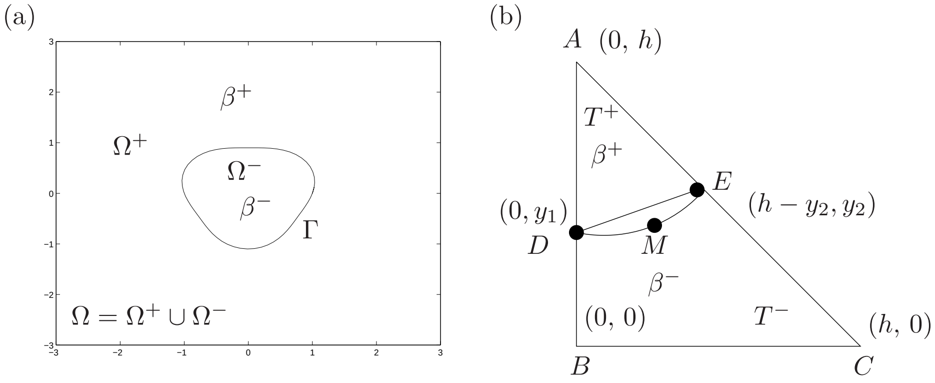

Figure 9.20. *Left figure: a configuration of a rectangular domain $\Omega = \Omega^+ \cup \Omega^-$ with an interface Γ from an IFEM test. The coefficient $p(\mathbf{x})$ may have a finite jump across the interface Γ. Right diagram: an interface triangle and the geometry after transformed to the standard right triangle.*

For simplicity, we assume that the coefficient $p(x)$ is a piecewise constant

$$p(x,y) = \begin{cases} \beta^+ & \text{if } (x,y) \in \Omega^+, \\ \beta^- & \text{if } (x,y) \in \Omega^-. \end{cases}$$

Again, across the interface Γ where the discontinuity occurs, the natural jump conditions hold

$$[u]_\Gamma = 0, \qquad \left[\beta \frac{\partial u}{\partial n}\right]_\Gamma = 0. \tag{9.55}$$

where the jump at a point $\mathbf{X} = (X, Y) \in \Gamma$ on the interface is defined as

$$[u]_\mathbf{X} = \left. u \right|_\mathbf{X}^+ - \left. u \right|_\mathbf{X}^- = \lim_{\mathbf{x} \to \mathbf{X}, \mathbf{x} \in \Omega^+} u(\mathbf{x}) - \lim_{\mathbf{x} \to \mathbf{X}, \mathbf{x} \in \Omega^-} u(\mathbf{x}),$$

and so on, where $\mathbf{x} = (x, y)$ is an interior point in the domain. Due to the discontinuity in the coefficient, the partial derivatives across the interface Γ are discontinuous although the solution and the flux (the second jump condition), are continuous. Such a problem is referred as a two-dimensional interface problem.

To solve such an interface problem using a finite element method, first a mesh needs to be chosen. One way is to use a fitted mesh as illustrated in Fig. 9.21 (a). A fitted mesh can be generated by many existing academic or commercial software packages, for example, Matlab PDE Toolbox, Freefem, Comsol, PLTMG, Triangle, Gmesh, etc. Usually there is no fixed pattern between the indexing of nodal points and elements, thus such a mesh is called an unstructured mesh. For such a mesh, the finite element method and most theoretical analysis are still valid for the interface problem.

However, it may be difficult and time consuming to generate a body fitted mesh. Such difficulty may become even severer for moving interface problems because a new mesh has to be generated at each time step, or every other time steps. A number of efficient software packages and methods that are based Cartesian meshes such as the FFT, the level set method, and others may not be applicable with a body fitted mesh.

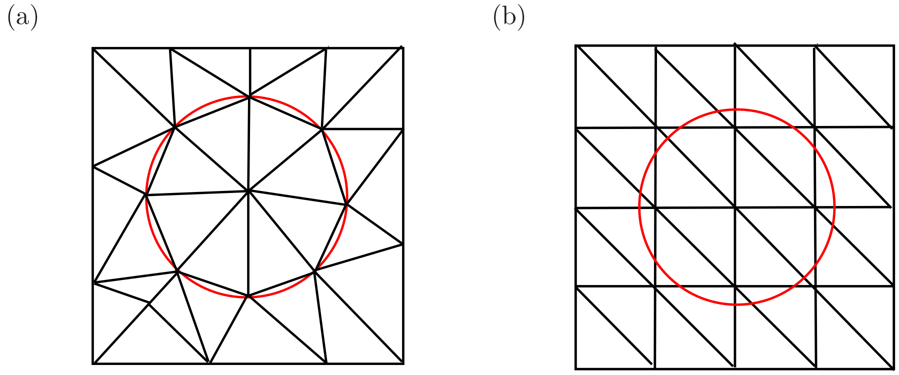

Figure 9.21. *(a). A diagram of a fitted mesh (unstructured). (b). A unfitted Cartesian mesh (structured).*

Another way to solve the interface problem is to use an unfitted mesh, for example, a uniform Cartesian mesh as illustrated in Fig. 9.21 (b). There is rich literature on unfitted meshes and related finite element methods. The non-conforming immersed finite element method (IFEM) [20] is one of early work in this direction. The idea is to enforce the natural jump conditions in triangles that the interface cuts through, which we call it an interface triangle. Without loss of generality, we consider a reference interface element T whose geometric configuration is given in Figure 9.20 (b) in which the curve between points D and E is a part of the interface. We assume that the coordinates at A, B, C, D, and E are

$$(0,h), \quad (0,0), \quad (h,0), \quad (0,y_1), \quad (h-y_2,y_2), \tag{9.56}$$

with the restriction

$$0 \le y_1 < h, \quad 0 \le y_2 < h. \tag{9.57}$$

Given the values at the three vertices we explain how to determine a piecewise linear function in the triangle that satisfies the natural jump conditions. Assume that the values at vertices A, B, and C of the element T are specified, we construct the following piecewise linear function:

$$u(\mathbf{x}) = \begin{cases} u^+(\mathbf{x}) = a_0 + a_1 x + a_2(y-h), & \text{if } \mathbf{x}=(x,y)\in T^+, \\ u^-(\mathbf{x}) = b_0 + b_1 x + b_2 y, & \text{if } \mathbf{x}=(x,y)\in T^-, \end{cases} \tag{9.58a}$$

$$u^+(D) = u^-(D), \quad u^+(E) = u^-(E), \quad \beta^+\frac{\partial u^+}{\partial n} = \beta^-\frac{\partial u^-}{\partial n}, \tag{9.58b}$$

where $\mathbf{n}$ is the unit normal direction of the line segment $\overline{DE}$. Intuitively, there are six constraints and six parameters, so we can expect the solution exists and is unique as confirmed in Theorem 8.4 in [21].

The dimension of the non-conforming IFE space is the number of interior points for a homogeneous Dirichlet boundary condition ($u|_{\partial\Omega} = 0$) as if there was no interface. The basis function centered at a node is defined as:

$$\phi_i(\mathbf{x}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise,} \end{cases} \qquad [\,\phi_i\,]_{\bar{\Gamma}} = 0, \quad \left[\beta \frac{\partial \phi_i}{\partial n}\right]_{\bar{\Gamma}} = 0, \quad \phi_i|_{\partial\Omega} = 0. \quad (9.59)$$

A basis function $\phi_i(\mathbf{x})$ is continuous in each element T except along some edges if $\mathbf{x}_i$ is a vertex of one or several interface triangles, see Figure 9.22. We use $\bar{\Gamma}$ to denote the union of the line segment that is used to approximate the interface.

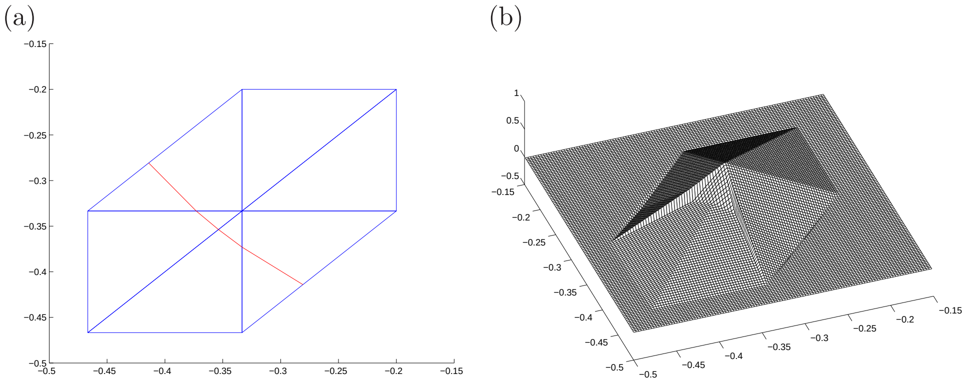

Figure 9.22. (*a*). *A standard domain of six triangles with an interface cutting through.* (*b*). *A global basis function on its support of the non-conforming immersed finite element space. The basis function has small jump across some edges.*

The basis functions in an interface triangle are continuous piecewise linear. However, it is likely discontinuous across the edges of neighboring interface triangles. Thus it is a non-conforming finite element space. Nevertheless, the corresponding non-conforming finite element method performs much better than the standard finite element method without any changes. Theoretical, an second order approximation property has been proved for the interpolation function in the L^∞ norm; and first order approximation for the partial derivatives except for the small mismatched region depicted as bounded by the points D, E, and M. It has been shown that the non-conforming IFEM is second order accurate in the L^2 norm. But its convergence order in the L^∞ norm is not so clear. Various variations, improvements, extensions, and applications can be found in the literature, particularly the symmetric and consistent IFEM that takes mismatched edge contributions into account in the variational form[15], and various penalty methods. Note that, a conforming IFEM can also be found in [23] although its implementation is not so straightforward.

9.8 The FE method for parabolic problems

We can apply the finite element method to solve time dependent problems using two different approaches. One approach is to discretize the space variables using the finite element method while discretizing the time variable using a finite difference method. This is possible if the PDE is separable. Another way is to discretize both the space and time variables using the finite element method. In this section, we briefly explain the first approach, since it is simple and easy to implement.

Let us consider the following parabolic problem in 2D,

$$\frac{\partial u}{\partial t} = \nabla \cdot (p\nabla u) + qu + f(x,y,t)\,,\ (x,y) \in \Omega,\ 0 < t \le T\,, \tag{9.60}$$

$$u(x,y,0) = 0\,,\ (x,y) \in \Omega\,, \quad \text{the initial condition}, \tag{9.61}$$

$$u(x,y,t)\Big|_{\partial\Omega} = g(x,y,t)\,, \quad \text{the boundary condition}, \tag{9.62}$$

where p, q, f and g are given functions with usual regularity assumptions. Multiplying the PDE by a test function $v(x,y) \in H^1(\Omega)$ on both sides, and then integrating over the domain, once again we obtain the weak form below,

$$\iint_\Omega u_t v\, dxdy = \iint_\Omega (qv - p\nabla u \cdot \nabla v)\, dxdy + \iint_\Omega fv\, dxdy\,, \tag{9.63}$$

where $u_t = \partial u/\partial t$. The weak form above can be simplified as

$$(u_t, v) = -a(u,v) + (f,v) \qquad \forall v \in H^1(\Omega)\,, \tag{9.64}$$

where $a(u,v) = \iint_\Omega (p\nabla u \cdot \nabla v - qv)\ dxdy$.

Given a triangulation T_h and finite element space $V_h \in H^1(\Omega) \cap C^0(\Omega)$, with $\phi_i(x,y)$, $i = 1, 2, \cdots, M$ denoting a set of basis functions for V_h, we seek the finite element solution of form

$$u_h(x,y,t) = \sum_{j=1}^{M} \alpha_j(t)\, \phi_j(x,y)\,. \tag{9.65}$$

Substituting this expression into (9.64), we obtain

$$\left(\sum_{j=1}^{M} \alpha_i'(t)\phi_i(x,y)\ ,\ v_h \right) = -a \left(\sum_{j=1}^{M} \alpha_i(t)\phi_i(x,y),\ v_h \right) + (f, v_h), \tag{9.66}$$

and then take $v_h(x,y) = \phi_i(x,y)$ for $i = 1, 2, \cdots, M$ to get the linear system of ordinary differential equations in the $\alpha_j(t)$:

$$\begin{bmatrix} (\phi_1,\phi_1) & (\phi_1,\phi_2) & \cdots & (\phi_1,\phi_M) \\ (\phi_2,\phi_1) & (\phi_2,\phi_2) & \cdots & (\phi_2,\phi_M) \\ \vdots & \vdots & \vdots & \vdots \\ (\phi_M,\phi_1) & (\phi_M,\phi_2) & \cdots & (\phi_M,\phi_M) \end{bmatrix} \begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \\ \vdots \\ \alpha_M'(t) \end{bmatrix} =$$

$$\begin{bmatrix} (f,\phi_1) \\ (f,\phi_2) \\ \vdots \\ (f,\phi_M) \end{bmatrix} - \begin{bmatrix} a(\phi_1,\phi_1) & a(\phi_1,\phi_2) & \cdots & a(\phi_1,\phi_M) \\ a(\phi_2,\phi_1) & a(\phi_2,\phi_2) & \cdots & a(\phi_2,\phi_M) \\ \vdots & \vdots & \vdots & \vdots \\ a(\phi_M,\phi_1) & a(\phi_M,\phi_2) & \cdots & a(\phi_M,\phi_M) \end{bmatrix} \begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \vdots \\ \alpha_M(t) \end{bmatrix}.$$

The corresponding problem can therefore be expressed as

$$B\frac{d\vec{\alpha}}{dt} + A\vec{\alpha} = F, \qquad \alpha_i(0) = u(N_i, 0),\ i = 1, 2, \cdots, M. \tag{9.67}$$

There are many methods to solve the above problem involving the system of first order ODEs. We can use the ODE Suite in Matlab, but note that the ODE system is known to be very stiff. We can also use finite difference methods that march in time, since we know the initial condition on $\vec{\alpha}(0)$. Thus with the solution $\vec{\alpha}^k$ at time t^k, we compute the solution $\vec{\alpha}^{k+1}$ at the time $t^{k+1} = t^k + \Delta t$ for $k = 0, 1, 2, \cdots$.

Explicit Euler method

If the forward finite difference approximation is invoked, we have

$$B\frac{\vec{\alpha}^{k+1} - \vec{\alpha}^k}{\Delta t} + A\vec{\alpha}^k = F^k, \tag{9.68}$$

$$\text{or} \quad \vec{\alpha}^{k+1} = \vec{\alpha}^k + \Delta t B^{-1}\left(F^k - A\vec{\alpha}^k\right). \tag{9.69}$$

Since B is a non-singular tridiagonal matrix, its inverse and hence $B^{-1}\left(F^k - A\vec{\alpha}^k\right)$ can be computed. However, the CFL (Courant-Friedrichs-Lewy) condition

$$\Delta t \le Ch^2\,, \tag{9.70}$$

must be satisfied to ensure the numerical stability. Thus we need to use a rather small time step.

Implicit Euler method

If we invoke the backward finite difference approximation, we get

$$B\frac{\vec{\alpha}^{k+1} - \vec{\alpha}^k}{\Delta t} + A\vec{\alpha}^{k+1} = F^{k+1}, \tag{9.71}$$

$$\text{or} \quad (B + \Delta t A)\,\vec{\alpha}^{k+1} = B\vec{\alpha}^k + \Delta t F^{k+1} \tag{9.72}$$

then there is no constraint on the time step and thus the method is called unconditionally stable. However, we need to solve a linear system of equations similar to that for an elliptic PDE at each time step.

The Crank-Nicolson method

Both of the above Euler methods are first order accurate in time and second order in space, *i.e.*, the error in computing $\vec{\alpha}$ is $O(\Delta t + h^2)$. We obtain a second order scheme in time as well in space if we use the central finite difference approximation at $t^{k+\frac{1}{2}}$:

$$B\frac{\vec{\alpha}^{k+1} - \vec{\alpha}^k}{\Delta t} + \frac{1}{2}A\left(\vec{\alpha}^{k+1} + \vec{\alpha}^k\right) = \frac{1}{2}\left(F^{k+1} + F^k\right), \tag{9.73}$$

$$\text{or} \quad \left(B + \frac{1}{2}\Delta t A\right)\vec{\alpha}^{k+1} = \left(B - \frac{1}{2}\Delta t A\right)\vec{\alpha}^k + \frac{1}{2}\Delta t\left(F^{k+1} + F^k\right). \tag{9.74}$$

This Crank-Nicolson method is second order accurate in both time and space, and it is unconditionally stable for linear parabolic PDEs. The challenge is to solve the resulting linear system of equations efficiently.

9.9 Exercises

1. Derive the weak form for the following problem:

$$
\begin{aligned}
&-\nabla\cdot(p(x,y)\nabla u(x,y)) + q(x,y)u(x,y) = f(x,y)\ ,\ (x,y)\in\Omega\,,\\
&u(x,y)=0\ ,\ (x,y)\in\partial\Omega_1\,, \qquad \frac{\partial u}{\partial n} = g(x,y),\ (x,y)\in\partial\Omega_2\,,\\
&a(x,y)u(x,y) + \frac{\partial u}{\partial n} = c(x,y)\,,\ (x,y)\in\partial\Omega_3\,,
\end{aligned}
$$

   where $q(x,y)\ge q_{min}>0$, $\partial\Omega_1\cup\partial\Omega_1\cup\partial\Omega_3=\partial\Omega$ and $\partial\Omega_i\cap\partial\Omega_j=\phi$. Provide necessary conditions so that the weak form has a unique solution. Show your proof using the Lax-Milgram Lemma but without using the Poincaré inequality.

2. Derive the weak form and appropriate space for the following problem involving the bi-harmonic equation:

$$
\begin{aligned}
\Delta\Delta u(x,y) &= f(x,y)\,,\ (x,y)\in\Omega\,,\\
u(x,y)|_{\partial\Omega} &= 0\,, \quad u_n(x,y)|_{\partial\Omega}=0\,.
\end{aligned}
$$

   What kind of basis function do you suggest, to solve this problem numerically?

   Hint: Use Green's theorem twice.

3. Consider the problem involving the Poisson equation:

$$
\begin{aligned}
-\Delta u(x,y) &= 1\ ,\ (x,y)\in\Omega\,,\\
u(x,y)|_{\partial\Omega} &= 0\,,
\end{aligned}
$$

   where Ω is the unit square. Using a uniform triangulation, derive the stiffness matrix and the load vector for $N=2$; in particular, take $h=1/3$ and consider

   (a) the nodal points ordered as $(1/3,1/3)$, $(2/3,1/3)$; $(1/3,2/3)$, and $(2/3,2/3)$; and

   (b) the nodal points ordered as $((1/3,2/3)$, $(2/3,1/3)$; $(1/3,1/3)$, and $(2/3,2/3)$.

   Write down each basis function explicitly.

4. Use the Matlab PDE toolbox to solve the following problem involving a parabolic equation for $u(x,y,t)$, and make relevant plots:

$$
\begin{aligned}
u_t &= u_{xx}+u_{yy}, \quad (x,y)\in(-1\ \ 1)\times(-1\ \ 1)\,,\\
u(x,y,0) &= 0\,.
\end{aligned}
$$

   The geometry and the BC are defined in Fig. 9.23. Show some plots of the solution (mesh, contour, *etc.*).

5. Download the Matlab source code *f.m, my_assemb.m, uexact.m* from

   `http://www4.ncsu.edu/~zhilin/FD\_FEM\_Book.`

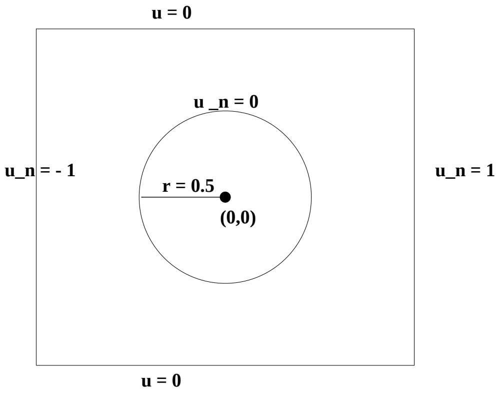

Figure 9.23. *Diagram for Exercise 2.*

Use the exported mesh of the geometry generated from Matlab, see Fig. 9.23 to solve the Poisson equation

$$-(u_{xx} + u_{yy}) = f(x, y)\,,$$

subject to the Dirichlet BC corresponding to the exact solution

$$u(x, y) = \frac{1}{4}\left(x^2 + y^4\right)\sin \pi x \cos 4\pi y\,.$$

Plot the solution and the error.

6. Modify the Matlab code to consider the generalized Helmholtz equation

$$-(u_{xx} + u_{yy}) + \lambda\, u = f(x, y)\,.$$

Test your code with $\lambda = 1$, with reference to the same solution, geometry and BC as in Problem 5. Adjust $f(x, y)$ to check the errors.

7. Modify the Matlab code to consider the Poisson equation

$$-\nabla\left(p(x, y)\cdot \nabla u(x, y)\right) = f(x, y)\ , \tag{9.75}$$

using a third order quadrature formula. Choose two examples with nonlinear $p(x, y)$ and $u(x, y)$ to show that your code is bug-free. Plot the solutions and the errors.

9.9.1 Matlab PDE-Toolbox lab exercises

Purpose: to learn the Matlab *Partial Differential Equation* toolbox.

Use the Matlab PDE toolbox to solve some typical second order PDE on some regions with various BC. Visualize the mesh triangulation and the solutions, and export the triangulation.

Reference: Partial Differential Equation Toolbox, MathWorks.

Test Problems

1. Poisson equation on a unit circle:

$$\begin{aligned} -\Delta u &= 1\,, \quad x^2+y^2<1\,,\\ u|_{\partial\Omega} &= 0\,, \quad x\le 0\,,\\ u_n|_{\partial\Omega} &= 1\,, \quad x>0\,. \end{aligned}$$

2. Wave equation on a unit square $x \in [-1,1] \times y \in [-1,1]$:

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} &= \Delta u\,,\\ u(x,y,0) &= \arctan\left(\cos\frac{\pi x}{2}\right),\\ u_t(x,y,0) &= 3\sin(\pi x)e^{\sin(\pi y/2)}\,, \end{aligned}$$

$$u = 0 \text{ at } x=-1 \text{ and } x=1\,, \quad u_n = 0 \text{ at } y=-1 \text{ and } y=1\,.$$

3. Eigenvalue problem on an L-shape:

$$-\Delta u = \lambda u, \quad u = 0 \text{ on } \partial\Omega.$$

The domain is the L-shape with corners (0,0) , $(-1,0)$, $(-1,-1)$, $(1,-1)$, (1,1) , and (0,1).

4. The heat equation:

$$\frac{\partial u}{\partial t} = \Delta u\,.$$

The domain is the rectangle $[-0.5 \quad 0.5] \times [-0.8 \quad 0.8]$, with a rectangular cavity $[-0.05 \quad 0.05] \times [-0.4 \quad 0.4]$; and the BC are:

- $u = 100$ on the left-hand side;
- $u = -10$ on the right-hand side; and
- $u_n = 0$ on all other boundaries.

5. Download the Matlab source code 2D.rar from

`http://www4.ncsu.edu/~zhilin/FD\_FEM\_Book`

Export the mesh of the last test problem from Matlab and run assemb.m to solve the example.

General Procedure

- Draw the geometry;

- define the BC;
- define the PDE;
- define the initial conditions if necessary;
- solve the PDE;
- plot the solution;
- refine the mesh if necessary; and
- save and quit.

Bibliography

[1] J. Adams, P. Swarztrauber, and R. Sweet. Fishpack: Efficient Fortran subprograms for the solution of separable elliptic partial differential equations. http://www.netlib.org/fishpack/.

[2] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, 3rd edition, 2007.

[3] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer New York, 2002.

[4] R. L. Burden and J. D. Faires. *Numerical Analysis*. 2010, PWS-Kent Publ. Co. 2006, Brooks &. Cool, 9 ed, 2010.

[5] D. Calhoun. A Cartesian grid method for solving the streamfunction-vorticity equation in irregular regions. *J. Comput. Phys.*, 176:231–275, 2002.

[6] G. F. Carey and J. T. Oden. *Finite Element, I-V*. Prentice-Hall, Inc, Englewood Cliffs, 1983.

[7] A. J. Chorin. Numerical solution of the Navier-Stokes equations. *Math. Comp.*, 22:745–762, 1968.

[8] P. G. Ciarlet. *The finite element method for elliptic problems*. North Holland, 1978, and SIAM Classic in Applied Mathematics 40, 2002.

[9] D. De Zeeuw. Matrix-dependent prolongations and restrictions in a blackbox multigrid solver. *J. Comput. Appl. Math.*, 33:1–27, 1990.

[10] L. C. Evans. *Partial Differential Equations*. AMS, 1998.

[11] G. Golub and C. Van Loan. *Matrix computations*. The Johns Hopkins University Press, 2nd ed., 1989.

[12] H. Huang and Z. Li. Convergence analysis of the immersed interface method. *IMA J. Numer. Anal.*, 19:583–608, 1999.

[13] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge, 2008.

[14] Jr. J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM, 1996.

[15] Haifeng Ji, Jinru Chen, and Zhilin Li. A symmetric and consistent immersed finite element method for interface problems. *J. Sci. Comput.*, 61(3):533–557, 2014.

[16] C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987.

[17] R. J. LeVeque. Clawpack and Amrclaw – Software for high-resolution Godunov methods. 4-th Intl. Conf. on Wave Propagation, Golden, Colorado, 1998.

[18] R. J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations, Steady State and Time Dependent Problems*. SIAM, 2007.

[19] Z. Li. *The Immersed Interface Method — A Numerical Approach for Partial Differential Equations with Interfaces*. PhD thesis, University of Washington, 1994.

[20] Z. Li. The immersed interface method using a finite element formulation. *Applied Numer. Math.*, 27:253–267, 1998.

[21] Z. Li and K. Ito. *The Immersed Interface Method – Numerical Solutions of PDEs Involving Interfaces and Irregular Domains*. SIAM Frontier Series in Applied mathematics, FR33, 2006.

[22] Z. Li and M-C. Lai. The immersed interface method for the Navier-Stokes equations with singular forces. *J. Comput. Phys.*, 171:822–842, 2001.

[23] Z. Li, T. Lin, and X. Wu. New Cartesian grid methods for interface problem using finite element formulation. *Numer. Math.*, 96:61–98, 2003.

[24] Z. Li and C. Wang. A fast finite difference method for solving Navier-Stokes equations on irregular domains. *J. of Commu. in Math. Sci.*, 1:180–196, 2003.

[25] M. Minion. A projection method for locally refined grids. *J. Comput. Phys.*, 127:158–178, 1996.

[26] K. W. Morton and D. F. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge press, 1995.

[27] J. W. Ruge and K. Stuben. Algebraic multigrid. In S. F. McCormick, editor, *Multigrid Method*. SIAM, Philadelphia, 1987.

[28] Y. Saad. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Stat. Comput.*, 7:856–869, 1986.

[29] Lawrence F. Shampine and Mark W. Reichelt. The MATLAB ODE suite. *SIAM J. Sci. Comput.*, 18(1):1–22, 1997.

[30] Z-C. Shi. Nonconforming finite element methods. *Journal of Computational and Applied Mathematics*, 149:221âĂŞ225, 2002.

[31] G. Strang and G. J. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, 1973.

[32] J. C. Strikwerda. *Finite Difference Scheme and Partial Differential Equations*. Wadsworth & Brooks, 1989.

[33] K. Stüben. Algebraic multigrid (AMG): An introduction with applications. *Gesellschaft für Mathematik und Datenveranbeitung*, Nr. 70, 1999.

[34] J. W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Springer New York, 1995.

[35] R. E. White. *An introduction to the FE method with applications to non-linear problems*. John Wiley & Sons, 1985.

Index