Finite Element Methods
Fall 2009
Dr. J. Howell*

Chris Almost†

Contents

Contents **1**

1 The Finite Element Method **3**
- 1.1 Model problem . . . 3
- 1.2 The finite difference method . . . 3
- 1.3 Finite element methods . . . 4
- 1.4 Weak statement of the model problem . . . 6
- 1.5 Lax-Milgram theorem and Poincaré lemma . . . 7
- 1.6 Well-posedness of weak approach . . . 8
- 1.7 From the continuous to the discrete . . . 8
- 1.8 Automating the computations . . . 10
- 1.9 Numerical integration (quadrature) . . . 11
- 1.10 Boundary Conditions . . . 12
- 1.11 Main Program . . . 12
- 1.12 Accuracy . . . 16
- 1.13 Rates of convergence . . . 17
- 1.14 Computing norms and errors . . . 18
- 1.15 Non-homogeneous boundary conditions . . . 18
- 1.16 Higher order elements . . . 19

*howell4@andrew.cmu.edu
†cdalmost@cmu.edu

2 Finite Element Approximation of Elliptic Problems **21**
2.1 Review 21
2.2 Model problem: the diffusion equation 22
2.3 Well-posedness 23
2.4 Approximation theory 26
2.5 Well-posedness of the model problem 27
2.6 Finite elements 28
2.7 Simplicial finite elements 29
2.8 Piecewise linear functions on triangles, $n = 2$ 30
2.9 Missed 3 lectures 31
2.10 Finite element meshes 31
2.11 Rectangular elements 32
2.12 Interpolation in Sobolev spaces 33

3 Parabolic problems **39**
3.1 Introduction 39
3.2 Semi-discrete formulation 40
3.3 Discretization in space and time 42
3.4 Missed 3 lectures 42

Index **43**

1 The Finite Element Method

1.1 Model problem

Consider the following two point boundary value problem. Given $f : [0,1] \to \mathbb{R}$, find $u : [0,1] \to \mathbb{R}$ satisfying $-u''(x) = f(x)$ for all $x \in (0,1)$ and $u(0) = u'(1) = 0$. This is the *strong form* (S) of the problem. It describes the heat distribution on a metal bar of unit length with the temperature fixed at the left end, insulated at the right end, and heat along the bar supplied by f.

1.2 The finite difference method

Discretize the problem with $N+2$ points $0 = x_0 < x_1 < \cdots < x_{N+1} = 1$, the *mesh points*. Let $h_i = x_{i+1} - x_i$, the *mesh spacing* be constant for this example (so $h_i = h = \frac{1}{N+1}$ and $x_i = ih$). Let $u_i = u(x_i)$ and expand with Taylor's Theorem.

$$u(x+h) = u(x) + hu'(x) + \frac{h^2}{2}u''(x) + \frac{h^3}{6}u'''(x) + O(h^4)$$

$$u(x-h) = u(x) - hu'(x) + \frac{h^2}{2}u''(x) - \frac{h^3}{6}u'''(x) + O(h^4)$$

Adding, $u''(x) = \frac{u(x+h)-2u(x)+u(x-h)}{h^2} + O(h^2)$. If we set $u''_i = \frac{u_{i+1}-2u_i+u_{i-1}}{h^2}$ then we have an $O(h^2)$ approximation for u'' on this mesh. Rearranging and applying the equation, $-u_{i+1} + 2u_i - u_{i-1} = h^2 f_i$, $i = 1, \ldots, N$, where $F_i = f(x_i)$. By the first boundary condition $u_0 = 0$.

Subtracting, $u(x+h) - u(x-h) = 2hu'(x) + O(h^3)$. Applying the approximation at $i = N+1$ we see, by the second boundary condition, we can use u_N for u_{N+2} wherever the latter appears.

This yields the linear system

$$\begin{bmatrix} 2 & -1 & 0 & \ldots & 0 \\ -1 & 2 & -1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & -1 & 2 & -1 \\ 0 & \ldots & 0 & -2 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \\ u_{N+1} \end{bmatrix} = h^2 \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \\ F_{N+1} \end{bmatrix}$$

A Python implementation of the above is provided below.

```
from numpy import *
from scipy.sparse import lil_matrix
from scipy.sparse.linalg import spsolve

N = 1000
h = 1.0/(N+1)
f = lambda x: sin(4*x)
```

```
x = linspace(0.0, 1.0, N+2)
F = f(x)
A = lil_matrix((N+1, N+1))
A.setdiag([2]*(N+1))
A.setdiag([-1]*N, 1)
A.setdiag([-1]*N, -1)
A[N, N-1] = -2
u = zeros(N+2)
u[1:N+2] = spsolve(A.tocsr(), h*h*F[1:N+2])

with open('output.dat', 'w') as o:
    for t in zip(x, u, F):
        o.write('%f %f %f\n' % t)
```

1.3 Finite element methods

Finite element methods are based on "weak" or "variational" statements of the problem. There are two main approaches. The first is the *minimization approach* (M), or *Rayleigh-Ritz approach*. Define

$$F(v) := \frac{1}{2}\int_0^1 (v')^2 dx - \int_0^1 f v dx.$$

We wish to find u in some appropriate space such that $F(u) \leq F(v)$ for all v in that space. The second approach is the *weak approach* (W), or *Galerkin approach*. We wish to find u in some space such that $\int_0^1 u'v' dx = \int_0^1 f v dx$ for all v in that space. The choice of space is what takes care of the boundary conditions.

1.3.1 Theorem. *Let $U := \{u \in C[0,1] \mid u'$ is piecewise continuous on $[0,1]$ and $u(0) = 0\}$. If the strong form of the model problem has a solution then the weak approach has a solution on U, and the weak and minimization approaches are equivalent.*

PROOF: Notice that U is a vector space.
$(S) \implies (W)$: Suppose that u satisfies (S). Then for any $v \in U$,

$$\int_0^1 -u'' v dx = \int_0^1 f v dx.$$

By integration-by-parts,

$$-\int_0^1 u'' v dx = \int_0^1 u'v' dx - u'(1)v(1) + u'(0)v(0) = \int_0^1 u'v' dx,$$

so u satisfies (W).

$(W) \iff (M)$: Let u be a solution to (W). Let $v \in U$ and set $w = v - u$, so $v = u + w$ and $w \in U$. Then

$$\begin{aligned} F(v) &= F(u+w) \\ &= \frac{1}{2}\int_0^1 ((u+w)')^2 dx - \int_0^1 (u+w) f\, dx \\ &= \underbrace{\frac{1}{2}\int_0^1 (u')^2 dx - \int_0^1 f u dx}_{F(u)} + \underbrace{\int_0^1 u'w' dx - \int_0^1 f w dx}_{0 \text{ since } w \in U} + \underbrace{\frac{1}{2}\int_0^1 (w')^2 dx}_{\geq 0} \\ &\geq F(u) \end{aligned}$$

Conversely, assume that $F(u) \leq F(v)$ for all $v \in U$. Let $v \in U$ and for $\varepsilon > 0$ define

$$\begin{aligned} G(\varepsilon) := F(u + \varepsilon v) = \frac{1}{2}\int_0^1 (u')^2 dx + \varepsilon \int_0^1 u'v' dx \\ + \frac{\varepsilon^2}{2}\int_0^1 (v')^2 dx - \int_0^1 f u dx - \varepsilon \int_0^1 f v dx \end{aligned}$$

Then G is differentiable and has a minimum at $\varepsilon = 0$, so $G'(0) = 0$. But $G'(0) = \int_0^1 u'v' dx - \int_0^1 f v dx$, so u satisfies (W). □

1.3.2 Proposition. *Solutions to the weak approach are unique.*

PROOF: Suppose that $u_1, u_2 \in U$ both solve (W). Then $\int_0^1 (u_1' - u_2')^2 v' dx = 0$ for all $v \in U$, so taking $v = u_1 - u_2 \in U$ it follows that $u_1' = u_2'$ a.e. It follows that $u_1 = u_2$ since they are continuous functions. □

But when does a solution to the weak approach solve the strong problem? If u is a solution to (W) then $\int_0^1 u'v' dx = \int_0^1 f v dx$ for all $v \in U$. If u'' exists and is continuous, then $\int_0^1 u'v' dx = \int_0^1 -u'' v dx = \int_0^1 f v dx$, so $\int_0^1 (u'' + f) v dx = 0$ for all $v \in U$. In particular, we may conclude $-u'' = f$ when $u'' + f$ is continuous (and possibly under other, weaker, conditions on f).

We will concern ourselves mostly with the weak approach. To write down the solution u (in the numerical sense of implementing the function u on a computer), it suffices to choose an appropriate finite-dimensional subspace of $U_h \leq U$ and consider the problem (W_h) on U_h: given $f : [0,1] \to \mathbb{R}$, find $u_h \in U_h$ such that $\int_0^1 u_h' v_h' dx = \int_0^1 f v_h dx$ for all $v_h \in U_h$.

The following flow-chart illustrates the general method we will use in this class to solve problems stated in a strong form.

Strong Problem

↓ multiply and integrate

Weak Problem (Continuous)

↓ choose appropriate subspaces

Weak Problem (Discretized)

↓ interpolate and approximate

Finite Elements

↓ numerical quadrature

System of Equations

↓ numerical linear algebra

Approximate Solution

1.4 Weak statement of the model problem

Define $L^2(0,1)$ in the usual way, and note that it is a Hilbert space with inner product $(f,g)_{L^2(0,1)} := \int_0^1 fg\,dx$. Let $H^1(0,1) := \{f \in L^2(0,1) \mid f' \in L^2(0,1)\}$, the Sobolev space $W^{1,2}(0,1)$. It too is a Hilbert space, with inner product

$$(f,g)_{H^2(0,1)} := \int_0^1 (fg + f'g')\,dx.$$

The weak form of the model problem can be stated as follows. Given $f \in L^2(0,1)$, find $u \in \{u \in H^1(0,1) \mid u(0) = 0\} =: U$ satisfying, for all $v \in U$,

$$\int_0^1 u'v'\,dx = \int_0^1 fv\,dx.$$

In general, the problem is to find $u \in U$ such that $a(u,v) = F(v)$ for all $v \in V$, where $a : U \times V \to \mathbb{R}$ is a bilinear form and $F : U \to \mathbb{R}$ is a linear functional, for some spaces U and V. In the model problem $a(u,v) = \int_0^1 u'v'\,dx$, $F(v) = \int_0^1 fv\,dx$, and $V = U = \{u \in H^1(0,1) \mid u(0) = 0\}$.

We can incorporate Dirichlet boundary conditions into the definition of U, but we cannot do this for Neumann boundary conditions. Dr. Howell offers the following cryptic statement, "The fact that u will have degrees of freedom on the Neumann portion of the boundary will take care of the Neumann boundary condition."

1.5 Lax-Milgram theorem and Poincaré lemma

There are some extra conditions on a and F in the problem above required for the problem to be well-defined in general. For now we will concern ourselves with the case when $U = V$.

1.5.1 Theorem. *Let U be a Hilbert space, $a : U \times U \to \mathbb{R}$ be a bilinear form, and let $F : U \to \mathbb{R}$ be a linear functional, such that*

Continuity: *There are constants $C > 0$, $M > 0$ such that $|a(u,v)| \leq C\|u\|\|v\|$ for all $u, v \in U$ and $|F(v)| \leq M\|v\|$ for all $v \in U$; and*

Coercivity: *There is a constant $\alpha > 0$ such that $a(u,u) \geq \alpha\|u\|^2$ for all $u \in U$.*

Then we may conclude there is a unique $u \in U$ such that $a(u,v) = F(v)$ for all $v \in U$, and $\|u\| \leq \frac{M}{\alpha}$.

1.5.2 Theorem. *Let $U = \{u \in H^1(0,1) \mid u(0) = 0\}$. Then there is a constant $c_P > 0$ such that, for each $u \in U$,*

$$\|u\|_{L^2}^2 = \int_0^1 u^2 dx \leq c_P^2 \int_0^1 (u')^2 dx = c_P^2 \|u'\|_{L^2} =: c_P^2 |u|_{H^1}^2.$$

Note that $|\cdot|_{H^1}$ is a semi-norm on $H^1(0,1)$.

PROOF: Using the fundamental theorem of calculus and the fact that $u(0) = 0$,

$$\begin{aligned}
\|u\|_{L^2}^2 &= \int_0^1 u^2(x) dx \\
&= \int_0^1 \int_0^x \frac{d}{ds} u^2(s) ds dx \\
&= \int_0^1 \int_0^x 2u(s)u'(s) ds dx \\
&\leq 2 \int_0^1 \left| \int_0^x u(s)u'(s) ds \right| dx \\
&\leq 2 \int_0^1 \|u\|_{L^2} \|u'\|_{L^2} dx \\
&= 2\|u\|_{L^2} \|u'\|_{L^2}
\end{aligned}$$

So the constant c_P in this case is at most 2. □

The above lemma does not hold if the condition that $u(0) = 0$ is dropped. Indeed, any non-zero constant function is a counterexample.

Notice that, for $u \in H^1(0,1)$,

$$|u|_{H^1} = \|u'\|_{L^2}^2 \leq \|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|u'\|_{L^2}^2 \leq (1 + c_P^2)\|u'\|_{L^2}^2,$$

so the $H^1(0,1)$-semi-norm is equivalent to the $H^1(0,1)$-norm. (This probably does not hold in general.)

1.6 Well-posedness of weak approach

A problem is said to be *well-posed* if it has a unique solution that "depends continuously on the data". For the model problem it suffices to show that the hypotheses of the Lax-Milgram theorem are satisfied by $F(v) = \int_0^1 f v\,dx$ and $a(u,v) = \int_0^1 u'v'\,dx$.

Continuity: We have $a(u,v) = \int_0^1 u'v'\,dx \le \|u'\|_{L^2}\|v'\|_{L^2} \le \|u\|_{H^2}\|v\|_{H^2}$ and $F(v) = \int_0^1 f v\,dx \le \|f\|_{L^2}\|v\|_{L^2} \le \|f\|_{L^2}\|v\|_{H^2}$, if $f \in L^2(0,1)$. In fact, later we will see that f may live in a much larger space, $H^{-1}(0,1)$.

Coercivity: We have $a(u,u) = \int_0^1 (u')^2 dx = \|u'\|_{L^2}^2 \ge \frac{1}{1+c_P^2}\|u\|_{H^1}^2$ by the Poincaré lemma.

By the Lax-Milgram theorem, when $f \in L^2(0,1)$ there is a unique $u \in U$ such that $\int_0^1 u'v'\,dx = \int_0^1 f v\,dx$ for all $v \in U$ and $\|u\|_{H^1} \le (1+c_P^2)\|f\|_{L^2}$. But how do we find it?

1.7 From the continuous to the discrete

To discretize the problem we introduce a finite dimensional subspaces, the *trial space* and the *test space*, $U_h \le U$ and $V_h \le V$, and seek a solution to an approximate problem in U_h. The *Galerkin method* requires us to find $u_h \in U_h$ such that $a(u_h, v_h) = F(v_h)$ for all $v_h \in V_h$. (Often in the engineering literature the Galerkin method also requires $U_h = V_h$. Otherwise the method may be said to be *non-conforming*.)

There are two important questions that must be addressed. If the continuous problem is well-posed then is the discrete problem also well-posed? And, if the discrete problem is well-posed then how do we compute solutions? The first of these questions is answered for the model problem in the affirmative by a trivial application of the Lax-Milgram theorem.

As for the second, in general, let $\{\phi_j, j = 1, \dots, N\}$ be any basis of U_h. Since it is a basis we can write $u_h = \sum_{i=1}^N u_j\phi_j$ and $v_h = \sum_{i=1}^N v_i\phi_i$. Applying a we get

$$a(u_h, v_h) = \sum_{i,j=1}^N v_i a(\phi_j, \phi_i) u_j = v^T A u$$

where u and v are the vectors of coefficients and A is an $N \times N$ matrix with $A_{ij} := a(\phi_i, \phi_j)$. Also, $F(v_h) = \sum_{i=1}^N v_i F_i$, where $F_i := \int_0^1 f\phi_i\,dx$. Since $v^T A u = a(u_h, v_h) = F(v_h) = v^T F$ must hold for all $v_h \in U_h$ (i.e. all $v \in \mathbb{R}^n$) the *discrete variational problem* reduces to the linear system of equations $Au = F$.

1.7.1 Example. Let $0 = x_0 < x_1 < \cdots < x_N = 1$ and U_h be the subspace of $C[0,1]$ consisting of functions u_h that are linear (i.e. affine) on each interval (x_{n-1}, x_n), for $n = 1, \ldots, N$, and $u_h(0) = 0$. Define

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

and truncate ϕ_N at $x = 1$. Notice that $\phi_i(x_j) = \delta_{ij}$, that $\{\phi_j, j = 1, \ldots, N\}$ is a basis of U_h, and that

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

Specifying to the uniform mesh, $x_i := i/N$, and to the model problem,

$$\phi_i(x) := \begin{cases} Nx - (i-1) & x \in [\frac{i-1}{N}, \frac{i}{N}) \\ (i+1) - Nx & x \in [\frac{i}{N}, \frac{i+1}{N}) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \phi_i'(x) = \begin{cases} N & x \in (\frac{i-1}{N}, \frac{i}{N}) \\ -N & x \in (\frac{i}{N}, \frac{i+1}{N}) \\ 0 & \text{otherwise} \end{cases}$$

and

$$A_{ij} = a(\phi_i, \phi_j) = \int_0^1 \phi_i' \phi_j' dx = \begin{cases} 2N & i = j < N \\ N & i = j = N \\ -N & |i - j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

and $F_i = \int_0^1 \phi_i f \, dx = \int_{x_{i-1}}^{x_{i+1}} \phi_i f \, dx$. Notice that F_i is a weighted average of f over a small interval centred at x_i.

A Python implementation of the above is provided below.

```
from numpy import *
from scipy.sparse import lil_matrix
from scipy.sparse.linalg import spsolve

N = 1000
f = lambda x: sin(4*x)
x = linspace(0.0, 1.0, N+1)
F = (1.0/N)*f(x) # incorrect
A = lil_matrix((N, N))
A.setdiag([2.0*N]*N)
A.setdiag([-N]*(N-1), 1)
A.setdiag([-N]*(N-1), -1)
A[N-1, N-1] = N
u = zeros(N+1)
```

```
u[1:N+1] = spsolve(A.tocsr(), F[1:N+1])

with open('output.dat', 'w') as o:
    for t in zip(x, u, f(x)):
        o.write('%f %f %f\n' % t)
```

1.8 Automating the computations

Let's recall what we had to do to arrive at the finite element approximation.
(i) Partition the problem domain;
(ii) Construct the finite element basis $\{\phi_j, j = 1,\dots,N\}$;
(iii) Compute $A_{ij} = a(\phi_i,\phi_j)$;
(iv) Compute $F_i = F(\phi_j)$;
(v) Solve the linear system;

Steps (iii) and (iv) are together referred to as the *assembly* of the problem. The first two steps are often referred to as *geometry*.

Partition the interval $[0,1]$ by $0 = x_0 < x_1 < \cdots < x_N = 1$ and consider the basis of "spike" functions for the subspace of piecewise continuous functions, as in 1.7.1. Then

$$A_{ij} = a(\phi_j,\phi_i) = \int_0^1 \phi_j'\phi_i' dx = \sum_{n=1}^N \int_{x_{n-1}}^{x_n} \phi_j'\phi_i' dx =: \sum_{n=1}^N A_{ij}^{(n)}.$$

Now ϕ_i and ϕ_i' are non-zero on (x_{n-1},x_n) if and only if $i = n-1$ or $i = n$. Whence the entries of $A^{(n)}$ are non-zero only for $(i,j) = (n-1,n-1)$, $(n-1,n)$, $(n,n-1)$, and (n,n). Similarly,

$$F_i = F(\phi_i) = \int_0^1 f\,\phi_i dx = \sum_{n=1}^n \int_{x_{n-1}}^{x_n} f\,\phi_i dx =: \sum_{n=1}^N F_i^{(n)}.$$

The information content of $A^{(n)}$ is the 2×2 *element matrix* A_e of non-zero entries, and of $F^{(n)}$ is the 2-dimensional *element vector* F_e of non-zero entries.

Define $\hat\phi_1$ and $\hat\phi_2$ on the *parent element* or *reference element* $[-1,1]$ by

$$\hat\phi_1(x) = \frac{1-x}{2}, \qquad \hat\phi_2(x) = \frac{1+x}{2}.$$

Basis functions on the element $[x_{n-1},x_n]$ are related to the functions on the parent element by the affine transformation

$$T_n : [-1,1] \to [x_{n-1},x_n] : \hat x \mapsto \frac{x_n + x_{n-1}}{2} + \frac{x_n - x_{n-1}}{2}\hat x,$$

where $\phi_{n-1}(T_n\hat x) = \hat\phi_1(\hat x)$ and $\phi_n(T_n\hat x) = \hat\phi_2(\hat x)$. This change of variables allows us to integrate over $[-1,1]$ when computing on any element, which will help simplify the automation of the integration. For an arbitrary g,

$$\int_{x_{n-1}}^{x_n} g(x)dx = \int_{-1}^1 g(T_n\hat x)\left|\frac{dx}{d\hat x}\right| d\hat x = \int_{-1}^1 g(T_n\hat x)\left(\frac{x_n - x_{n-1}}{2}\right) d\hat x.$$

The derivatives of ϕ_{n-1} and ϕ_n are computed using the chain rule:

$$\frac{d\hat{\phi}_1}{d\hat{x}} = \frac{d\phi_{n-1}}{dx}\frac{dx}{d\hat{x}},$$

so $\phi'_{n-1} = (\frac{dx}{d\hat{x}})^{-1}\hat{\phi}'_1$, and similarly $\phi'_n = (\frac{dx}{d\hat{x}})^{-1}\hat{\phi}'_2$.

On the element $[x_{n-1}, x_n]$, $(\frac{dx}{d\hat{x}})^{-1} = \frac{2}{x_n - x_{n-1}}$. Let $h := \frac{x_n - x_{n-1}}{2} = \frac{dx}{d\hat{x}}$ and $\bar{x} := \frac{x_n + x_{n-1}}{2}$. Notice that h is independent of N for the uniform partition. On each element $[x_{n-1}, x_n]$, for $1 \le i, j \le 2$

$$(A_e)_{ij} = \int_{x_{n-1}}^{x_n} \phi'_{n-2+j}\phi'_{n-2+i}dx = \frac{1}{h}\int_{-1}^{1} \hat{\phi}'_i\hat{\phi}'_j d\hat{x},$$

and

$$(F_e)_i = \int_{x_{n-1}}^{x_n} f(x)\phi_{n-2+i}dx = \int_{-1}^{1} f(\bar{x} + h\hat{x})\hat{\phi}_i(\hat{x})hd\hat{x}.$$

1.9 Numerical integration (quadrature)

In general we are not necessarily going to be able to evaluate the integrals that appear in the last expressions for A_e and F_e exactly. We would like to automate the computation of a numerical approximation to the integral. The *general quadrature rule* says

$$\int_a^b g(x)dx \approx \sum_{k=1}^{K} g(x_k)w_k,$$

where each x_k is a *quadrature point* and w_k is the corresponding *weight*.

1.9.1 Example. On the interval $[a, b]$ we have the following quadrature rules.

(i) Midpoint rule: $x_1 = \frac{a+b}{2}$, $w_1 = b - a$.

(ii) Trapezoid rule: $x_1 = a$, $x_2 = b$, $w_1 = w_2 = \frac{b-a}{2}$.

(iii) Simpson's rule, adaptive Simpson's rule, etc.

On the interval $[-1, 1]$ we have the 2-point Gaussian quadrature rule: $w_1 = w_2 = 1$ and $x_1 = -x_2 = \frac{1}{\sqrt{3}}$.

The midpoint and trapezoid rules integrate polynomials of degree one exactly. The 2-point Gaussian rule integrates cubic polynomials exactly. There is a 3-point Gaussian rule that integrates quintic polynomials exactly. A good strategy for choosing a quadrature rule is to use a rule that will integrate inner products of basis functions exactly, but not too much more. For example, for our piecewise linear basis functions, the 2-point Gaussian rule will do. It is important to note that the values of the basis functions at quadrature points can be precomputed and stored.

1.10 Boundary Conditions

There is a small hitch when dealing with the first element $[x_0, x_1]$, since there is no "ϕ_0". To get around this, define ϕ_0 and add it to the basis. Then compute the assembly in the same manner for all elements. After the coefficient matrix and right hand side vector are computed, go back and correct the linear system to account for the appropriate boundary conditions. There are two approaches. The first is to replace A_{00} with a very large value and set $F_0 = 0$. The other is to set $A_{00} = 1$ and $A_{0j} = 0$ for $j \geq 1$ and set $F_0 = 0$.

1.11 Main Program

The program implementing the finite element method as we've discussed it will require the following steps, with various associated subroutines.

Step	Components
Initialization	Geometry
	Quadrature rule
	Construct basis
Assembly	Construct coefficient matrix
	Construct RHS vector
	Deal with boundary conditions
Solve	Linear solver
Postprocessing	Generate data
	Visualization
	Compute norms, errors, etc.

```
from numpy import *

def geometry(l, r, ne, nbf):
    """Returns the "geometry" associated with the interval
    [l, r] with ne elements and nbf basis functions per element.

    Returns a 3-tuple consisting of:
    1) number of nodes required (nbf-1)*ne+1 (integer)
    2) array of x-coordinates of nodes
    3) "node" function, defined by:
        node(ele, fn) = global node number of fn^th
                        local node in ele^th element
    """
    nx = (nbf-1) * ne + 1
    node = lambda e, f: (nbf-1) * e + f
    return nx, linspace(l, r, nx), node

def quadrature(nqp):
```

```
    """Returns arrays of quadrature points and weights."""
    if (1 == nqp): # trivial quadrature
        return array([0.0]), \
               array([2.0])
    elif (2 == nqp): # 2-point Gaussian
        return array([-sqrt(1.0/3), sqrt(1.0/3)]), \
               array([1.0, 1.0])
    elif (3 == nqp): # 3-point Gaussian
        return array([-sqrt(3.0/5), 0.0, sqrt(3.0/5)]), \
               array([5.0/9, 8.0/9, 5.0/9])
    elif (4 == nqp): # 4-point Gaussian
        ip = sqrt((3 - 2*sqrt(1.2))/7)
        iw = (18 + sqrt(30))/36
        op = sqrt((3 + 2*sqrt(1.2))/7)
        ow = (18 - sqrt(30))/36
        return array([-op, -ip, ip, op]), \
               array([ow, iw, iw, ow])
    elif (5 == nqp): # 5-point Gaussian
        ip = sqrt(5 - 2*sqrt(10.0/7))/3
        iw = (322 + 13*sqrt(70))/900
        op = sqrt(5 + 2*sqrt(10.0/7))/3
        ow = (322 - 13*sqrt(70))/900
        return array([-op, -ip, 0.0, ip, op]), \
               array([ow, iw, 128.0/225, iw, ow])
    else: # Higher order rules not implemented yet
        raise NotImplementedError

def precompute_basis(nbf, qp):
    """Returns basis function values at the quad points.

    Input qp should be a numpy array.  Returns two lists, each
    the same length as qp.  The q^th element of each list is a
    list of length nbf of numbers that are the basis functions
    (resp. the derivatives of the basis functions) evaluated at
    the q^th quadrature point.
    """
    if (2 == nbf): # linear basis functions
        ph0  = lambda x: 0.5 * (1.0-x)
        ph1  = lambda x: 0.5 * (1.0+x)
        dph0 = lambda x: -0.5 * ones(x.shape)
        dph1 = lambda x:  0.5 * ones(x.shape)
        return zip(ph0(qp), ph1(qp)), zip(dph0(qp), dph1(qp))
    if (3 == nbf): # quadratic basis functions
        ph0  = lambda x: 0.5 * x * (x-1)
        ph1  = lambda x: 1 - x * x
        ph2  = lambda x: 0.5 * x * (x+1)
```

```
        dph0 = lambda x: x - 0.5
        dph1 = lambda x: -2.0 * x
        dph2 = lambda x: x + 0.5
        return zip(ph0(qp), ph1(qp), ph2(qp)), \
               zip(dph0(qp), dph1(qp), dph2(qp))
    else: # Other basis functions not implemented yet
        raise NotImplementedError

def assembly(ne, nbf, xc, node, qp, qw, ph, dph, f, b=lambda x: 0.0):
    """Returns the coefficient matrix and the RHS vector for the
    equation -u'' + bu = f.

    ne - number of elements
    nbf - number of basis functions per element
    xc - numpy array of x-coords of nodes
    node - as descibed in geometry function doc string
    qp - numpy array of quadrature points
    qw - numpy array of quadrature weights
    ph - list of arrays of basis functions evaluated at
         quadrature points
    dph - list of arrays of derivatives of basis functions
          evaluated at quadrature points
    """
    nx  = len(xc) # number of nodes
    nqp = len(qp) # number of quadrature points

    Amat = matrix(zeros((nx, nx)))
    Fvec = zeros(nx)
    for n in range(ne):
        #elem_node_coord = xc[node(n, array(range(nbf)))]
        Jmat = (xc[node(n, nbf-1)] - xc[node(n, 0)])/2
        Jinv = 1.0/Jmat
        xbar = (xc[node(n, nbf-1)] + xc[node(n, 0)])/2

        Ae = mat(zeros((nbf, nbf)))
        Fe = zeros(nbf)
        for q in range(nqp):
            x = xbar + Jmat * qp[q]
            weight = abs(Jmat) * qw[q]
            dphidx = [dfq * Jinv for dfq in dph[q]]
            for i in range(nbf):
                Fe[i] += f(x) * ph[q][i] * weight
                for j in range(nbf):
                    Ae[i, j] += weight * (dphidx[i] * dphidx[j] + \
                            b(x) * ph[q][i] * ph[q][j]) # NEW!!!
        k = node(n, 0)
```

```
        Fvec[k:k+nbf] += Fe
        Amat[k:k+nbf, k:k+nbf] += Ae

    return Amat, Fvec

def errors(ne, nbf, uh, xc, node, u, udx, nqp):
    """Computes the error of the approximation uh with respect to
    the true solution u in the L^2 and H^1 norms.

    ne - number of elements
    nbf - number of basis functions associated with uh
    uh - numpy array of ceofficients of approximate solution
    xc - numpy array of x-coords of nodes
    node - as descibed in geometry function doc string
    u - exact solution (function)
    udx - first derivative of u (function)
    npq - number of quadrature points to use in computing errors
    """
    qp, qw = quadrature(nqp)
    ph, dph = precompute_basis(nbf, qp)
    L2error2 = 0.0
    H1semi2 = 0.0

    for n in range(ne):
        # get local node coordinates elem_node_coord
        Jmat = (xc[node(n, nbf-1)] - xc[node(n, 0)])/2
        Jinv = 1.0/Jmat
        xbar = (xc[node(n, nbf-1)] + xc[node(n, 0)])/2

        for q in range(nqp):
            x = xbar + Jmat * qp[q]
            weight = Jmat * qw[q]
            dphidx = [dphi * Jinv for dphi in dph[q]]
            approx = 0.0
            approxdx = 0.0
            for i in range(nbf):
                approx += uh[node(n, i)] * ph[q][i]
                approxdx += uh[node(n, i)] * dphidx[i]
            L2error2 += (u(x)-approx) * (u(x)-approx) * weight
            H1semi2 += (udx(x)-approxdx)*(udx(x)-approxdx)*weight

    return sqrt(L2error2), sqrt(L2error2 + H1semi2)

def norms(ne, nbf, uh, xc, node, nqp):
    """Computes the norms of the approximation uh."""
    return errors(ne, nbf, uh, xc, node, \
```

```
        lambda x: 0.0, lambda x: 0.0, nqp)
```

1.12 Accuracy

1.12.1 Theorem. *Let U be a Hilbert space, and let $a : U \times U \to \mathbb{R}$ and $F : U \to \mathbb{R}$ satisfy the hypotheses of the Lax-Milgram theorem. Let $U_h \le U$ be a closed subspace and suppose $u \in U$ and $u_h \in U_h$ satisfy $a(u,v) = F(v)$ for all $v \in V$ and $a(u_h, v_h) = F(v_h)$ for all $v_h \in V_h$. Then*

$$\|u - u_h\| \le \left(\frac{C}{\alpha}\right) \inf_{w_h \in U_h} \|u - w_h\|.$$

Recall that a is bilinear, F is linear, and there are constants α, C, and M such that $|a(u,v)| \le C\|u\|\|v\|$, $|a(u,u)| \ge \alpha\|u\|^2$, and $|F(v)| \le M\|v\|$ for all $u, v \in U$.

PROOF: We have $a(u - u_h, v_h) = a(u, v_h) - a(u_h, v_h) = 0$ for all $v_h \in U_h$, i.e. $u - u_h$ is a-orthogonal to U_h. Conclude by noting

$$\begin{aligned}
\alpha\|u - u_h\|^2 &\le a(u - u_h, u - u_h) && \text{by coercivity} \\
&= a(u - u_h, u - w_h + w_h - u_h) && \text{for any } w_h \in U_h \\
&= a(u - u_h, u - w_h) + a(u - u_h, w_h - u_h) && \text{by linearity} \\
&= a(u - u_h, u - w_h) && \text{by orthogonality} \\
&\le C\|u - u_h\|\|u - w_h\| && \text{by continuity} \qquad \square
\end{aligned}$$

1.12.2 Lemma. *Let $u \in H^1(0,1)$ and let w_h be the piecewise constant function on the partition $0 = x_0 < x_1 < \cdots < x_N = 1$ that assumes the values*

$$w_h|_{(x_{n-1},x_n)} = \frac{1}{x_n - x_{n-1}} \int_{x_{n-1}}^{x_n} u(x)dx.$$

Then

$$\|u - w_h\|_{L^2}^2 \le \sum_{n=1}^{N} (x_n - x_{n-1})^2 \int_{x_{n-1}}^{x_n} |u'(x)|^2 dx.$$

In particular, on the uniform partition, $\|u - w_h\| \le h\|u'\|_{L^2} = h|u|_{H^1}$.

PROOF: Since $u \in H^1(0,1)$, u is continuous. So there is $z \in (x_{n-1}, x_n)$ such that $u(z) = \frac{1}{x_n - x_{n-1}} \int_{x_{n-1}}^{x_n} u(x)dx = w_h|_{(x_{n-1},x_n)}$. Then for any $x \in (x_{n-1}, x_n)$,

$$
\begin{aligned}
u(x) - w_h(x) &= u(z) - w_h(z) + \int_z^x (u - w_h)'(t)dt \\
&= \int_z^x u'(t)dt \\
\text{so } |u(x) - w_h(x)|^2 &\le \left(\int_z^x |u'(t)|dt \right)^2 \\
&\le |x - z| \int_z^x |u'(t)|^2 dt \\
&\le (x_n - x_{n-1}) \int_{x_{n-1}}^{x_n} |u'(t)|^2 dt
\end{aligned}
$$

on the interval (x_{n-1}, x_n). It follows that

$$
\int_{x_{n-1}}^{x_n} |u(x) - w_h(x)|^2 dx \le (x_n - x_{n-1})^2 \int_{x_{n-1}}^{x_n} |u'(x)|^2 dx.
$$

□

In the model problem, $(-u'' = f, u(0) = u'(1) = 0)$, on a uniform N element partition of $[0,1]$ $(h = \frac{1}{N})$, provided the solution u satisfies $u'' \in L^2(0,1)$,

$$
\|u - u_h\|_{H^1} \le \left(\frac{C}{\alpha} \right) \inf_{w_h \in U_h} \|u - w_h\|_{H^1} \le \frac{C}{\alpha} h \sqrt{1 + h^2} \|u''\|_{L^2} \le \tilde{C} h.
$$

1.13 Rates of convergence

From the considerations in the previous section we have seen that our finite element approximation u_h should satisfy $\|u - u_h\|_{H^1} \le C_1 h$. Further, from the homework, $\|u - u_h\|_{L^2} \le C_2 h^2$. In general we will obtain an estimate of the form $\|u - u_h\| \le C h^r$, for some norm and some r. How can we verify this experimentally?

Suppose we have computed the approximation on two different meshes, of sizes $h_1 \ne h_2$. Then

$$
\frac{e_1}{e_2} := \frac{\|u - u_{h_1}\|}{\|u - u_{h_2}\|} \approx \frac{C h_1^r}{C h_2^r} = \left(\frac{h_1}{h_2} \right)^r,
$$

provided these estimates are "sharp". It follows that $\log \frac{e_1}{e_2} = r \log \frac{h_1}{h_2}$, giving an estimate for r. The theoretical rate should be observed, for small enough h, provided the code is written properly. How can we find a u to use to compute $\|u - u_h\|$? Apply the *method of manufactured solutions*: choose some convenient $\tilde{u}$ and then, using the strong form of the problem statement, determine the f that would force the solution to be your $\tilde{u}$.

1.13.1 Example. Let $\tilde{u}(x) = \sin(\frac{5\pi}{2}x)$, so that $\tilde{u}'(x) = \frac{5\pi}{2}\cos(\frac{5\pi}{2}x)$ and $\tilde{u}''(x) = -\frac{25\pi^2}{4}\sin(\frac{5\pi}{2}x)$. Taking $f = \frac{25\pi^2}{4}\sin(\frac{5\pi}{2}x)$ in the model problem will force $\tilde{u}$ to be a solution.

1.14 Computing norms and errors

Notice that

$$\|u\|_{L^2} = \int_0^1 u^2 dx = \sum_{n=1}^{N}\int_{x_{n-1}}^{x_n} = \sum_{n=1}^{N}\int_{-1}^{1} u^2(T_n(\hat{x}))\left|\frac{dx}{d\hat{x}}\right| d\hat{x} \approx \sum_{n=1}^{N}\sum_{k=1}^{K} u^2(T_n(\hat{x}_k))\left|\frac{dx}{d\hat{x}}\right| w_k$$

and

$$\|u_h\|_{L^2}^2 \approx \sum_{n=1}^{N}\sum_{k=1}^{K}(\underbrace{u_{n-1}\hat{\phi}_1(\hat{x}_k) + u_n\hat{\phi}_2(\hat{x}_k)}_{u_n(\hat{x}_k)})^2\left|\frac{dx}{d\hat{x}}\right| w_k$$

1.15 Non-homogeneous boundary conditions

Consider the problem $-u'' = f$ on $(0,1)$, $u(0) = u_0$, $u'(1) = \gamma$. The variational form of the problem has

$$\int_0^1 u'v'dx - u'(1)v(1) + u'(0)v(0) = \int_0^1 f v dx.$$

If $v(0) = 0$ then $u'(0)v(0) = 0$. The problem becomes to find

$$u \in U(u_0) := \{u \in H^1(0,1) \mid u(0) = u_0\}$$

such that

$$a(u,v) := \int_0^1 u'v'dx = \int_0^1 f v dx + \gamma v(1) =: F(v)$$

for all $v \in U(0)$. Note that $U(u_0) = \hat{u}_0 + U(0)$, where $\hat{u}_0$ is any function in $H^1(0,1)$ such that $\hat{u}_0(0) = u_0$. Hence it suffices to find $u \in U(0)$ such that $a(u,v) = F(v) - a(\hat{u}_0, v)$ for all $v \in V$.

1.15.1 Lemma. *Let H be a Hilbert space with semi-norm $|\cdot|_H$ and let $U \le H$ be a closed subspace such that $|\cdot|_H$ is a norm on U. Let $a : H \times U \to \mathbb{R}$ be continuous and assume that, when restricted to $U \times U$, a is coercive.*

If $F : U \to \mathbb{R}$ is continuous and $u_0 \in H$ is specified then the problem of finding $u \in u_0 + U$ such that $a(u,v) = F(v)$ for all $v \in U$ has a unique solution satisfying $|u|_H \le \frac{M}{\alpha} + (1 + \frac{C}{\alpha})|u_0|_H$.

PROOF: Let $\tilde{F}(v) := F(v) - a(u_0, v)$ and notice that $|\tilde{F}(v)| \le M|v|_H + C|u_0|_H|v|_H = (M + C|u_0|_h)|v|_H$, so $\tilde{F}$ is still continuous on U.

Skipped steps.

Let $u \in H$ solve $a(u,v) = F(v)$ and let $\tilde{u} = u - u_0 \in U$. Then $\alpha\|\tilde{u}\|^2 \le a(\tilde{u},\tilde{u})$ implies

$$\|\tilde{u}\|^2 \le \frac{1}{\alpha}(\tilde{u},\tilde{u}) = \frac{1}{\alpha}\tilde{F}(\tilde{u}) \le \frac{1}{\alpha}(M + C|u_0|_H)|\tilde{u}|_H. \qquad \square$$

For the discretization, recall that we originally defined $U_h \le U$ with basis functions at all of the nodes except $x_0 = 0$. A natural discrete subset of $U(u_0)$ is

$$U_h(u_0) = \left\{ \sum_{n=0}^{N} u_n \phi_n \mid u_n \in \mathbb{R}, i = 1, \ldots, N, \phi_n|_{(x_{n-1},x_n)} \text{ is linear} \right\}.$$

This is a translate of U_n. The discrete problem is to find $u_h \in U_h(u_0)$ such that $a(u_h, v_h) = F(v_h)$ for all $v_h \in U_h$.

In practise, the only modification to the code is altering the boundary condition manipulation step. Here we set $F_0 = u_0$ (instead of 0) and $F_N = F_N + \gamma$. That's it!

1.16 Higher order elements

Intuitively, piecewise quadratic polynomial functions will approximate the solution better than piecewise linear functions, so we can use them to increase the rate of convergence. Let P_2 denote the collection of polynomials of degree at most two. Define

$$U_h = \{u \in C[0,1] \mid u|_{(x_{n-1},x_n)} \in P_2(x_{n-1},x_n), n = 1,\ldots,N; \text{and } u(0) = 0\}.$$

Once we have a basis $\{\phi_i\}_{i=0}^N$ for U_h, the problem reduces to the same form as earlier. Note that quadratic basis elements will have three degrees of freedom, as opposed to just two for linear basis functions. There are two ways of describing the basis elements in terms of the partition points and nodes.

One way is to let an element consist be the interval between partition points. In this case we need to create an additional "node" between each consecutive pair of partition points. Say $x_{n-\frac{1}{2}} \in (x_{n-1}, x_n)$. Often we will take this node to be the average of the endpoints. In this method the transformation $x = T_n(\hat{x})$ is the same as before. Indices need to be manipulated for unknowns and basis functions, but the element numbering is easy.

Another way to do it is to have an element to consist of three consecutive partition points. In this case the map $x = T_n(\hat{x})$ is different from the linear case. Element numbering needs to be manipulated, but nodes, unknowns, and basis functions will all have the same indices.

1.16.1 Lagrange interpolating polynomials. *To interpolate planar points* $(x_1, x_1), \ldots, (x_n, y_n)$ *with an* $(n-1)$*-degree polynomial, one can use the formula*

$$p(x) = \sum_{i=1}^{n} y_i \ell_i(x) := \sum_{i=1}^{n} y_i \prod_{\substack{j=1 \\ j\ne i}}^{n} \frac{x - x_j}{x_i - x_j}.$$

Note that $\ell_i(x_j) = \delta_{i,j}$.

As before, we take $[-1, 1]$ as the parent element, but now the relevant points are $\{-1, 0, 1\}$ instead of just the endpoints. We define three basis functions $\hat{\phi}_0$, $\hat{\phi}_1$, and $\hat{\phi}_2$ on the parent element. $\hat{\phi}_0$ is the Lagrange interpolating polynomial interpolating $\{(-1, 1), (0, 0), (1, 0)\}$, so $\hat{\phi}_0(\hat{x}) = \frac{1}{2}\hat{x}(\hat{x} - 1)$. $\hat{\phi}_1$ is the Lagrange interpolating polynomial interpolating $\{(-1, 0), (0, 1), (1, 0)\}$, so $\hat{\phi}_1(\hat{x}) = 1 - \hat{x}^2$. $\hat{\phi}_2$ is the Lagrange interpolating polynomial interpolating $\{(-1, 0), (0, 0), (1, 1)\}$, so $\hat{\phi}_2(\hat{x}) = \frac{1}{2}\hat{x}(\hat{x} + 1)$.

We take the second approach in the relationship between elements and nodes. An element is an interval $[x_{2n}, x_{2n+2}]$ for $n = 0, 1, \ldots, \frac{N}{2} - 1$, where N must now be even. Supposing that $x_{2n+1} = \frac{1}{2}(x_{2n} + x_{2n+2})$ for all n, the transformation

$$x = T_n(\hat{x}) = \frac{x_{2n+2} + x_{2n}}{2} + \frac{x_{2n+2} - x_{2n}}{2}\hat{x}$$

gives the correct mapping from the parent element to the elements.

It can be seen that the theoretical convergence rates become

$$\|u - u_h\|_{L^2} \leq Ch^3\|u'''\|_{L^2} \quad \text{and} \quad \|u - u_h\|_{H^1} \leq Ch^2\|u'''\|_{L^2}.$$

Note that, to compare linear basis functions with quadratic basis functions, with this numbering scheme, more nodes are required. Linear on $N + 1$ nodes gives N elements, but to get N elements with quadratic we need $2N + 1$ nodes.

Aside: $M_{ij} := \int_{-1}^{1} \phi_i \phi_j dx$ is known as the *mass matrix* or *Gramian*, and $K_{ij} = \int_{-1}^{1} \phi_i' \phi_j' dx$ is known as the *stiffness matrix*.

2 Finite Element Approximation of Elliptic Problems

2.1 Review

Please refer to my notes 21-832, Partial Differential Equations 2, for a rigorous introduction to Sobolev spaces. We list important results below.

2.1.1 Theorem. *For any domain Ω, $C^\infty(\Omega) \cap W^{k,p}(\Omega)$ is dense in $W^{k,p}(\Omega)$.*

From now on let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain (i.e. a bounded, connected, open set) with Lipschitz boundary.

2.1.2 Theorem. *Let $k > m \geq 0$ be integers, and let $1 \leq p < \infty$ such that*
(i) $k - m \geq n$ if $p = 1$; or
(ii) $k - m > n/p$ if $p > 1$.
Then there is a constant c such that, for all $u \in W^{k,p}(\Omega)$,

$$\|u\|_{W^{m,\infty}(\Omega)} \leq c\|u\|_{W^{k,p}(\Omega)}.$$

2.1.3 Theorem (Trace). *Let $1 \leq p < \infty$. Then for all $u \in W^{1,p}(\Omega)$, $u|_{\partial\Omega}$ is well-defined and there is a constant c_t depending only on Ω such that*

$$\|u|_{\partial\Omega}\|_{L^p(\partial\Omega)} \leq c_t\|u\|_{W^{1,p}(\Omega)}.$$

For $1 \leq p \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, and $k \geq 1$, denote the dual space of $W^{k,p}(\Omega)$ by $W^{-k,q}(\Omega)$. Denote $(H_0^1(\Omega))'$ by $H^{-1}(\Omega)$. These space may be referred to as a *negative Sobolev spaces*. These spaces can be very large and contain interesting objects. E.g. $\delta \in W^{-k,p}(\Omega)$ provided that $k > n - \frac{n}{p}$.

2.1.4 Theorem (Divergence). *Let $A : \Omega \to \mathbb{R}^n$ be a continuously differentiable vector field. Then*

$$\int_\Omega div(A) d\Omega = \int_{\partial\Omega} A \cdot \vec{n} d\Gamma,$$

where $\vec{n}$ is the outward pointing normal vector for $\partial\Omega$.

Apply the *Divergence Theorem* to $A^{(i)} = (0, \dots, 0, vw, 0, \dots, 0)$, where vw appears in the i^{th} coördinate, to see that

$$\int_\Omega \left(\frac{\partial v}{\partial x_i} w + v \frac{\partial w}{\partial x_i} \right) d\Omega = \int_\Omega \text{div}(A^{(i)}) d\Omega = \int_{\partial\Omega} vw\vec{n}_i d\Gamma.$$

Replace w with $\frac{\partial w}{\partial x_i}$ in the i^{th} expression and sum over $i = 1, \dots, n$ to see

$$\int_\Omega \nabla v \cdot \nabla w d\Omega = -\int_\Omega v \Delta w d\Omega + \int_{\partial\Omega} v \nabla w \cdot \vec{n} d\Gamma.$$

This is known as *Green's Theorem*.

2.1.5 Theorem (Poincaré Inequality). *Let $\Gamma \subseteq \partial\Omega$ have positive boundary measure. There is a constant c_P depending only on Ω and Γ such that*

$$\|u\|_{L^2(\Omega)} \leq c_P \|\nabla u\|_{L^2(\Omega)}$$

for all $u \in H^1(\Omega)$ such that $u|_\Gamma = 0$. It follows that

$$\|u\|_{H^1(\Omega)} \leq \sqrt{1+c_P^2}\|\nabla u\|_{L^2(\Omega)}$$

on the space $\{u \in H^1(\Omega) : u|_\Gamma = 0\}$.

2.1.6 Theorem (Riesz Representation). *Let H be a Hilbert space. For each $f \in H^*$ there is a unique $h_f \in H$ such that $f(h) = (h_f, h)_H$ for all $h \in H$.*

We write $R : H^* \to H : f \mapsto h_f$ for the *Riesz map*. R is a bijective linear isometry.

2.2 Model problem: the diffusion equation

Let $\Omega \subseteq \mathbb{R}^n$ be an open, connected, bounded, Lipschitz domain with boundary Γ. Write $\Gamma = \overline{\Gamma}_0 \cup \overline{\Gamma}_1$, where Γ_0 and Γ_1 are connected and open in the relative topology of Γ. The problem we would like to solve is

$$-\text{div}(K\nabla u) + bu = f,$$

the *steady-state* of a *diffusion equation*. For example, suppose u is the temperature inside the body Γ and $K : \Omega \to \mathbb{R}^{n\times n}$ is the *diffusivity matrix* for the body: it describes how easily heat flows through Γ. f and b are real-valued functions describing an external heat source and rate of heat loss to the surrounding region, respectively. Suppose that the temperature is held fixed to be u_0 on Γ_0. Denote $K\nabla u \cdot \vec{n}$ on Γ_1 by g, the *heat flux* through that part of the boundary. Note that in the particular case $K = I$ then

$$-\text{div}(K\nabla u) = -\nabla\cdot\nabla u = -\Delta u.$$

Multiply the equation by a smooth function v that is zero on Γ_0 and integrate.

$$\begin{aligned}
\int_\Omega -\text{div}(K\nabla u) v d\Omega + \int_\Omega buv d\Omega &= \int_\Omega f v d\Omega \\
\int_\Omega K\nabla u\cdot\nabla v d\Omega - \int_\Gamma vK\nabla\cdot\vec{n} d\Gamma + \int_\Omega buv d\Omega &= \int_\Omega f v d\Omega \\
\int_\Omega K\nabla u\cdot\nabla v d\Omega + \int_\Omega buv d\Omega &= \int_\Omega f v d\Omega + \int_{\Gamma_1} gv d\Gamma
\end{aligned}$$

Let $V := \{v \in L^2(\Omega) \mid \nabla v \in (L^2(\Omega))^n \text{ and } v = 0 \text{ on } \Gamma_0\}$, the natural test space for this problem. We are looking for u in the space $U(u_0) := u_0 + V$, an affine space. The *weak formulation* is to find $u \in U(u_0)$ such that $a(u,v) = f(v)$ for all $v \in V$, where

$$a(u,v) := \int_\Omega K\nabla u\cdot\nabla v d\Omega + \int_\Omega buv d\Omega \quad \text{and} \quad f(v) := \int_\Omega f v d\Omega + \int_{\Gamma_1} gv d\Gamma$$

2.3 Well-posedness

2.3.1 Theorem (Generalized Lax-Milgram). *Let U be a Banach space, V a Hilbert space, and let $a : U \times V \to \mathbb{R}$ be bilinear and continuous. For any $\alpha > 0$, the following are equivalent.*

(C) *(Coercivity) For each $u \in U$,*

$$\sup_{\substack{v \in V \\ v \neq 0}} \frac{a(u,v)}{\|v\|_V} \geq \alpha \|u\|_U$$

and for each $v \in V \setminus \{0\}$, $\sup_{u \in U} a(u,v) > 0$.

(E) *(Existence of solutions) For each $f \in V^*$ there is a unique $u \in U$ such that $a(u,v) = f(v)$ for all $v \in V$, and $\|u\|_U \leq \frac{1}{\alpha}\|f\|_{V^*}$.*

(E') *(Existence of solutions for the adjoint problem) For each $g \in U^*$ there is a unique $v \in V$ such that $a(u,v) = g(u)$ for all $u \in U$, and $\|v\|_V \leq \frac{1}{\alpha}\|g\|_{U^*}$.*

PROOF: Let $R : V' \to V$ be the Riesz map. For each $u \in U$, $a(u,\cdot)$ is a continuous linear functional on V. By the Riesz representation theorem there is $Au \in V$ such that $(Au, v)_V = a(u,v)$. The operator $A : U \to V$ is linear by the bilinearity of a.

$(C) \implies (E)$: Coercivity of a implies that

$$\|Au\|_V = \sup_{\substack{v \in V \\ v \neq 0}} \frac{(Au,v)_V}{\|v\|_V} = \sup_{\substack{v \in V \\ v \neq 0}} \frac{a(u,v)}{\|v\|_V} \geq \alpha \|u\|_U.$$

If $v \in \text{range}(A)$ then we can write $\|A^{-1}v\|_U \leq \frac{1}{\alpha}\|v\|_V$.

We claim that range(A) is closed. Let $\{u_n\}_{n=1}^\infty$ be a sequence in U such that $Au_n \to y \in V$. Then

$$\alpha \|u_m - u_n\|_U \leq \|Au_m - Au_n\|_V \to 0,$$

so $\{u_n\}_{n=1}^\infty$ has the Cauchy property. Since U is complete, there is $x \in U$ such that $u_n \to x$. By the continuity of a, for each $v \in V$ we find that

$$(Ax, v)_V = a(x,v) = \lim_{n\to\infty} a(u_n, v) = \lim_{n\to\infty} (Au_n, v)_V = (y,v)_V.$$

It follows that $y = Ax \in \text{range}(A)$.

We claim further that A is surjective. Since the range of A is closed we can write $V = \text{range}(A) \oplus \text{range}(A)^\perp$. Let $v^\perp$ be in the orthogonal complement of the range. For each $u \in U$, $a(u, v^\perp) = (Au, v^\perp)_V = 0$, so $\sup_{u \in U} a(u, v^\perp) = 0$. By the second condition in (C), $v^\perp = 0$ and so $V = \text{range}(A)$.

Let $f \in V^*$ and let $v_f = R(f)$. Set $u = A^{-1}v_f$, and notice that, by coercivity,

$$\|u\|_U \leq \frac{1}{\alpha}\|Au\|_V = \frac{1}{\alpha}\|v_f\|_V = \frac{1}{\alpha}\|f\|_{V^*}.$$

For any $v \in V$,

$$a(u,v) = (Au, v)_V = (AA^{-1}v_f, v)_V = (v_f, v)_V = f(v).$$

Uniqueness of the solution u follows from this.

$(E) \implies (E')$: Suppose v_1 and v_2 are solutions to the adjoint problem. Set $f = R^{-1}(v_2 - v_1) \in V^*$. By (E) there is $u \in U$ such that $a(u,v) = f(v)$ for all $v \in V$. In particular,

$$0 = a(u, v_2 - v_1) = f(v_2 - v_1) = (v_2 - v_1, v_2 - v_1) = \|v_2 - v_1\|_V^2.$$

This establishes uniqueness for the adjoint problem.

To establish existence of solutions to the adjoint problem, recall the definition of $A : U \to V$ characterized by $(Au, v)_V = a(u, v)$. We claim that (E) implies that A is a bijection and $\|A^{-1}u\|_U \le \frac{1}{\alpha}\|v\|_V$ for all $v \in V$. Fix $\tilde{v} \in V$. By (E) there is $u \in U$ such that $a(u, v) = (\tilde{v}, v)_V$ for all $v \in V$. It follows that $Au = \tilde{v}$, so A is surjective. To show that A is injective, recall that if $a(u, v) = f(v)$ then (E) states that $\|u\|_U \le \frac{1}{\alpha}\|f\|_{V^*}$. For any $u \in U$, let $f_u(v) := (Au, v)_V = a(u, v)$ to conclude that

$$\|u\|_U \le \frac{1}{\alpha}\|f_u\|_{V^*} = \|(Au, \cdot)\|_{V^*} = \|Au\|_V.$$

Let u be the solution to (E). Set $v = Au$. Then $\|A^{-1}u\|_U \le \frac{1}{\alpha}\|v\|_V$. Let $g \in U^*$. Observe that $g \circ A^{-1} \in V^*$ and has norm bounded by $\frac{1}{\alpha}\|g\|_{U^*}$. Let $v_g := R(g \circ A^{-1})$. For $u \in U$ we compute

$$a(u, v_g) = (Au, v_g)_V = g \circ A^{-1}Au = g(u)$$

and $\|v_g\|_V = \|g \circ A^{-1}\|_{V^*} \le \frac{1}{\alpha}\|g\|_{U^*}$.

$(E') \implies (C)$: Fix $u_0 \in U$ and let $g \in U^*$ satisfy $\|g\|_{U^*} = 1$ and $g(u_0) = \|u_0\|_U$. Such a g exists by the Hahn-Banach theorem. (E') implies the existence of $v_0 \in V$ such that $\|v_0\| \le \frac{1}{\alpha}\|g\|_{U^*} = \frac{1}{\alpha}$ and $a(u, v_0) = g(u)$ for all $u \in U$. In particular,

$$\|u_0\|_U = g(u_0) = a(u_0, v_0) = \frac{a(u_0, v_0)}{\|v_0\|_V}\|v_0\|_V \le \frac{a(u_0, v_0)}{\|v_0\|_V}\left(\frac{1}{\alpha}\right).$$

Therefore

$$\sup_{\substack{v \in V \\ v \ne 0}} \frac{a(u_0, v)}{\|v\|_V} \ge \alpha\|u_0\|_U.$$

The second coercivity condition also follows, since if there is $v \in V$ such that $a(u, v) = 0$ for every $u \in U$ then the uniqueness guaranteed by (E') implies that $v = 0$. □

2.3.2 Corollary. *If U is a Hilbert space then (C) is equivalent to*

(C') *For each $v \in V$,*

$$\sup_{\substack{u \in U \\ u \ne 0}} \frac{a(u, v)}{\|u\|_U} \ge \alpha\|v\|_V$$

and for each $u \in U \setminus \{0\}$, $\sup_{v \in V} a(u, v) > 0$.

Remark.

(i) $(C) \implies (E)$ is originally due to I. Babuška, and the form of the generalized Lax-Milgram theorem is due to F. Brezzi, thus the coercivity condition is sometimes called the Babuška-Brezzi condition.
(ii) If the condition that $\sup_{u\in U} a(u,v) > 0$ for all $v \in V\setminus\{0\}$ is dropped then the theorem remains intact provided that f is suitably restricted, but uniqueness may not hold for the adjoint problem.
(iii) The theorem holds as stated if V is merely a reflexive Banach space.

2.3.3 Theorem (J.-L. Lions). *Let U be a normed linear space and V be a Hilbert space. Suppose that $a : U \times V \to \mathbb{R}$ is bilinear (but not necessarily continuous) and that $a(u,\cdot) \in V^*$ for each $u \in U$. The following are equivalent.*
(C) *(Coercivity) There is $\alpha > 0$ such that, for each $u \in U$,*

$$\sup_{\substack{v\in V\\ v\neq 0}} \frac{|a(u,v)|}{\|v\|_V} \geq \alpha \|u\|_U.$$

(E') *(Existence of solutions for the adjoint problem) For each $g \in U^*$ there is $v \in V$ such that $a(u,v) = g(u)$ for all $u \in U$.*

PROOF: See *Monotone operators in Banach spaces and non-linear partial differential equations* by Showalter. □

2.3.4 Example. The following is an illustration of the application of Lions theorem to "parabolic" problems. Consider the problem $v'(t) + b(t)v(t) = g(t)$, $t \in (0,1)$, with $v(0) = v_0$, for $b \geq \beta > 0$ and $g \in L^2(0,1)$. Let u be smooth and vanish at $t = 1$. Multiply by u and integrate over $t \in (0,1)$.

$$a(u,v) := \int_0^1 (-u'v + buv)dt = \int_0^1 gu dt + v_0 u(0) =: G(v)$$

Set $U := \{u \in H_0^1(0,1) \mid u(1) = 0\}$ with norm $\|u\|_U^2 := \|u\|_{L^2}^2 + u(0)^2$. U is not complete in $\|\cdot\|_U$. Let $V := L^2(0,1)$ with the usual norm. Note that $G \in U^*$ and $a : U \times V \to \mathbb{R}$ is bilinear, but not continuous. Fix $u \in U$. Since $U \subseteq V$ we can set $v = u$ to get

$$a(u,u) = \int_0^1 -u'u dt + \int_0^1 bu^2 dt = -\frac{1}{2}u^2\Big|_0^1 + \int_0^1 bu^2 dt \geq \frac{1}{2}u(0)^2 + \beta\|u\|_{L^2}^2,$$

so a satisfies the coercivity condition with $\alpha = \min\{\frac{1}{2}, \beta\}$. The existence of solutions v now follows from the theorem. Further, we may say that at least one of the solutions satisfies

$$\|v\|_{L^2} \leq \frac{1}{\alpha}\sqrt{\|g\|_{L^2}^2 + v_0^2}.$$

It can be shown that the solution is unique (exercise).

2.4 Approximation theory

We wish to find $u \in U(u_0)$ such that $a(u,v) = f(v)$ for all $v \in V$, where $U \leq X$ is a Banach space, $u_0 \in X$, $a : U \times V \to \mathbb{R}$ is bilinear, and $f \in V^*$. The *Galerkin approximation* involves choosing $U_h \leq U$ and $V_h \leq V$, and finding $u_h \in U_h(u_{0h})$ such that $a(u_h, v_h) = f(v_h)$ for all $v_h \in V_h$. Typically u_{0h} is the "interpolant" of u_0.

Even if $U = V$, selecting $U_h = V_h$ sometimes can give rise to poor numerical approximations for certain problems (e.g. in convection-diffusion problems of the form $b \cdot \nabla u - \mu \Delta u = f$). A method where $U_h, V_h \leq U$ are such that $U_h \neq V_h$ is sometimes called a *Petrov-Galerkin method*. If $U_h \not\subseteq U$ or $V_h \not\subseteq V$ then the may be called a *non-conforming method*.

If we are using a conforming method, u solves the continuous problem, u_h solves the discrete problem, and $a(u - u_h, v_h) = 0$ for all $v_h \in V_h$ then this is referred to as *Galerkin orthogonality*.

2.4.1 Theorem (Cea's Lemma). *Let $U \leq X$, and V all be normed linear spaces. Suppose $a : X \times V \to \mathbb{R}$ and $F : V \to \mathbb{R}$ satisfy $|a(u,v)| \leq C\|u\|_U\|v\|_V$ and $|F(v)| \leq M\|v\|_V$ for all $u \in U$ and $v \in V$. Let $U_h \leq U$ and $V_h \leq V$ be closed subspaces for which there is a constant $\alpha_h > 0$ such that*

$$\sup_{\substack{v_h \in V_h \\ v_h \neq 0}} \frac{a(u_h, v_h)}{\|v_h\|} \geq \alpha_h \|u_h\|_U$$

for all $u_h \in U_h$. If $u \in U(u_0)$ and $u_h \in U_h(u_{0h})$ satisfy $a(u,v) = F(v)$ for all $v \in V$ and $a(u_h, v_h) = F(v_h)$ for all $v_h \in V_h$, then

$$\|u - u_h\|_U \leq \left(1 + \frac{C}{\alpha_h}\right) \inf_{w_h \in U_h(u_{0h})} \|u - w_h\|_U.$$

PROOF: Notice that, for all $v_h \in V_h$, $a(u_h, v_h) = F(v_h) = a(u, v_h)$. Let $w_h \in U_h(u_{0h})$. Then $u_h - w_h \in U_h$, so coercivity of the discrete problem shows

$$\alpha_h \|u_h - w_h\|_U \leq \sup_{\substack{v_h \in V_h \\ v_h \neq 0}} \frac{a(u_h - w_h, v_h)}{\|v_h\|_V} = \sup_{\substack{v_h \in V_h \\ v_h \neq 0}} \frac{a(u - w_h, v_h)}{\|v_h\|_V} \leq C\|u - w_h\|_U$$

Then $\|u - u_h\|_U \leq \|u - w_h\| + \|w_h - u_h\| \leq (1 + \frac{C}{\alpha_h})\|u - w_h\|_U$. □

2.4.2 Lemma (Aubin-Nitsche). *Let $U = V$ be Hilbert spaces and assume that both the continuous and discrete problems are well-posed. Also assume the following.*

- *(i) There is a Hilbert space L with a continuous, symmetric, positive bilinear form (i.e. an inner product) $\ell(\cdot,\cdot)$ defining a (semi?)norm $|\cdot|_L = \sqrt{\ell(\cdot,\cdot)}$ such that V is continuously embedded into L.*
- *(ii) There is a Banach space $Z \subseteq V$ and a constant $c_s > 0$ such that the solution ϕ_g to the adjoint problem "$a(v, \phi_g) = \ell(g, v)$ for all $v \in V$" satisfies $\|\phi_g\|_Z \leq c_s |g|_L$.*

(iii) *There is an interpolation constant* $c_i > 0$ *such that, for all* h *and all* $z \in Z$, $\inf_{v_h \in V_h} \|z - v_h\|_V \le c_i h \|z\|_Z$.

Then for all h, $|u - u_h|_L \le (c_i c_s M) h \|u - u_h\|_V$.

2.4.3 Example. In practical applications we might take $Z = H^2(\Omega)$, $V = H^1(\Omega)$, and $L = L^2(\Omega)$ if we are looking at a second order elliptic PDE such as $-\Delta u = f$.

2.5 Well-posedness of the model problem

Recall that the model problem is as follows. Let $f \in L^2(\Omega)$ and $g \in L^2(\Gamma_1)$. Find $u \in U(u_0) = \{u \in H^1(\Omega) \mid u|_{\Gamma_0} = u_0\}$ such that

$$a(u,v) := \int_\Omega k \nabla u \cdot \nabla v \, d\Omega + \int_\Omega b u v \, d\Omega = \int_\Omega f v \, d\Omega + \int_{\Gamma_1} v g \, d\Gamma =: F(v)$$

for all $v \in U = \{u \in H^1(\Omega) \mid u|_{\Gamma_0} = 0\}$.

2.5.1 Theorem. *Let* $\Omega \subseteq \mathbb{R}^d$ *be a bounded Lipschitzian domain and assume that* $\Gamma_0, \Gamma_1 \in \partial\Omega$ *are open and* $\overline{\Gamma}_0 \cup \overline{\Gamma}_1 = \partial\Omega$. *Let* $U = \{u \in H^1(\Omega) \mid u|_{\Gamma_0} = 0\}$. *Let* $k \in (L^\infty(\Omega))^{d \times d}$ *be uniformly positive definite, i.e. there is* $\gamma > 0$ *such that* $z^T k(x) z \ge \gamma |z|^2$ *for all* $z \in \mathbb{R}^d$, *for all* $x \in \Omega$. *Let* $b \in L^\infty(\Omega)$ *be non-negative.*

If $f \in L^2(\Omega)$ *and* $g \in L^2(\Gamma_1)$ *then there are constants* $C, M > 0$ *such that* $|a(u,v)| \le C \|u\|_{H^1} \|v\|_{H^1}$ *and* $|F(v)| \le M \|v\|_{H^1}$. *If either (a)* $|\Gamma_0| > 0$ *(i.e.* $\Gamma_0 \ne \emptyset$, *since it is open) or (b)* $b(x) \ge b_0 > 0$, *then there is* $\alpha > 0$ *such that* $a(u,u) \ge \alpha \|u\|_{H^1}^2$ *for all* $u \in U$, *i.e.* a *is coercive.*

PROOF: Continuity of a:

$$\begin{aligned} |a(u,v)| &\le \int_\Omega (|k|_{\ell^2} |\nabla u||\nabla v| + |b||u||v|) d\Omega \\ &\le \max\{\|k\|_{L^\infty}, \|b\|_{L^\infty}\} \int_\Omega (|\nabla u||\nabla v| + |u||v|) d\Omega \\ &\le \max\{\|k\|_{L^\infty}, \|b\|_{L^\infty}\} \|u\|_{H^1} \|v\|_{H^1} \end{aligned}$$

Continuity of F:

$$\begin{aligned} |F(v)| &\le \int_\Omega |f||v| d\Omega + \int_{\Gamma_1} |v||g| d\Gamma \\ &\le \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|v\|_{L^2(\Gamma_1)} \|g\|_{L^2(\Gamma_1)} \\ &\le (\|f\|_{L^2(\Omega)} + c_t \|g\|_{L^2(\Gamma_1)}) \|v\|_{H^1} \end{aligned}$$

Coercivity of a:

$$\begin{aligned}
a(u,u) &= \int_\Omega k\nabla \cdot u \nabla u d\Omega + \int_\Omega b u^2 d\Omega \\
&= \int_\Omega (\nabla u)^T k \nabla u d\Omega + \int_\Omega b u^2 d\Omega \\
&\geq \gamma \int_\Omega |\nabla u|^2 d\Omega + \int_\Omega b u^2 d\Omega \\
&= \gamma \|\nabla u\|_{L^2}^2 + \int_\Omega b u^2 d\Omega
\end{aligned}$$

In case (a) $|\Gamma_0| > 0$, then as $b \geq 0$, we can write

$$a(u,u) \geq \gamma \|\nabla u\|_{L^2}^2 \geq \frac{\gamma}{\sqrt{1+c_P^2}} \|u\|_{H^1}$$

by Poincaré's inequality. In case (b) $b \geq b_0 > 0$, then we can write

$$a(u,u) \geq \min\{\gamma, b_0\} \|u\|_{H^1}^2. \qquad \square$$

2.6 Finite elements

As usual, let $\Omega \subseteq \mathbb{R}^d$ be a Lipschitzian domain. We would like to write the closure $\overline{\Omega} = \Omega \cup \partial\Omega$ as a union of a finite number of subsets K_j. This is often called a *triangulation*, even if the K_j are not simplices. We assume for now that $\overline{\Omega}$ is *polygonal*, i.e. that $\overline{\Omega}$ is an intersection of finitely many half-spaces. We will take each K_j to be polygonal, closed, and with non-empty interior. Further, we require that $\mathrm{int}(K_i) \cap \mathrm{int}(K_j) = \varnothing$ for $i \neq j$. Note that ∂K_j is Lipschitz because K_j is convex. Importantly, we will require that the triangulation is *face-to-face*, i.e. any face shared by two regions has the same "boundary" for both regions. "Degenerate" regions will not be allowed, so we will avoid angles near 0 and π.

2.6.1 Definition (Ciarlet). Let

(i) $K \subseteq \mathbb{R}^d$ be a bounded closed set with non-empty interior and piecewise smooth boundary, the *element domain*.

(ii) P be a finite dimensional space of functions on K, the *shape functions*.

(iii) $\mathcal{N} = \{N_1, \dots, N_k\}$ be a basis for P', the *nodal variables*.

Then (K, P, N) is called a *finite element*. The basis $\{\phi_1, \dots, \phi_k\}$ of P dual to $\mathcal{N}$ is called the *nodal basis*.

The set of points $\{a_1, \dots, a_k\} \subseteq K$ such that $N_i(\phi) = \phi(a_i)$ for all $\phi \in P$ (if there are such points) are called the *nodes*.

2.6.2 Example. Let $K = [0,1]$, P = set of linear polynomials on K, and $\mathcal{N} = \{N_0, N_0\}$, where $N_0(v) = v(0)$ and $N_0(v) = v(1)$ for all $v \in P$. Then $(K, P, \mathcal{N})$ is a finite element and the nodal basis consists of $\phi_0(x) = 1 - x$ and $\phi_0(x) = x$. This is the 1-D Lagrangian P_1 element.

In general, $K = [a, b]$, P_ℓ = set of polynomials of degree at most ℓ, and $\mathcal{N} = \{N_0, \dots, N_\ell\}$, where $N_i(v) = v(a + \frac{i}{\ell}(b-a))$ for all $v \in P_\ell$, for $i = 0, \dots, \ell$, defines a finite element.

2.7 Simplicial finite elements

In $\mathbb{R}^n$, an *n-simplex* K is the convex hull of $n+1$ points $\{a^{(0)}, \dots, a^{(n)}\}$, no three of which are collinear. Each $a^{(i)}$ is a *vertex* of the simplex. The *unit simplex* of $\mathbb{R}^n$ is the set $\{x \in \mathbb{R}^n \mid x \geq 0, x \cdot e \leq 1\}$, where e is the vector of all ones. Alternatively, the unit simplex is seen to be the simplex generated by the standard basis and the origin. Any simplex can be defined as the image of the unit simplex under a bijective affine transformation.

We denote the face opposite vertex $a^{(i)}$ by $F^{(i)}$, and the outward normal to this face by $\underline{n}^{(i)}$. For $0 \leq i \leq n$ define $\lambda_i : \mathbb{R}^n \to \mathbb{R}$ by

$$\lambda_i(x) = 1 - \frac{(x - a^{(i)}) \cdot \underline{n}^{(i)}}{(a^{(j)} - a^{(i)}) \cdot \underline{n}^{(i)}},$$

where $a(j)$ is an arbitrary vertex on the face $F^{(i)}$ (it turns out that λ_i does not depend on the choice of $a^{(j)}$). The λ_i are the *barycentric coördinates* of x with respect to the $a^{(i)}$. Note that λ_i is an affine function that is 1 at $a^{(i)}$ and 0 on $F^{(i)}$, and its level sets are hyperplanes parallel to $F^{(i)}$. We can also define the λ_i as the solution to the linear system

$$\begin{bmatrix} a_1^{(0)} & a_1^{(1)} & \dots & a_1^{(n)} \\ \vdots & \vdots & & \vdots \\ a_n^{(0)} & a_n^{(1)} & \dots & a_n^{(n)} \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_{n-1} \\ \lambda_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$$

The λ_i satisfy the following.

(i) $0 \leq \lambda_i(x) \leq 1$ if and only if $x \in K$. If x is on $F^{(i)}$ then $\lambda_i(x) = 0$. If x is in the interior of K then $0 < \lambda_i < 1$.

(ii) For all $x \in \mathbb{R}^n$, $\sum_{i=0}^n \lambda_i(x) = 1$.

(iii) $\lambda_i(a^{(j)}) = \delta_{ij}$ for all $0 \leq i, j \leq n$.

(iv) The *barycentre* or *centre of mass* of K has barycentric coördinates

$$\Big(\underbrace{\frac{1}{n+1}, \dots, \frac{1}{n+1}}_{n+1}\Big).$$

For the unit 2-simplex, defined by $\{(0,0),(1,0),(0,1)\}$, $\lambda_0 = 1 - x_1 - x_2$, $\lambda_1 = x_1$, and $\lambda_2 = x_2$.

Let $K \subseteq \mathbb{R}^n$ be a polygon and define $P_\ell(K)$ to be the collection of polynomials in n variables of degree at most ℓ, i.e.

$$P_\ell(K) = \left\{ p(x) = \sum_{\substack{0 \le i_1,\dots,i_n \le \ell \\ i_1+\cdots+i_n \le \ell}} \alpha_{i_1,\dots,i_n} x_1^{i_1} \cdots x_n^{i_n} \mid \alpha_{i_1,\dots,i_n} \in \mathbb{R}, x \in K \right\}.$$

It can be shown that $\dim P_\ell(K) = \binom{n+\ell}{\ell}$. The number of degrees of freedom per element increases rapidly with the degree of the approximations. Continuity restricts the degrees of freedom somewhat. In the field of finite element method researchers, the "KISS" principle is obeyed, so $\ell > 2$ is rarely seen.

2.7.1 Proposition. *Let $K \subseteq \mathbb{R}^n$ be a simplex and let $P = P_\ell(K)$ for some $\ell \ge 1$, and let $k = \dim P$. Consider the set of nodes $\{a^{(j)}\}_{j=1}^k$ with barycentric coördinates $(i_0/\ell, \dots, i_n/\ell)$, with $0 \le i_0, \dots, i_n \le \ell$ and $i_1 + \cdots + i_n = \ell$. Let $\Sigma = \{\sigma_1, \dots, \sigma_k\}$ be the linear forms such that $\sigma_j(p) = p(a^{(j)})$ for $1 \le j \le k$. Then (K, P, Σ) is a (Lagrange) finite element.*

2.7.2 Example ($n = 2$, $\ell = 1$). In this case $k = 3$ and the nodes are

$$\{(1,0,0),(0,1,0),(0,0,1)\}.$$

A basis for $P_1(K)$ is $\{\ell_i, 1 \le i \le n+1\}$.

2.7.3 Example ($n = 2$, $\ell = 2$). In this case $k = 6$ and the nodes are

$$\{(\tfrac{1}{2},0,0),(0,\tfrac{1}{2},0),(0,0,\tfrac{1}{2}),(\tfrac{1}{2},\tfrac{1}{2},0),(0,\tfrac{1}{2},\tfrac{1}{2}),(\tfrac{1}{2},0,\tfrac{1}{2})\}.$$

A basis for $P_2(K)$ is

$$\begin{cases} \lambda_i(2\lambda_i - 1) & 1 \le i \le n+1 \\ 4\lambda_i\lambda_j & 1 \le i < j \le n+1 \end{cases}$$

2.7.4 Example ($\ell = 3$). A basis for $P_3(K)$ is

$$\begin{cases} \frac{1}{2}\lambda_i(3\lambda_i - 1)(3\lambda_i - 2) & 1 \le i \le n+1 \\ \frac{9}{2}\lambda_i(3\lambda_i - 1)\lambda_j & 1 \le i,j \le n+1 \text{ and } i \ne j \\ 27\lambda_i\lambda_j\lambda_k & 1 \le i < j < k \le n+1 \end{cases}$$

2.8 Piecewise linear functions on triangles, $n = 2$

Given a triangulation $\mathcal{T}_h$ of Ω, the simplest finite element subspace of $H^1(\Omega)$ is the space of continuous, piecewise linear functions

$$U_h := \{u_h \in C(\overline{\Omega}) \mid u_h|_K \in P_1(K) \text{ for all } K \in \mathcal{T}_h\}.$$

Define $\{\phi_i\}_{i=1}^3$ to be the basis such that $\phi_i(a^{(j)}) = \delta_{i,j}$ on each $K \in \mathcal{T}_h$. Then we can write $u_h|_K = \sum_{i=1}^3 u_i\phi_i$ for any $u_h \in U_h$.

2.8.1 Lemma. *Let $\mathscr{T}_h$ be a triangulation of Ω and assign to each vertex of $\mathscr{T}_h$ a real value. Then the function $u_h : \Omega \to \mathbb{R}$ constructed by piecewise linear extension of the vertex values to the simplices of $\mathscr{T}_h$ is continuous on $\overline{\Omega}$.*

2.8.2 Lemma. *Let $\Omega \subseteq \mathbb{R}^d$ and suppose $\overline{\Omega} = \bigcup_{i=1}^n \overline{\Omega}_i$, where $\{\Omega_i\}_{i=1}^n$ are pairwise disjoint open subsets of Ω. Suppose further that each Ω_i satisfies the regularity assumptions of Gauss's divergence theorem. If $u \in C(\overline{\Omega})$ and $u|_{\Omega_i} \in H^1(\Omega_i)$ then $u \in H^1(\Omega)$.*

PROOF: Recall that $H^1(\Omega) = \{u \in L^2 \mid \nabla u \in (L^2)^d\}$. If u is smooth and $\phi \in C_c^\infty(\Omega)$ then $\int_\Omega \phi \nabla u dx = -\int_\Omega u \nabla \phi dx$. Motivated by this, we say that $\nabla u \in (L^2)^d$ if there is $p \in (L^2)^d$ such that $\int_\Omega \phi p dx = -\int_\Omega u \nabla \phi dx$ for all $\phi \in C_c^\infty(\Omega)$, and we write $\nabla u = p$.

Let $\phi \in C_c^\infty(\Omega)$ and put $u_i = u|_{\Omega_i} \in H^1(\Omega_i)$. Then

$$\begin{aligned}
\int_\Omega u \nabla \phi dx &= \sum_{i=1}^n \int_{\Omega_i} u_i \nabla \phi dx \\
&= \sum_{i=1}^n \left(\int_{\Omega_i} -\phi \nabla u_i dx + \int_{\partial \Omega_i} u \phi \underline{n}_i d\hat{x} \right) \\
&= \int_\Omega -\phi p dx + \sum_{i=1}^n \int_{\partial \Omega_i} u \phi \underline{n}_i d\hat{x}
\end{aligned}$$

where $p \in (L^2)^d$ is a function satisfying $p|_{\Omega_i} = \nabla u_i$. Since ϕ vanishes on $\partial\Omega$, only the portions of $\partial\Omega_i$ common with some $\partial\Omega_j$, $i \neq j$, contribute to the boundary integral. Since $\underline{n}_{ij} = -\underline{n}_{ji}$,

$$\sum_{i=1}^n \int_{\partial \Omega_i} u \phi \underline{n}_i d\hat{x} = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{\partial\Omega_i \cap \partial\Omega_j} u \phi \underline{n}_{ij} d\hat{x} = \sum_{1 \le i < j \le n} \int_{\partial\Omega_i \cap \partial\Omega_j} u \phi (\underline{n}_{ij} + n_{ji}) d\hat{x} = 0$$

□

2.9 Missed 3 lectures

2.10 Finite element meshes

Let N_e be the number of elements. Recall that $\mathscr{T}_h = \{K_j : j = 1, \dots, N_e\}$, and $\bigcup_{j=1}^{N_e} K_j = \overline{\Omega}$, with $K_M^\circ \cap K_n^\circ = \varnothing$ if $n \neq m$.

For $k \in \mathscr{T}_h$, define $h_K = \operatorname{diam}(K) = \max_{x,y \in K} \|x - y\|_2$, the *diameter* of the element and $h = \max_{K \in \mathscr{T}_h} h_K$, the *mesh size*. A triangulation $\mathscr{T}_h$ is said to be *geometrically conformal* if, for all K_m, K_n having non-empty $(d-1)$-dimensional intersection $F = K_m \cap K_n$, there is a face $\hat{F}$ of $\hat{K}$ and there are renumberings of the vertices of K_m and K_n, respectively, such that $T_m|_{\hat{F}} = T_n|_{\hat{F}}$ (and in particular, $F = T_m(\hat{F}) = T_n(\hat{F})$).

Let $\mathscr{T}_h$ be geometrically conformal with no holes. Let N_v, N_e, and N_f be the numbers of elements, faces, and vertices, respectively. Then in 2 dimensions we have the *Euler relations* $N_e - N_f + N_v = 1$, $N_v^\partial - N_f^\partial = 0$, and $2N_f - N_f^\partial = 3N_e$. (Note that these are related to *Euler's formula* $v - e + f = 2$ for polyhedra, but the notation is different!)

What can we expect a mesh generator to give us? Three things:

- coördinates of the vertices (`x_coord`);
- vertices of each simplex (`node`); and
- boundary elements (`bdry_node(Nbf, 2)`).

and sometimes also a "neighbours" array. Once we have this data we need to do some post-processing to add extra nodes, say.

The *aspect ratio* of a triangle is the radius of the circumscribed circle, R_K, divided by the radius of the inscribed circle, ρ_K. It is a measure of how "well proportioned" is the triangle.

2.10.1 Lemma. *Let* $J_K = \frac{\partial T_K}{\partial \hat{x}}$*, the Jacobian of the reference mapping.*

(i) $|\det(J_K)| = |K|/|\hat{K}|$;
(ii) $\|J_K\| \leq h_K/\rho_K$;
(iii) $\|J_K^{-1}\| \leq h_{\hat{K}}/\rho_K$.

A mesh is said to be a *quasi-uniform mesh* if there are constants β_1 and β_2 such that, for all $K \in \mathscr{T}_h$, $\beta_1 h \leq h_K \leq \beta_2 \rho_K$. Of course, we can always find β_1 and β_2 for any given mesh (consisting of a finite number of elements). A *quasi-uniform family* of meshes is a set of meshes that use the same constants as $h \to 0$.

2.11 Rectangular elements

Sometimes the domain $\Omega \subseteq \mathbb{R}^d$ has a natural Cartesian (grid) structure which allows for simple decomposition into quadrilaterals ($d = 2$) or hexahedra ($d = 3$). Take the rectangle $\{x \in \mathbb{R}^2 : \|x\|_\infty \leq 1\}$ to be the reference element, with vertices $\hat{a}^{(i)}$ numbered counterclockwise from the lower left corner, in coördinates $\hat{x}_1$ and $\hat{x}_2$. Let $x = T(\hat{x})$ be the coördinates of the problem element K. We seek basis functions on K which interpolate at the vertices: $\phi_i(a^{(j)}) = \delta_{i,j}$. A natural choice is $\phi_i(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 + a_3(\text{quad})$. The form of the quadratic term is dictated by the constraint that the basis functions must be globally continuous. If two elements share an edge then they share exactly two vertices, so the restriction of each basis function to the shared edge must be linear to guarantee that it is continuous across the border. Since our rectangles are axis-aligned, the quadratic term must be $x_1 x_2$. The basis on the reference element $\hat{K}$ is hence

$$\begin{aligned}
\hat{\phi}_1 &= \tfrac{1}{4}(1 - \hat{x}_1)(1 - \hat{x}_2) \\
\hat{\phi}_2 &= \tfrac{1}{4}(1 + \hat{x}_1)(1 - \hat{x}_2) \\
\hat{\phi}_3 &= \tfrac{1}{4}(1 + \hat{x}_1)(1 + \hat{x}_2) \\
\hat{\phi}_4 &= \tfrac{1}{4}(1 - \hat{x}_1)(1 + \hat{x}_2).
\end{aligned}$$

which has the desire property, that $\hat{\phi}_i(\hat{a}^{(j)}) = \delta_{i,j}$. These basis functions can be viewed as the tensor product of the one dimensional basis functions $\frac{1}{2}(1 \pm x)$. If $K = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2]$ then $K = T(\hat{K})$, where

$$T \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} x_1^c \\ x_2^c \end{bmatrix} + \begin{bmatrix} \frac{1}{2}(\alpha_2 - \alpha_1) & 0 \\ 0 & \frac{1}{2}(\beta_2 - \beta_1) \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}.$$

Unlike rectangular K, general quadrilateral K may not be the image of $\hat{K}$ under an affine transformation, as $T : \hat{K} \to \mathbb{R}^2$ is uniquely determined by the image of three points. So we relax the requirement that T be affine. Since $\hat{\phi}_i(\hat{a}^{(j)}) = \delta_{i,j}$, T maps vertices of $\hat{K}$ to vertices of K. Provided K is convex, we will have $\phi_i(a^{(j)}) = \delta_{i,j}$ and $\phi_i|_{k \subset \partial K}$ is linear. This implies ϕ_i(global basis) will be continuous and in $H^1(\Omega)$. T contains products of the from $\hat{x}_1 \hat{x}_2$, so its inverse will not be a polynomial in general, thus $\phi_i = \hat{\phi}_i \circ T^{-1}$ are not polynomials. This does not cause any problems though, as we only compute using $\hat{\phi}_i$ and T.

Basis functions of degree k on $[-1,1]^d$ can be constructed as tensor products of the one dimensional interpolating polynomials of degree k on $[-1,1]$. Let Q_k denote the set $\{u_k \in P_{d,k}([-1,1]^d) \mid$ maximum degree of any one variable is no greater than $k\}$. Interpolation theory will show that extra variables in Q_k do not increase the rate of convergence, so little is gained from the computational cost due to extra variables. *Serendipity elements* are quadrilateral elements that locate all of the their interpolation points on $\partial \hat{K}$.

2.12 Interpolation in Sobolev spaces

The abstract approximation theory shows that Galerkin approximations $u_h \in U_h$ of $u \in U$ satisfy an estimate of the form

$$\|u - u_h\|_U \in\leq \left(1 + \frac{c}{\alpha_h}\right) \inf_{w_h \in U_h} \|u - w_h\|_U.$$

Our goal is to construct an interpolant $w_h = I_h u \in U_h$ and estimate $\|u - w_h\|_U$.

Recall Taylor's theorem. If $u : [-1,1] \to \mathbb{R}$ is sufficiently smooth then for any $x \in [-1,1]$, $u(x) = p_k(x) + \frac{1}{(k+1)!} u^{(k+1)}(\xi) x^{k+1}$ where $|\xi| \leq |x|$ and p_k is the polynomial $p_k(x) = u(0) + u'(0)x + \cdots + \frac{1}{k!} u^{(k)}(0) x^k$. Yet otherwise said,

$$\inf_{p \in P_k([-1,1])} \|u - p\|_\infty \leq c \|u^{(k+1)}\|_\infty.$$

In multiple dimensions, in multi-index notation, Taylor's theorem may be stated

$$u(x+h) = \sum_{i=0}^{k} \sum_{|\alpha|=i} \frac{1}{\alpha!} D^\alpha u(x) h^\alpha + O(|h|^{k+1}).$$

It is a fact that if $u : \Omega \to \mathbb{R}$ satisfies $D^\alpha u = 0$ a.e. for all α with $|\alpha| = k+1$ then $u \in P_k(\Omega)$. Recall that $H^k(\Omega) = \{u \in L^2(\Omega) \mid D^\alpha u \in L^2$ for all $|\alpha| \leq k\}$ is a Hilbert space with inner product

$$(u,v)_{H^k} = \sum_{|\alpha|\le k} (D^\alpha u, D^\alpha v)_{L^2}.$$

The norm is given by this inner product, and the H^k semi-norm is defined by

$$|u|_{H^k}^2 = \sum_{|\alpha|=k} (D^\alpha u, D^\alpha v)_{L^2} =: |u|_k^2.$$

Let $N = \dim P_k(\Omega)$ and let $\{q_n\}_{n=1}^N$ be an orthonormal basis for $P_k(\Omega)$ (orthonormal with respect to the L^2-inner product). Then $p \in P_k(\Omega)$ can be written as $p(x) = \sum_{n=1}^N a_n q_n(x)$, where $a_n = (p, q_k)_{L^2}$. Let $\{\ell_n\}_{n=1}^N$ be a set of continuous linear functional on $L^2(\Omega)$ defined by $\ell_n(u) = (u, q_n)_{L^2}$. Clearly $|\ell_n(u)| \le \|u\|_{L^2} \le \|u\|_{H^k}$. So the ℓ_n are continuous on all of the $H^k(\Omega)$, $k \ge 0$. We also have the following properties.

(i) If $\ell_n(p) = 0$ for $n = 1, 2, \dots, N$ and $p \in P_k(\Omega)$ then $p = 0$.

(ii) If $u \in L^2(\Omega)$ then there is $p \in P_k(\Omega)$ satisfying $\ell_n(p) = \ell_n(u)$ for all $n = 1, \dots, N$ (namely p is the orthogonal projection of u onto $P_k(\Omega)$).

2.12.1 Theorem (J.-L. Lions). *Let $\Omega \subseteq \mathbb{R}^d$ be a bounded, Lipschitz domain. Let $\{\ell_n\}_{n=1}^N$ be a set of continuous linear functionals on $H^{k+1}(\Omega)$ satisfying (i) and (ii) above. Then there is a constant $C = C(\Omega, k)$ (depending only upon Ω and k) such that for $u \in H^{k+1}(\Omega)$,*

$$\|u\|_{k+1} \le C\left(|u|_{k+1} + \sum_{n=1}^N |\ell_n(u)| \right).$$

2.12.2 Corollary (Bramble-Hilbert Lemma). *Let V be a normed linear space and $\Pi : H^{k+1}(\Omega) \to V$ be a continuous linear mapping such that $\Pi(p) = 0$ for all $p \in P_k(\Omega)$. Then for $u \in H^{k+1}(\Omega)$, $\|\Pi(u)\|_V \le \|\Pi\| C(\Omega, k) |u|_{k+1}$.*

Proof (of Corollary): Let $\{\ell_n\}_{n=1}^N$ be a set of linear functionals satisfying the hypotheses of Lion's theorem. Fix $u \in H^{k+1}(\Omega)$ and let $p \in P_k(\Omega)$ satisfy $\ell_n(p) = \ell_n(u)$ for $n = 1, \dots, N$. Then

$$\begin{aligned} \|\Pi(u)\|_V = \|\Pi(u-p)\|_V &\le \|\Pi\| \|u-p\|_{k+1} \\ &\le \|\Pi\| C(\Omega,k)\left(|u-p|_{k+1} + \sum_{n=1}^n |\ell_n(u-p)| \right) \\ &\le \|\Pi\| C(\Omega,k) |u|_{k+1} \end{aligned}$$

The last line uses linearity of ℓ_n and the triangle inequality. □

PROOF (OF THEOREM): Note that $H^k(\Omega)$ is a reflexive Banach space, so bounded subsets are weakly sequentially precompact. Let $(\cdot,\cdot)_{k+1}$ denote the semi-inner product on $H^{k+1}(\Omega)$, i.e.

$$(u,v)_{k+1} = \sum_{|\alpha|=k+1} \int_\Omega D^\alpha u D^\alpha v\, dx.$$

Then $f(v) := (u,v)_{k+1}$ is a continuous linear functional on H^{k+1}. If $\{u_i\}$ converges weakly to u then

$$|u|_{k+1}^2 = (u,u)_{k+1} = \lim_{i\to\infty}(u,u_i)_{k+1} \le \liminf_i |u|_{k+1}|u_i|_{k+1}$$

so $|u|_{k+1} \le \liminf_i |u_i|_{k+1}$. The $(k+1)$-seminorm is said to be *lower semicontinuous*.

Suppose for contradiction that there is no such constant. Then there is a sequence $\{u_n\}_{n=1}^\infty \subseteq H^{k+1}(\Omega)$ such that $\|u_n\|_{k+1} = 1$ for all $n \ge 1$ and $|u_n|_{k+1} + \sum_{i=1}^N \ell_i(u_n)| \to 0$ as $n \to \infty$. Suppose without loss of generality that the sequence converges weakly to a limit $u \in H^{k+1}(\Omega)$ and converges strongly in $H^k(\Omega)$. (The latter happens because the embedding of H^{k+1} into H^k is compact.) Lower semicontinuity of the $(k+1)$-seminorm implies that $|u|_{k+1} \le \liminf_n |u_n|_{k+1} = 0$, so $D^\alpha u = 0$ for all α with $|\alpha| = k+1$. Thus $u \in P_k(\Omega)$. Then

$$\|u-u_n\|_{k+1}^2 = |u-u_n|_{k+1}^2 + \|u-u_n\|_k^2 = |u_n|_{k+1}^2 + \|u-u_n\|_k^2 \to 0$$

so $u_n \to u$ in H^{k+1}. Since each $\ell_i : H^{k+1} \to \mathbb{R}$ is continuous, it follows that $\ell_i(u) = \lim_{n\to\infty} \ell_i(u_n) = 0$. Since $u \in P_k(\Omega)$, by the property of the ℓ_i, $u = 0$. But then $0 = \|u\|_{k+1} = \lim_{n\to\infty}\|u_n\|_{k+1} = 1$ a contradiction. □

We pause to introduce some notation. Let $u : K \to \mathbb{R}$ be smooth and m be a non-negative integer. Define

$$D^m u(x) : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{m} \to \mathbb{R} \text{ by } D^m u(x)(y^{(1)},\dots,y^{(m)}) = \sum_{i_1,\dots,i_m=1}^d \frac{\partial^m u(x)}{\partial x_{i_1}\cdots\partial x_{i_m}} y_{i_1}^{(1)}\cdots y_{i_m}^{(m)}$$

The "natural norm" is

$$|D^m u(x)| = \sup_{y^{(1)},\dots,y^{(m)}\neq 0} \frac{D^m u(x)(y^{(1)},\dots,y^{(m)})}{|y^{(1)}|\cdots|y^{(m)}|}$$

2.12.3 Example. $D^1 u(x) y = \nabla u(x)\cdot y =$ the directional derivative of u evaluated at x in the direction y. $D^2 u(x)(y,z) = y^T H(x) z$, where H is the Hessian. Here $|D^2 u(x)|$ is the usual ℓ^2 operator norm of a matrix.

In this notation, Taylor's theorem states

$$u(x+h) = u(x) + Du(x)h + \frac{1}{2!}D^2(x)(h,h) + \cdots + \frac{1}{m!}D^m u(x)(h,\dots,h) + O(|h|^{m+1})$$

The space of m-multilinear forms $\mathbb{R}^d$ is finite dimensional so there are constants $0 < c_m < C_m$ such that

$$c_m|D^m u(x)|^2 \le \sum_{|\alpha|=m} |D^\alpha u(x)|^2 \le C_m|D^m u(x)|^2.$$

It follows that $(\sum_{m=0}^{k} \|D^m u\|_{L^2}^2)^{1/2}$ is a norm equivalent to the H^k norm.

2.12.4 Lemma. *Let $T : \hat{K} \to K$ be an invertible affine map, so that $T(\hat{x}) = x^{(0)} + B\hat{x}$, where $B \in \mathbb{R}^{d\times d}$ is nonsingular. Let $u : K \to \mathbb{R}$ be smooth and define $\hat{u} : \hat{K} \to \mathbb{R}$ by $\hat{u} = u \circ T$. If $x = T(\hat{x})$ then*

$$D^m\hat{u}(\hat{x})(y^{(1)}, \dots, y^{(m)}) = D^m u(x)(By^{(1)}, \dots, By^{(m)}).$$

2.12.5 Corollary. $|D^m\hat{u}(\hat{x})| \le |B|^m |D^m u(x)|$ *and* $|D^m u(x)| \le |B^{-1}|^m |D^m\hat{u}(\hat{x})|$.

2.12.6 Corollary. $\|D^m\hat{u}\|_{L^2(\hat{K})} \le (|B|^m/\sqrt{\det B})\|D^m u\|_{L^2(K)}$ *and* $\|D^m u\|_{L^2(K)} \le |B^{-1}|^m\sqrt{\det B}\|D^m\hat{u}\|_{L^2(\hat{K})}$.

Proof: $x = T(\hat{x})$, so

$$\begin{aligned}
\|D^m u\|_{L^2(K)}^2 &= \int_K |D^m u(x)|^2 dx \\
&= \int_{\hat{K}} |D^m u(T^{-1}x)|^2 \det(B) d\hat{x} \\
&= \int_{\hat{K}} |D^m\hat{u}(\hat{x})|^2 |B^{-1}|^{2m} \det(B) d\hat{x} \\
&\le |B|^{2m} \det(B) \|D^m\hat{u}\|_{L^2(\hat{K})}^2 \qquad \square
\end{aligned}$$

Note that $\det(B) = |K|/|\hat{K}|$.

2.12.7 Lemma. *Let K and $\hat{K}$ be bounded domains in $\mathbb{R}^d$ and suppose $T : \hat{K} \to K$ is invertible, affine, and $T(\hat{x}) = x^{(0)} + B\hat{x}$. Let h_K ($\hat{h}$) be the diameter of K ($\hat{K}$) and ρ_K ($\hat{\rho}$) be the radius of the largest inscribed sphere within K ($\hat{K}$). Then $|B| \le h_K/\hat{\rho}$ and $|B^{-1}| \le \hat{h}/\rho_K$.*

Proof:

$$\begin{aligned}
|B| = \sup_{y\ne 0} \frac{|By|}{|y|} &= \frac{1}{\hat{\rho}} \sup_{|y|=\hat{\rho}} |By| \\
&= \frac{1}{\hat{\rho}} \sup_{|\hat{x}-\hat{x}(0)|=\hat{\rho}} |B(\hat{x} - \hat{x}(0)| \\
&= \frac{1}{\hat{\rho}} \sup_{|\hat{x}-\hat{x}(0)|=\hat{\rho}} |T\hat{x} - T\hat{x}(0)| \\
&\le \frac{\operatorname{diam}(K)}{\hat{\rho}} = \frac{h_K}{\hat{\rho}} \qquad \square
\end{aligned}$$

Recall the definition of *finite element*: $(\hat{K}, \hat{P}, \hat{N})$ with $\hat{K} \subseteq \mathbb{R}^d$, $\hat{P}$ a finite dimensional space of *shape functions* with basis $\{\hat{\phi}_i\}_{i=1}^m$, and $\hat{N} = \{\hat{x}^{(i)}\}_{i=1}^m$ the *nodal basis*, dual to the basis of $\hat{P}$.

For Lagrange finite elements the nodal basis are the interpolation points. In this case let $\hat{I} : C(\hat{K}) \to \hat{P}$ be defined by $\hat{I}u = \sum_{i=1}^m \hat{\phi}_i u(\hat{x}^{(i)})$. An arbitrary finite element (K, P, N) is *affine equivalent* to the *reference element* $(\hat{K}, \hat{P}, \hat{N})$ if there is an invertible affine map $T : \mathbb{R}^n \to \mathbb{R}^d$ for which $T(\hat{K}) = K$, $P = P(K) = \{\hat{u} \circ T^{-1} \mid \hat{u} \in \hat{P}\}$, and $N = N(K) = \{T(\hat{x}) \mid \hat{x} \in \hat{N}\}$. This implies $\hat{\phi}_i \circ T^{-1}(x^{(j)}) = \hat{\phi}_i(\hat{x}^{(j)}) = \delta_{i,j}$ if $x^{(j)} \in N(K)$, so $\phi_i = \hat{\phi}_i \circ T^{-1}$.

2.12.8 Definition. Let $\hat{K}$ and K be domains in $\mathbb{R}^d$ and $T : \hat{K} \to K$ be a homeomorphism. Then $\hat{\ }: C(K) \to C(\hat{K})$ is the mapping $u \mapsto u \circ T$.

2.12.9 Lemma. *Let $(\hat{K}, \hat{P}, \hat{N})$ and (K, P, N) be affine equivalent under $T : \hat{K} \to K$ and let $I_K : C(K) \to P(K)$ and $\hat{I} : C(\hat{K}) \to \hat{P}$ be their interpolation operators. The following diagram commutes.*

$$\begin{array}{ccc} C(K) & \xrightarrow{\ \hat{}\ } & C(\hat{K}) \\ \downarrow{\scriptstyle I_K} & & \downarrow{\scriptstyle \hat{I}} \\ P(K) & \xrightarrow{\ \hat{}\ } & \hat{P} \end{array}$$

PROOF: Let $u \in C(K)$ and recall that $x^{(i)} = T(\hat{x}^{(i)})$.

$$\begin{aligned} \widehat{I_K u} &= \sum_{i=1}^m \widehat{\phi_i u(x^{(i)})} \\ &= \sum_{i=1}^m (\phi_i \circ T) u(x^{(i)}) \\ &= \sum_{i=1}^m (\phi_i \circ T)(u \circ T)(\hat{x}^{(i)}) \\ &= \sum_{i=1}^m \hat{\phi}_i \hat{u}(\hat{x}^{(i)}) = \hat{I}\hat{u} \end{aligned}$$ □

$|u - I_K u|_{H^m(K)}$.

2.12.10 Theorem. *Let $(\hat{K}, \hat{P}, \hat{N})$ and $(K, P(K), N(K))$ be affine equivalent and let $I_K : C(K) \to P(K)$ be the interpolation operator onto $P(K)$. If $k \geq 1$ is an integer and $P_k(K) \subseteq P(K)$ then there is $C = C(\hat{K}, k)$ such that, for $0 \leq m \leq k+1$,*

$$|u - I_K u|_{H^m(K)} \leq C \left(\frac{h_K^{k+1}}{\rho_K^m} \right) |u|_{H^{k+1}(K)}|.$$

Proof: Let $T : \hat{K} \to K$ be affine and invertible, $T(\hat{x}) = x^{(0)} + B\hat{x}$. Then we have

$$|u - I_K u|_{H^m(K)} \le (|B^{-1}|^m \sqrt{\det(B)})|\widehat{u - I_K u}|_{H^m(\hat{K})} \qquad 2.12.6$$
$$\le (|B^{-1}|^m \sqrt{\det(B)})|\hat{u} - \hat{I}\hat{u}|_{H^m(\hat{K})} \qquad 2.12.7$$

Let $\hat{\Pi} : H^{k+1}(\hat{K}) \to H^m(\hat{K})$ be defined by $\hat{\Pi}(\hat{u}) = \hat{u} - \hat{I}\hat{u}$. Since $\hat{I}\hat{u}_h = \hat{u}_h$ for all $\hat{u}_h \in \hat{P}$ and $P_k(\hat{K}) \subseteq \hat{P}$, it follows that $\hat{\Pi}(\hat{p}) = 0$ for all $p \in P_k(\hat{K})$. The Brambel-Hilbert lemma asserts that there is $C = C(\hat{K}, k)$ such that

$$\|\hat{u} - \hat{I}\hat{u}\|_{H^m(\hat{K})} = \|\hat{\Pi}(\hat{u})\|_{H^m(\hat{K})} \le C\|\hat{\Pi}\|_{\mathscr{L}}|\hat{u}|_{k+1}.$$

Therefore

$$\begin{aligned}|u - I_K u|_{H^m(K)} &\le C(|B^{-1}|^m \sqrt{\det(B)})\|\hat{\Pi}\|_{\mathscr{L}}|\hat{u}|_{k+1}\\ &\le C\|\hat{\Pi}\|_{\mathscr{L}}|B^{-1}|^m|B|^{k+1}|\hat{u}|_{k+1} \qquad 2.12.6\\ &\le C\|\hat{\Pi}\|_{\mathscr{L}}\left(\frac{h_K}{\hat{\rho}}\right)^m\left(\frac{\hat{h}}{\hat{\rho}_K}\right)^{k+1}|u|_{H^{k+1}(K)}|\\ &= C\|\hat{\Pi}\|_{\mathscr{L}}\left(\frac{h_K^{k+1}}{\rho_K^m}\right)|u|_{H^{k+1}(K)}|\\ &\le C\left(\frac{h_K^{k+1}}{\rho_K^m}\right)|u|_{H^{k+1}(K)}|\end{aligned}$$

since $\|\hat{\Pi}\|_{\mathscr{L}}$ is a finite number that depends on...

Since $\hat{\Pi}(\hat{u}) = \hat{u} - \hat{I}\hat{u}$ and $m \le k+1$ it suffices to show that $\|\hat{I}\|_{\mathscr{L}}$ is finite.

$$\begin{aligned}\|\hat{I}\hat{u}\|_{\mathscr{L}} &= \left\|\sum_{i=1}^m \hat{\phi}_i \hat{u}(\hat{x}^{(i)})\right\|_{H^m(K)}\\ &\le \sum_{i=1}^m \|\hat{\phi}_i\|_{H^m(K)}|\hat{u}(\hat{x}^{(i)})|\\ &\le \left(\sum_{i=1}^m \|\hat{\phi}_i\|_{H^m(K)}\right)\|\hat{u}\|_{C(\hat{K})}\\ &\le C\|\hat{u}\|_{C(\hat{K})}\end{aligned}$$

The Sobolev embedding theorem states that there is $c > 0$ such that $\|\hat{u}\|_{C(\hat{K})} \le c\|u\|_{H^2(\hat{K})}$ and the hypothesis that $k \ge 1$ ensures that $H^2(\hat{K}) \hookrightarrow H^{k+1}(\hat{K})$. ☐

2.12.11 Theorem (Sobolev embedding). *$H^k(\Omega) \hookrightarrow C^s(\overline{\Omega})$, where $s = \lfloor k - \frac{d}{2} \rfloor$, i.e. $\max_{x \in \overline{\Omega}} |D^\alpha u(x)| \le c\|u\|_k$ for all α with $|\alpha| \le s$.*

In dimensions $d = 2$ and $d = 3$ we have $H^k(\Omega) \hookrightarrow C^{k-2}(\overline{\Omega})$.

2.12.12 Definition. The *aspect ratio* of $K \subseteq \mathbb{R}^d$ is h_K/ρ_K. A family $\{\tau_h\}_{h>0}$ of triangulations of $\overline{\Omega} \subseteq \mathbb{R}^d$ with the diameter of K at most h for each $K \in \tau_h$ is a *regular triangularization* if there is $\sigma > 0$ such that $h_K/\rho_K < \sigma$ for all $K \in \bigcup_{h>0} \tau_h$. $\{(K, P(K), N(K)) \mid K \in \tau_h|\}_{h>0}$ is an *affine family* if each element is affine equivalent to the same *reference element* $(\hat{K}, \hat{P}, \hat{N})$. For each $h > 0$, $I_h : C(\overline{\Omega}) \to L^\infty(\Omega)$ is defined by $I_h(u|_K) = I_K u$, $K \in \tau_h$ and $U_h = I_h(C(\overline{\Omega}))$.

2.12.13 Corollary. *If $u_h \in H^k(\Omega)$ and $P_k(\hat{K}) \subseteq \hat{P}$ for an integer $k \geq 1$ then there is $C > 0$, independent of u and h, such that, for all $0 \leq m \leq k+1$,*

$$\|u - I_h u\|_{H^m(\Omega)} \leq \left(\sum_{K \in \tau_h} h_K^{2(k+1-m)} |u|^2_{H^{k+1}(K)} \right)^{\frac{1}{2}} \leq C h^{k+1-m} |u|_{k+1}.$$

If $u_h \in H^\ell(\Omega)$ then estimates hold for $0 \leq m \leq \min(\ell, k+1)$.

$$\|u - u_h\|_{H^m(\Omega)} \leq \inf_{w_h \in U_h} \|u - w_h\|_{H^m(\Omega)}$$

We need $k \geq 1$ in order to guarantee that $H^{k+1}(\Omega) \hookrightarrow C(\hat{K})$, so that $\hat{I}$ is well-defined. This excludes the important case of estimating $\inf_{w_h \in U_h} \|u - w_h\|_{L^2(\Omega)}$. This was considered by Clément who constructed $\tilde{I}_h : H^1(\Omega) \to U_h$ satisfying $\|u - \tilde{I}_h u\|_{H^m(\Omega)} \leq c h^{1-m} |u|_{H^1(\Omega)}$ for $m = 0$ or $m = 1$. If U_h is consists of piecewise constant (hence in particular discontinuous), and $k = 0$ then $\overline{I}_h : L^2(\Omega) \to U_h$ is defined by $\overline{I}_h u|_K = \frac{1}{|K|} \int_K u$ satisfies $\|u - \overline{I}_h u\|_{L^2} \leq ch\|u\|_1$.

2.12.14 Example. Let $u(x, y) = x^2$ on the triangle $(0, \varepsilon)$, $(\pm h/2), 0)$. The gradient of an affine function is constant so

$$\begin{aligned} \frac{\partial u_h}{\partial y} &= \frac{1}{\varepsilon}(u_h(0,\varepsilon) - u_h(0,0)) \\ &= \frac{1}{\varepsilon}(u_h(0,\varepsilon) - \frac{1}{2}(u(-h/2,0) - u(h/2,0))) \\ &= -\frac{h^2}{2\varepsilon} \end{aligned}$$

As $\varepsilon \to 0$ then $\frac{\partial u_h}{\partial y} \to \infty$, but $u_y = 0$.

Inverse inequality $\|u_h\|_{H^1} \leq \frac{C}{h} \|u_h\|_{L^2}$.

3 Parabolic problems

3.1 Introduction

Find $u : [0, \pi] \times (0, T) \to \mathbb{R}$ such that $\dot{u} - u_x x = f$ for $0 < x < \pi$, $t \in (0, T)$, $u(0, t) = u(\pi, t) = 0$ for all t and $u(x, 0) = u^0(x)$ for on $(0, \pi)$. If $f = 0$ then we can construct a solution using separation of variables, giving a solution in terms of Fourier series

$$u(x,t)=\sum_{j=1}^{\infty}u_j^0 e^{-j^2 t}\sin(jx)$$

where $u_j^0 = \sqrt{\frac{2}{\pi}}\int_0^\pi u^0(x)\sin(jx)dx$ for $j=1,2,\ldots$. The solution is an infinite sum of sine waves, with frequencies j and amplitude $u_j^0 e^{-j^2 x}$. Each component $\sin(jx)$ lives on a timescale of $O(-j^2)$ since $e^{-j^2 t}$ is small for $j^2 t$ moderately large. However, we may have $\|\dot{u}(t)\|_{L^2(0,\pi)} \to \infty$ as $t \to 0$. The size of the derivatives of u for small t will depend on how quickly the Fourier coefficients decay with increasing j.

3.1.1 Example. If $u^0(x)=\pi - x$ then $u_j^0 = \frac{c}{j}$, so $\|\dot{u}(t)\|_{L^2} \sim Ct^{-\frac{3}{4}}$ as $t \to 0$.

If $u^0(x)=\min\{x,\pi-x\}$ then $u_j^0=\frac{c}{j^2}$, so $\|\dot{u}(t)\|_{L^2} \sim Ct^{-\frac{1}{4}}$ as $t \to 0$.

An initial phase for small t where certain derivatives of u are large is called an *initial transient*. The basic stability estimates are $\|u(t)\|_{L^2} \le \|u^0\|_{L^2}$ for all $t\in(0,T)$ and $\|\dot{u}(t)\|_{L^2} \le \frac{C}{t}\|u^0\|_{L^2}$ for all $t \in (0,T)$.

3.2 Semi-discrete formulation

Let $\Omega \subseteq \mathbb{R}^2$, $T>0$, $I=(0,T)$. Find $u:\Omega\times I \to \mathbb{R}$ satisfying $u_t - \Delta u = f$ in $\Omega \times I$, $u=0$ on $\partial\Omega\times I$, and $u(\cdot,0)=u^0$ on Ω.

Let $U=H_0^1(\Omega)$. Then we have: find $u(t)\in U$, $t\in I$, such that

$$\int_\Omega \dot{u}(t)v\,d\Omega + \int_\Omega \nabla u(t)\cdot\nabla v\,d\Omega = \int_\Omega f v\,d\Omega$$

for all $v\in U$, $t\in I$. This is essentially $(\dot{u}(t),v)+a(u(t),v)=(f,v)$ in terms of the L^2 inner product, with $u(0)=u^0$. The *semi-discrete problem* is as follows. Let $U_h \subseteq U$ be finite dimensional with basis $\{\phi_1,\ldots,\phi_m\}$. Find $u_h(t)\in U_h$, $t\in I$, such that

$$(\dot{u}_h(t),v_h)+a(u_h(t),v_h)=(f,v_h) \text{ for all } v_h\in U_h, t\in I$$

and $(u_h(0),v_h)=(u^0,v_h)$ for all $v_h\in U_h$ (i.e. $u_h(0)$ is the L^2-projection of u^0 on U_h).

Write $u_h(x,t)=\sum_{i=1}^m u_i(t)\phi_i(x)$. Then

$$\sum_{i=1}^m \dot{u}_i(t)(\phi_i,\phi_j)+\sum_{i=1}^m u_i(t)a(\phi_i,\phi_j)=(f,\phi_j) \qquad j=1,\ldots,m.$$

This is of the form $B\dot{\underline{u}}(t)+A\underline{u}(t)=F(t)$, $t\in I$, with $B\underline{u}(0)=\underline{u}^0$, where $B_{ij}=\int_\Omega \phi_i\phi_j$, $A_{ij}=\int_\Omega \nabla\phi_i\cdot\nabla\phi_j$, and $F_j(t)=(f(t),\phi_j)$.

B is symmetric positive definite, so we write $B = E^T E$ (Cholesky decomposition). Let $\underline{\tilde{u}} = E\underline{u}$. Then we have $\dot{\underline{\tilde{u}}}(t) + \tilde{A}\underline{\tilde{u}} = g(t)$ for $t \in I$ and $\underline{\tilde{u}}(0) = \tilde{u}^0$, where $\tilde{A} = E^{-T}AE^{-1}$, $g = E^{-T}F$, and $\tilde{u}^0 = E^{-T}u^0$. The solution is

$$\tilde{u}(t) = e^{-\tilde{A}t}u^0 + \int_0^t e^{-\tilde{A}(t-s)}g(s)ds.$$

Unfortunately, this is a stiff system.

If we set $v = u_h(t)$ in the semi-descretized problem then

$$\begin{aligned}
(\dot{u}_h(t), u_h(t)) + a(u_h(t), u_h(t)) &= (f, u_h(t)) \\
\frac{1}{2}\frac{d}{dt}\|u_h(t)\|_{L^2}^2 + a(u_h(t), u_h(t)) &\le \|f\|_{L^2}\|u_h(t)\|_{L^2} \\
\text{so } \|u_h(t)\|_{L^2}\|\frac{d}{dt}\|u_h(t)\|_{L^2} &\le \|f\|_{L^2}\|u_h(t)\|_{L^2}
\end{aligned}$$

Whence $\frac{d}{dt}\|u_h(t)\|_{L^2} \le \|f\|_{L^2}$, so integrating with respect to time,

$$\|u_h(t)\|_{L^2} \le \|u_h(0)\|_{L^2} + \int_0^t \|f(s)\|_{L^2}ds$$

3.2.1 Theorem. *Let $u(t)$ satisfy the weak formulation and $u_h(t)$ satisfy the semi-discrete weak formulation. Then for all $t \ge 0$,*

$$\|u(t) - u_h(t)\|_{L^2} \le \|u^0 - u_h(0)\|_{L^2} + Ch^2\left(\|u^0\|_{H^2} + \int_0^t \|\dot{u}(s)\|_{H^2}ds\right).$$

PROOF: Let $R_h : H_0^1(\Omega) \to U_h$ be the *Ritz projection operator* defined by $a(R_h u, v) = a(u, v)$ for all $v \in U_h$. Then it can be shown (and we will do so later) that $\|u - R_h u\|_{L^2} \le Ch^2\|u\|_{H^2}$.

Let $t \in (0, T)$ and $\tilde{u}_h(t) = R_h u(t)$.

$$u(t) - u_h(t) = \underbrace{u(t) - \tilde{u}_h(t)}_{\eta(t)} + \underbrace{\tilde{u}_h(t) - u_h(t)}_{\phi_h(t)}$$

Now by the property of the Ritz projection,

$$\|\eta(t)\|_{L^2} = \|u(t) - \tilde{u}_h(t)\|_{L^2} \le Ch^2\|u(t)\|_{H^2}.$$

Since u and u_h satisfy the weak formulations, for all $v \in U_h$,

$$\begin{aligned}
(\dot{u}(t) - \dot{u}_h(t), v) + a(u(t) - u_h(t), v) &= 0 \\
(\dot{\phi}_h(t) - \dot{\eta}(t), v) + a(\phi_h(t) - \eta(t), v) &= 0 \\
(\dot{\phi}_h(t), v) + a(\phi_h(t), v) &= -(\dot{\eta}(t), v) - a(\eta(t), v) \\
(\dot{\phi}_h(t), v) + a(\phi_h(t), v) &= -(\dot{\eta}(t), v)
\end{aligned}$$

Thus, by the stability estimate,

$$\|\phi_h(t)\|_{L^2} \leq \|\phi_h(0)\|_{L^2} + \int_0^t \|\dot{\eta}(s)\|_{L^2} ds.$$

And so forth... □

The error estimate for the semi-discrete parabolic problem is a consequence of two things: the error estimate for the elliptic problem, and the stability estimate. Loosely stated, the *Lax principle* is that stability plus consistency (i.e. small spatial discretization error) equals convergence.

3.3 Discretization in space and time

Recall the *backward Euler method* for $y' = f(t, y)$, obtain the next iterate by solving the equation

$$\frac{y^{n+1} - y^n}{\Delta t} = f(t^{n+1}, y^{n+1}).$$

This method is $O(\Delta t)$. Applied to our problem, given u_h^n, find u_h^{n+1} satisfying

$$\left(\frac{u_h^{n+1} - u_h^n}{\Delta t}, v\right) + a(u_h^{n+1}, v) = (f(t^{n+1}), v)$$

for all $v \in U_h$. For the initial iterate we require $(u_h^0, v) = (u^0, v)$ for all $v \in U_h$. Rewriting,

$$(u_h^{n+1}, v) + \Delta t a(u_h^{n+1}, v) = (u_h^n, v) + \Delta t (f(t^{n+1}), v)$$
$$(B + \Delta t A)\underline{u}^{n+1} = B\underline{u}^n + \Delta t \underline{F}^{n+1}$$

3.4 Missed 3 lectures

Index

$H^1(0,1)$, 6
$L^2(0,1)$, 6

affine equivalent, 37
affine family, 39
aspect ratio, 32, 39
assembly, 10

backward Euler method, 42
barycentre, 29
barycentric coördinates, 29

centre of mass, 29

diameter, 31
diffusion equation, 22
diffusivity matrix, 22
discrete variational problem, 8
Divergence Theorem, 21

element domain, 28
element matrix, 10
element vector, 10
Euler relations, 32
Euler's formula, 32

face-to-face, 28
finite element, 28, 37

Galerkin approach, 4
Galerkin approximation, 26
Galerkin method, 8
Galerkin orthogonality, 26
general quadrature rule, 11
geometrically conformal, 31
geometry, 10
Gramian, 20
Green's Theorem, 21

heat flux, 22

initial transient, 40
Inverse inequality, 39

Lax principle, 42
lower semicontinuous, 35

mass matrix, 20
mesh points, 3
mesh size, 31
mesh spacing, 3
method of manufactured solutions, 17
minimization approach, 4

negative Sobolev spaces, 21
nodal basis, 28, 37
nodal variables, 28
nodes, 28
non-conforming, 8
non-conforming method, 26

parent element, 10
Petrov-Galerkin method, 26
polygonal, 28

quadrature point, 11
quasi-uniform family, 32
quasi-uniform mesh, 32

Rayleigh-Ritz approach, 4
reference element, 10, 37, 39
regular triangularization, 39
Riesz map, 22
Ritz projection operator, 41

semi-discrete problem, 40
Serendipity elements, 33
shape functions, 28, 37
simplex, 29
steady-state, 22
stiffness matrix, 20
strong form, 3

test space, 8
trial space, 8
triangulation, 28

unit simplex, 29

vertex, 29

weak approach, 4

weak formulation, 22
weight, 11
well-posed, 8