

Discretization of the Poisson
Problem in \mathbb{R}^1 : Theory and
Implementation

April 7 & 9, 2003

1 Theory

1.1 Goals

1.1.1 *A priori*

A priori error estimates:
bound various “measures”
of u [exact] – u_h [approximate];
in terms of $C(\Omega, \text{problem parameters})$,
 h [mesh diameter], and u .

N1

SLIDE 1

Note 1

A priori theory

Clearly, since *a priori* estimates will be expressed in terms of the *unknown* exact solution, u , they are not useful in determining *in practice* whether u_h is accurate enough. *A priori* estimates are, however, useful to compare different discretizations (which converge faster in which norms? which is more efficient?), to understand what conditions must be satisfied for rapid convergence (is u smooth enough?), and to understand if a method has been properly implemented (for a test problem, does $u_h \rightarrow u$ at the correct rate?)

SLIDE 2

u : $-u_{xx} = f, u(0) = u(1) = 0$

$$a(u, v) = \ell(v), \quad \forall v \in X$$
$$a(w, v) = \int_0^1 w_x v_x dx, \quad \ell(v) = \int_0^1 f v dx$$
$$X = \{v \in H^1(\Omega) \mid v(0) = v(1) = 0\}$$

Recall that $\ell(v)$ can in fact be more general — any linear functional in $H^{-1}(\Omega)$, that is, any linear functional which satisfies $|\ell(v)| \leq C \|v\|_{H^1(\Omega)}$ for any $v \in H_0^1(\Omega)$. For example, $\ell(v) = \langle \delta_{x_0}, v \rangle = v(x_0)$ is admissible.

SLIDE 3

u_h :

$$a(u_h, v) = \ell(v), \quad \forall v \in X_h$$
$$a(w, v) = \int_0^1 w_x v_x dx, \quad \ell(v) = \int_0^1 f v dx$$
$$X_h = \{v \in X \mid v|_{T_h} \in \mathbb{P}_1(T_h), \quad \forall T_h \in \mathcal{T}_h\}$$

In fact, the theory presented applies equally well to the Neumann problem and (at least in \mathbb{R}^1) the inhomogeneous Dirichlet case.

1.1.2 *A posteriori*

SLIDE 4

A posteriori error estimates:

N2

bound various “measures”
of u [exact] – u_h [approximate];
in terms of $C(\Omega, \text{problem parameters})$,
 h [mesh diameter], and u_h .

Note 2

A posteriori theory

A posteriori error estimates are arguably more useful than *a priori* estimates since we *know* u_h . Bear in mind, however, that (i) in most methods for *a posteriori* error estimation the constants C are *not* known, and (ii) for those methods which do attempt to better quantify the constants C , additional computational effort is required. Nevertheless, *a posteriori* error analysis is an increasingly important aspect of finite element practice: even when the C are not known precisely, local estimators can provide guidance as to how best to refine a triangulation. We shall restrict attention in these lectures to the simpler case of *a priori* estimates.

1.2 Projection

We need several concepts to make the subsequent analysis flow smoothly: projection (general) and interpolation (specific to our particular space X_h).

1.2.1 Definition

SLIDE 5

Given Hilbert spaces Y and $Z \subset Y$,

$$\underbrace{(\Pi y, v)}_{\in Z} = \underbrace{(y, v)}_{\in Y}, \quad \forall v \in Z$$

defines the *projection* of y onto Z , Πy ;

$$\Pi: Y \rightarrow Z .$$

1.2.2 Property

SLIDE 6

The projection Πy minimizes $\|y - z\|_Y^2$, $\forall z \in Z$.

Why?

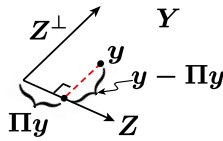
$$\begin{aligned} \|y - \underbrace{(\Pi y + v)}_{\text{any } z \in Z}\|_Y^2 &= ((y - \Pi y) - v, (y - \Pi y) - v)_Y \\ &= \|y - \Pi y\|_Y^2 - 2 \underbrace{(y - \Pi y, v)_Y}_{0: v \in Z} + \|v\|_Y^2, \quad \forall v \in Z. \end{aligned}$$

Note $z = \Pi y + v \in Z$ and $\Pi y \in Z$ implies $v \in Z$, and hence since $(\Pi y, v)_Y = (y, v)_Y$ for all $v \in Z$, $(y - \Pi y, v)_Y = 0$. The above result states that $\|y - \Pi y\|_Y^2 < \|y - z\|_Y^2$ for all $z \neq \Pi y$. In words, Πy is the best approximation in Z of y in the $\|\cdot\|_Y$ norm.

1.2.3 Geometry

SLIDE 7

Geometry of projection:



Orthogonality: $(y - \Pi y, v)_Y = 0$, $\forall v \in Z$. E1

Not surprisingly, if we wish to find the $z = \Pi y$ on the Z axis closest to y , $y - \Pi y$ should be perpendicular to the Z axis — in the $(\cdot, \cdot)_Y$ inner product. This analogue to our usual notion of projection in \mathbb{R}^n should be self-evident. In the above picture, Z^\perp is the orthogonal complement of Z in Y : the space of all members of Y orthogonal to all members of Z .

▷ **Exercise 1**

- (a) Show that $\|\Pi y\|_Y \leq \|y\|_Y$ and $\|y - \Pi y\|_Y \leq \|y\|_Y$, and interpret this result geometrically.
- (b) Show that $\Pi(\Pi y) = \Pi y$.

■

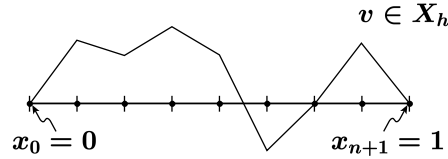
1.3 The Interpolant

1.3.1 Definition

SLIDE 8

Recall

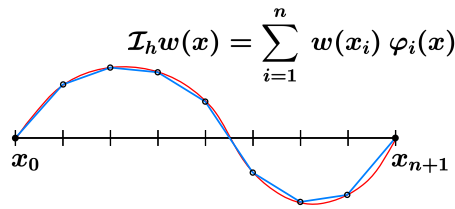
$$X_h = \{v \in X \mid v|_{T_h} \in \mathbb{P}_1(T_h), \quad \forall T_h \in \mathcal{T}_h\}$$



SLIDE 9

Given $w \in X$, the *interpolant* $\mathcal{I}_h w$ satisfies:

$$\mathcal{I}_h w \in X_h; \text{ and } \mathcal{I}_h w(x_i) = w(x_i), \quad i = 0, \dots, n+1.$$



1.3.2 Approximation Theory

SLIDE 10

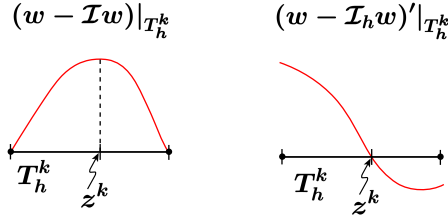
If $w \in X$, and $w|_{T_h} \in C^2(T_h)$, $\forall T_h \in \mathcal{T}_h$, then

$$\begin{aligned} \|w - \mathcal{I}_h w\|_{H^1(\Omega)} &\leq h \max_{T_h \in \mathcal{T}_h} \left(\max_{x \in T_h} |w''| \right) \\ \|w - \mathcal{I}_h w\|_{L^2(\Omega)} &\leq h^2 \max_{T_h \in \mathcal{T}_h} \left(\max_{x \in T_h} |w''| \right). \end{aligned}$$

Recall $\|v\|_{H^1(\Omega)}^2 = \int_0^1 v_x^2 dx$, $\|v\|_{L^2(\Omega)}^2 = \int_0^1 v^2 dx$,
and $\|v\|_{H^1(\Omega)}^2 = \|v\|_{H^1(\Omega)}^2 + \|v\|_{L^2(\Omega)}^2$.

SLIDE 11

Sketch of proof:



SLIDE 12

$$\begin{aligned}
 |(w - \mathcal{I}_h w)'|_{T_h^k}(x) &= \left| \int_{z^k}^x (w - \mathcal{I}_h w)''|_{T_h^k} dx \right| = \left| \int_{z^k}^x w'' dx \right| \\
 &\leq h \max_{x \in T_h^k} |w''| \\
 \sum_{k=1}^K \int_{T_h^k} (w - \mathcal{I}_h w)'|_{T_h^k}^2 dx &\leq \frac{1}{h} h \left(h \max_{k=1, \dots, K} \max_{x \in T_h^k} |w''| \right)^2
 \end{aligned}$$

E2

The first line follows from Rolle's Theorem (which requires $w|_{T_h} \in C^1(T_h)$, as is the case here). The second line bounds the $K = \frac{1}{h}$ integrals by $h \times$ the maximum of the integrand. Note, however, that we only require w'' to be defined in the elements, not at the nodes, so if we place our delta distribution loads at nodes, this hypothesis is still satisfied for solutions u of our Poisson problem even in this case.

Since $(\mathcal{I}_h w)'|_{T_h^k}$ is a constant, we are effectively approximating w' by a constant. Not surprisingly, this will not work very well if w' has jumps (w'' infinite) in T_h^k ; also, the larger the w'' , the larger the error, since the more w' will vary away from a constant. (In general, if w has strong singularities, $|w - \mathcal{I}_h w|_{H^1(\Omega)}$ will only converge as some fractional power of h .)

▷ **Exercise 2** Prove the L^2 estimate of Slide 10. *Hint:* write $(w - \mathcal{I}_h w)|_{T_h^k}$ as a definite integral in terms of $(w - \mathcal{I}_h w)'|_{T_h^k}$; then express $(w - \mathcal{I}_h w)'|_{T_h^k}$ as in the H^1 -seminorm proof. ■

SLIDE 13

If $w \in X$, and $w \in H^2(\Omega, \mathcal{T}_h)$,

$$\begin{aligned}
 \|w - \mathcal{I}_h w\|_{H^1(\Omega)} &\leq \frac{h}{\pi} \|w\|_{H^2(\Omega, \mathcal{T}_h)} \\
 \|w - \mathcal{I}_h w\|_{L^2(\Omega)} &\leq \frac{h^2}{\pi^2} \|w\|_{H^2(\Omega, \mathcal{T}_h)},
 \end{aligned}$$

where
$$\|w\|_{H^2(\Omega, \mathcal{T}_h)}^2 \equiv \sum_{k=1}^K \|w\|_{H^2(T_h^k)}^2 = \sum_{k=1}^K \int_{T_h^k} w_{xx}^2 + w_x^2 + w^2 dx .$$

Note again that jumps in the derivative (e.g., due to a delta distribution or change in conductivity) at nodes are fine — the function is still in H^2 over each element. (In fact, the above result is true with just the H^2 seminorm.) Norms which have been broken up over elements or subdomains are sometimes known as “broken” norms.

The proof of the above is not difficult, but involves the Rayleigh quotient for a fourth-order eigenvalue problem. As we have not introduced these concepts, the demonstration would require a major digression at this stage. The reader is referred to pages 45–47 of Strang & Fix. Note we prefer this second result to that of Slide 10 since the norm on u is “weaker,” and consistent with the energy notions that underly the finite element method.

1.4 Error: Energy Norm

1.4.1 Definition

SLIDE 14

Define the energy, or “ a ”, norm $|||v|||$ as

$$\begin{aligned} |||v|||^2 &= a(v, v) && \text{(generally)} \\ &= \int_0^1 v_x^2 dx = |v|_{H^1(\Omega)}^2 && \text{(here) .} \end{aligned}$$

Note: $||| \cdot |||$ is *problem-dependent*.

Since $a(\cdot, \cdot)$ is an SPD bilinear form, $(a(v, v))^{1/2}$ does indeed satisfy all the requirements of a proper norm. (Recall that the H^1 seminorm is in fact a norm over $H_0^1(\Omega)$.)

SLIDE 15

Of interest: for

$u(x)$ (exact solution)

$u_h(x)$ (finite element approximation)

$\Rightarrow e(x) = (u - u_h)(x)$ (discretization error)

find bound for $|||e|||$ in terms of h, u .

1.4.2 Orthogonality

SLIDE 16

Since $a(u, v) = \ell(v), \forall v \in X$

then

$$\boxed{\begin{aligned} a(u, v) &= \ell(v), \quad \forall v \in X_h && (X_h \subset X), \\ - [a(u_h, v) &= \ell(v)], \quad \forall v \in X_h \end{aligned}}$$

but

so

$$\boxed{a(u - u_h, v) = 0, \quad \forall v \in X_h} \quad (\text{bilinearity}).$$

1.4.3 General Bound

SLIDE 17

For any $w_h = u_h + v_h \in X_h$,

$v_h \in X_h$

$$\begin{aligned} \underbrace{a(u - w_h, u - w_h)}_{\|u - w_h\|^2} &= a((u - u_h) - v_h, (u - u_h) - v_h) \\ &= \underbrace{a(u - u_h, u - u_h)}_{\|e\|^2} - \underbrace{2a(u - u_h, v_h)}_{0: \text{orthogonality}} + \underbrace{a(v_h, v_h)}_{>0 \text{ if } v_h \neq 0} \end{aligned}$$

\Rightarrow

$$\boxed{\|e\| = \inf_{w_h \in X_h} \|u - w_h\| .}$$

SLIDE 18

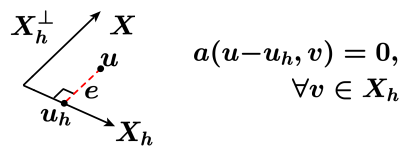
In words: even if you *knew* u ,
you could not find a w_h in X_h
more accurate than u_h

in the energy norm.

So we see that the finite element procedure does as well without knowledge of the exact solution as you can do with knowledge of the exact solution — so long as we speak of the energy (or “a”) norm. The finite element method has transformed the problem of discretization of a PDE into a problem of approximation.

SLIDE 19

Geometry



$$a(u - u_h, v) = 0, \quad \forall v \in X_h$$

$\Rightarrow u_h = \Pi_h^a u$: the projection of (closest point to)

u on X_h in the a norm.

SLIDE 20

Miracle ? : $a(\underbrace{\Pi_h^a u}_{u_h}, v) = a(u, v), \forall v \in X_h$;

but we do not know $u \dots$

NO: $a(u, v) = \underbrace{\ell(v)}_{\text{can evaluate}} \Rightarrow a(\underbrace{\Pi_h^a u}_{u_h}, v) = \ell(v), \forall v \in X_h$.

Only in the energy inner product can we

compute $\Pi_h u$ without knowing u . N3

Note 3

Generality of abstract result

We note that our bound

$$|||e||| = \inf_{w_h \in X_h} |||u - w_h||| ,$$

that is, that $u_h = \Pi_h^a u$, is in fact true for any SPD bilinear form a , and any boundary conditions, and any finite element space X_h , and any space dimension.

For any particular SPD problem (that is, any linear problem for which the bilinear form a in the weak formulation is SPD), the only thing that changes is the definition of the norm; for our particular problem $|||e||| = |e|_{H^1(\Omega)}$, though in general that will not be the case.

Obviously, however, we are not yet quite done; we must understand how

$$\inf_{w_h \in X_h} |||u - w_h|||$$

depends on h , the smoothness of u , and the parameters of the problem. For that we need to introduce the particulars of our finite element approximation space.

1.4.4 Particular Bound

SLIDE 21

We know $|u - \mathcal{I}_h u|_{H^1(\Omega)} \leq \frac{h}{\pi} \|u\|_{H^2(\Omega, \mathcal{T}_h)}$.

Thus

$$\begin{aligned} |||e||| &= \inf_{w_h \in X_h} |||u - w_h||| \leq |||u - \mathcal{I}_h u||| \\ &= |u - \mathcal{I}_h u|_{H^1(\Omega)} \leq \frac{h}{\pi} \|u\|_{H^2(\Omega, \mathcal{T}_h)} \quad \boxed{\text{E3}} \quad \boxed{\text{N4}} \end{aligned}$$

(assuming $\|u\|_{H^2(\Omega, \mathcal{T}_h)}$ finite).

We would, of course, prefer to directly use the projection u_h for w_h , rather than the interpolant. However, the latter is much easier to work with, and will, in general, yield the correct h dependence. In fact, for our particular problem, $u_h = \mathcal{I}_h u$ (see Exercise 3), but this is a bit of a “coincidence.”

We say the above estimate is “optimal” in the sense that the power of h can not be improved — there exist problems (in fact, almost all problems) for which $\|e\|$ decreases no faster than h . (The term “sharp” is usually reserved for the case in which, for some problem, the bound obtains with strict equality — that is not the case here, though we could tighten things up a bit to ensure sharpness.)

All the above requires essentially no modification for the Neumann problem.

▷ **Exercise 3** Show that, for our particular problem, $u_h = \mathcal{I}_h u$. *Hint:* Show that $a(u - \mathcal{I}_h u, v) = 0, \forall v \in X_h$, by integration by parts over each element. (We do not dwell on this “miracle” — nodal superconvergence — since it is rather special to $-u_{xx} = f, \mathbb{R}^1$, and exact quadrature.) ■

Note 4

Convergence rate and smoothness

First, the above estimate tells us that u_h converges to u (at least in the a norm). Second, it tells us that it converges as h . Third, it tells us that u must be sufficiently smooth — finite in the broken H^2 norm, $\|\cdot\|_{H^2(\Omega, \mathcal{T}_h)}$ — to achieve this convergence rate.

It is important to recognize that although we exploit the weak form to look for *finite element approximations* u_h that are only in $H^1(\Omega)$, we do require additional smoothness on the part of the *exact solution* u if we are to obtain rapid convergence. Furthermore, as we consider higher order finite elements, we will require additional smoothness to achieve the best convergence rates: for example, for quadratic finite elements, $\|e\| \leq C h^2 \|u\|_{H^3(\Omega)}$ — a higher power of h , but also a higher norm of u .

1.5 Error: H^1 Norm

1.5.1 Reminders

The H^1 norm:

$$\begin{aligned} \|v\|_{H^1(\Omega)}^2 &= |v|_{H^1(\Omega)}^2 + \|v\|_{L^2(\Omega)}^2 \\ &= \int_0^1 v_x^2 dx + \int_0^1 v^2 dx ; \end{aligned}$$

$\|e\|_{H^1(\Omega)}$ measures e and e_x .

SLIDE 22

SLIDE 23

Coercivity of $a(\cdot, \cdot)$:

$$\exists \alpha > 0 \text{ such that } a(v, v) \geq \alpha \|v\|_{H^1(\Omega)}^2, \quad \forall v \in X$$

$$\left(\int_0^1 v_x^2 dx \geq \alpha \left(\int_0^1 v_x^2 dx + \int_0^1 v^2 dx \right) \right).$$

(Recall this Poincaré-Friedrichs inequality follows from the fact that $v \in X$ satisfy $v(0) = v(1) = 0$. For our problem, $\alpha = \frac{1}{2}$ “works,” as can be shown (say) by considering the Rayleigh quotient.)

Continuity of $a(\cdot, \cdot)$:

$$\exists \beta (= 1) > 0 \text{ such that } a(w, v) \leq \beta \|w\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}.$$

(Recall this is derived from the Cauchy-Schwarz inequality.)

1.5.2 General Result

SLIDE 24

The error $e = u - u_h$ satisfies

$$\|e\|_{H^1(\Omega)} \leq \underbrace{\left(1 + \frac{\beta}{\alpha}\right)}_{\text{degradation}} \underbrace{\inf_{w \in X_h} \|u - w_h\|_{H^1(\Omega)}}_{\text{error in } H^1 \text{ projection of } u \text{ on } X_h};$$

in general u_h is *not* the H^1 projection of u on X_h .

E4 N5

▷ **Exercise 4** For what particular problem (give the strong form) is u_h the H^1 projection of u on X_h ? ■

Note 5 Proof of H^1 norm general bound (Optional)

To begin, we note that for any $w_h \in X_h$,

$$\begin{aligned} \alpha \|u_h - w_h\|_{H^1(\Omega)}^2 &\leq a(u_h - w_h, u_h - w_h) && \text{(coercivity)} \\ &= a(u_h - w_h + (u - u_h), u_h - w_h) && \text{(orthogonality)} \\ &= a(u - w_h, u_h - w_h) \\ &\leq \beta \|u - w_h\|_{H^1(\Omega)} \|u_h - w_h\|_{H^1(\Omega)} && \text{(continuity)} \end{aligned}$$

so that

$$\|u_h - w_h\|_{H^1(\Omega)} \leq \frac{\beta}{\alpha} \|u - w_h\|_{H^1(\Omega)} .$$

But then

$$\begin{aligned} \|u - u_h\|_{H^1(\Omega)} &= \|u - w_h + w_h - u_h\|_{H^1(\Omega)} \\ &\leq \|u - w_h\|_{H^1(\Omega)} + \|w_h - u_h\|_{H^1(\Omega)} \quad (\text{triangle inequality}) \\ &\leq \left(1 + \frac{\beta}{\alpha}\right) \|u - w_h\|_{H^1(\Omega)} \quad \forall w_h \in X_h ; \end{aligned}$$

and thus

$$\|u - u_h\|_{H^1(\Omega)} \leq \left(1 + \frac{\beta}{\alpha}\right) \inf_{w_h \in X_h} \|u - w_h\|_{H^1(\Omega)} ,$$

as desired. Note in this proof we in fact *did not use symmetry* — the proof applies to any linear problem for which the bilinear form a of the weak formulation is *coercive* (and continuous).

In fact, for our current case, which is also symmetric and thus has a minimization statement, we can improve this result: in the energy norm we know that

$$a(u - u_h, u - u_h) = \inf_{w_h \in X_h} a(u - w_h, u - w_h) ;$$

thus from coercivity and continuity

$$\alpha \|u - u_h\|_{H^1(\Omega)}^2 \leq \inf_{w_h \in X_h} \beta \|u - w_h\|_{H^1(\Omega)}^2$$

or

$$\|u - u_h\|_{H^1(\Omega)} \leq \sqrt{\frac{\beta}{\alpha}} \inf_{w \in X_h} \|u - w_h\|_{H^1(\Omega)} ,$$

which is sharper than our previous result since β can not be less than α (take $v = w$ in the continuity condition).

Note if we compare the above proofs to similar finite difference proofs, we see that coercivity plays the role of stability, and orthogonality the role of consistency. Together they imply convergence.

1.5.3 Particular Result

SLIDE 25

We know $\|u - \mathcal{I}_h u\|_{H^1(\Omega)} \leq \sqrt{2} \frac{h}{\pi} \|u\|_{H^2(\Omega, \tau_h)}$. Thus

The $\sqrt{2}$ is simply a sloppy bound for $(1 + \frac{h^2}{\pi^2})^{1/2}$ which arises because the H^1 norm has contributions from both the H^1 seminorm and L^2 norm of Slide 13.

$$\begin{aligned}
\|e\|_{H^1(\Omega)} &= \left(1 + \frac{\beta}{\alpha}\right) \inf_{w_h \in X_h} \|u - w_h\|_{H^1(\Omega)} \\
&\leq \left(1 + \frac{\beta}{\alpha}\right) \|u - \mathcal{I}_h u\|_{H^1(\Omega)} \\
&\leq \sqrt{2} \left(1 + \frac{\beta}{\alpha}\right) \frac{h}{\pi} \|u\|_{H^2(\Omega, \mathcal{T}_h)}.
\end{aligned}$$

The error in the H^1 norm converges at the same rate as the error in the energy norm. (This must be the case since the two norms are equivalent.)

1.6 Error: L^2 Norm

1.6.1 Reminder

SLIDE 26

The L^2 norm:

$$\|v\|_{L^2(\Omega)} = \left(\int_0^1 v^2 dx\right)^{1/2};$$

$\|e\|_{L^2(\Omega)}$ measures e .

In the L^2 -like norm, we found for finite differences that the error converged as h^2 ; given the similarity of \underline{A}_h for finite elements to the corresponding system matrix for finite differences, we expect h^2 behaviour here as well. Note the h dependence (as opposed to h^2 dependence) of the H^1 norm is not surprising — it measures the error in a stronger norm.

1.6.2 Particular Result

SLIDE 27

A General Result is possible, but not as transparent as in the other cases, so we go directly to the particular result.

The L^2 error satisfies

$$\begin{aligned}
\|e\|_{L^2(\Omega)} &\leq C h \|e\|_{H^1(\Omega)} \\
&\leq C h^2 \|u\|_{H^2(\Omega, \mathcal{T}_h)},
\end{aligned}$$

for C independent of h and u .

N6

The proof is by what is known as the “Aubin-Nitsche” trick, an application of duality. We will see it again for linear functional error estimates.

To begin, we introduce an auxiliary problem: find $\Phi \in X = H_0^1(\Omega)$ such that

$$a(v, \Phi) = \int_0^1 e v \, dx$$

where e is the error $u - u_h$. We now set $v = e$, so that

$$\begin{aligned} \|e\|_{L^2(\Omega)}^2 &= \int_0^1 e e \, dx = a(e, \Phi) \\ &= a(e, \Phi - \mathcal{I}_h \Phi) && \text{(orthogonality)} \\ &\leq \beta \|e\|_{H^1(\Omega)} \|\Phi - \mathcal{I}_h \Phi\|_{H^1(\Omega)} && \text{(continuity)} \\ &\leq \beta \|e\|_{H^1(\Omega)} \frac{h}{\pi} \|\Phi\|_{H^2(\Omega)}, \end{aligned}$$

where the last line follows from our interpolation result of Slide 10. Now we note from Slide 6 of the last lecture that, since $e \in L^2(\Omega)$ (in fact, $e \in H_0^1(\Omega)$), Φ satisfies $\|\Phi\|_{H^2(\Omega)} \leq C \|e\|_{L^2(\Omega)}$ (note the strong form for Φ is simply $-\Phi_{xx} = e$). Using this fact and dividing by $\|e\|_{L^2(\Omega)}$ gives

$$\|e\|_{L^2(\Omega)} \leq C h \|e\|_{H^1(\Omega)},$$

from which the rest directly follows. (Note C in different expressions need not be the same: C is a generic constant independent of h and u .)

Note the L^2 result appears relatively unimportant (apart from confirming our intuition). However, that is not the case: the fact that $\|e\|_{L^2(\Omega)}$ converges faster than $\|e\|_{H^1(\Omega)}$ has important ramifications in many different contexts (e.g., *a posteriori* error estimation). Our proof here needs only continuity — *not symmetry, not coercivity* — and is thus quite general, though the regularity hypothesis on Φ requires more attention in \mathbb{R}^2 .

1.7 Linear Functionals

1.7.1 Motivation

SLIDE 28

A *linear-functional “output”* s is defined by

$$s = \ell^O(u) + c^O;$$

where

$$\ell^O: H_0^1(\Omega) \rightarrow \mathbb{R}$$

is a bounded linear functional

$$|\ell^O(v)| \leq C \|v\|_{H^1(\Omega)}, \quad \forall v \in H_0^1(\Omega).$$

Strictly speaking, due to the c^O , our outputs are affine, not linear.

SLIDE 29

Very relevant: engineering quantities of interest.

For example:

s : average over $\mathcal{D} \subset \Omega$, with

$$\ell^O(v) = \int_{\mathcal{D}} v \, dx;$$

s : flux at boundary, $u_x(0)$, with

$$\ell^O(v) = - \int_0^1 (1-x)_x v_x, \quad c^O = \int_0^1 f(1-x) \, dx.$$

N7

Note 7 **Boundedness of output functionals ℓ^O (Optional)**

We shall see that it is very important in theory *and* practice that our output functionals be bounded. In the first case above, it is clear that $\ell^O \in H^{-1}(\Omega)$ (that is, is bounded); in fact, $\ell^O \in L^2(\Omega)$, which we can exploit (see subsequent slides).

In the second case, had we simply written the obvious choice $\ell^O(v) = v_x(0)$, this would not be a bounded functional for v in $H_0^1(\Omega)$. For example, if $v \sim x^{3/4}$ as $x \rightarrow 0$, v is in $H_0^1(\Omega)$, yet $v_x(0)$ is infinite. If we were to use this functional, $v_x(0)$, to compute $u_x(0)$ in practice, poor convergence would result. In contrast, it is clear that our choice of Slide 29 *is* bounded:

$$\ell^O(v) = - \int_0^1 (1-x)_x v_x \, dx = \int_0^1 v_x \, dx \leq \|v\|_{H^1(\Omega)}$$

by the Cauchy-Schwarz inequality. But does $\ell^O(u) + c^O = u_x(0)$, as desired?

To show this, we recall that u satisfies $-u_{xx} = f$, $u(0) = u(1) = 0$ (assuming f is in $L^2(\Omega)$ for simplicity). Then

$$\begin{aligned} \ell^O(u) + c^O &= - \int_0^1 (1-x)_x u_x - (1-x) f \, dx \\ &= -(1-x) u_x|_0^1 - \int (1-x)(-u_{xx} - f) \, dx = u_x(0), \end{aligned}$$

as desired. It may seem that our $\ell^O(v)$ of Slide 29 and $v_x(0)$ are thus equivalent — that is *not* the case. They both evaluate to $u_x(0)$ for $v = u$, but for general $v \in H_0^1(\Omega)$ (e.g., our finite element approximation u_h) the bounded choice behaves much better.

(Note for this particular very simple case, $\ell^O(v) = \int_0^1 v_x \, dx = 0$, and thus we can obtain the exact result $u_x(0) = \int_0^1 f(1-x) \, dx$ on any mesh (indeed,

without any numerical calculation). Of course this will not be generally true for more difficult problems.)

SLIDE 30

Of interest: $s = \ell^O(u) + c^O,$
 $s_h = \underbrace{\ell^O(u_h)}_{\text{finite element prediction of output}} + c^O;$

error in output is thus

$$\begin{aligned} |s - s_h| &= |\ell^O(u) - \ell^O(u_h)| = |\ell^O(u - u_h)| \\ &= |\ell^O(e)|. \end{aligned}$$

Recall that $e = u - u_h$; note the second step above follows from linearity of $\ell^O(v)$.

1.7.2 General Result

SLIDE 31

If $\ell^O \in H^{-1}(\Omega)$, then

$$|\ell^O(e)| \leq C \|e\|_{H^1(\Omega)} \text{ (boundedness).}$$

If $\ell^O \in L^2(\Omega)$, then

$$|\ell^O(e)| \leq C \|e\|_{L^2(\Omega)} \text{ (boundedness).}$$

SLIDE 32

In fact: for any $\ell^O \in H^{-1}(\Omega)$,

$$|\ell^O(e)| \leq C \|e\|_{H^1(\Omega)} \|\psi - \psi_h\|_{H^1(\Omega)}$$

where

$$a(v, \psi) = -\ell^O(v), \quad \forall v \in X \quad \boxed{\text{N8}}$$

$$a(v, \psi_h) = -\ell^O(v), \quad \forall v \in X_h,$$

and ψ is an adjoint, or dual, variable.

Note 8

Role of adjoint (Optional)

The variable ψ above is denoted an adjoint, or dual, variable, though we will not dwell on the meaning of the term here. We note only that the strong form for ψ is $-\psi_{xx} = -\ell^O$, which can be viewed as our original equation but now with the output functional as data; in the case of a *nonsymmetric* operator, ψ would be acted upon by the dual, or transpose, operator, in which the signs of the odd derivatives would be flipped relative to the original equation.

Quite apart from what we call ψ , the proof of the above result is simple. Since $e \in X$, we have that

$$\ell^{\mathcal{O}}(e) = a(e, \psi) .$$

But from orthogonality, $a(e, \psi) = a(e, \psi - \psi_h)$, since $a(e, \psi_h) = 0$ as $\psi_h \in X_h$ (see Slide 16). Thus, by continuity,

$$|\ell^{\mathcal{O}}(e)| \leq C \|e\|_{H^1(\Omega)} \|\psi - \psi_h\|_{H^1(\Omega)} ,$$

as advertised. (This applies even to nonsymmetric operators, and even noncoercive operators, since we have been careful to put ψ as the *second* argument of $a(\cdot, \cdot)$.)

1.7.3 Particular Result

SLIDE 33

From our earlier bounds for $\|e\|_{H^1(\Omega)}$ and $\|e\|_{L^2(\Omega)}$ for linear finite elements:

$$\text{for } \ell^{\mathcal{O}} \in H^{-1}(\Omega): \quad |\ell^{\mathcal{O}}(e)| \leq C h \|u\|_{H^2(\Omega, \mathcal{T}_h)}$$

$$\text{for } \ell^{\mathcal{O}} \in L^2(\Omega): \quad |\ell^{\mathcal{O}}(e)| \leq C h^2 \|u\|_{H^2(\Omega, \mathcal{T}_h)} .$$

Better yet: for $\ell^{\mathcal{O}} \in H^{-1}(\Omega)$

$$|\ell^{\mathcal{O}}(e)| \leq C \boxed{h^2} \|u\|_{H^2(\Omega, \mathcal{T}_h)} \|\psi\|_{H^2(\Omega, \mathcal{T}_h)} .$$

In this last step we simply apply our H^1 error estimate to the dual problem. Note this dual problem never appears in practice (at least in the a priori context) — it is only used in the theory to demonstrate the h^2 convergence rate. (It can also be used to prove superconvergence.) In practice, the h^2 convergence rate is very attractive — the quantities of interest converge quickly.

Note in the above we require that u be in the broken H^2 norm, and also that ψ be in the broken H^2 norm, to obtain the h^2 convergence rate. This will be the case if the distributional part of $\ell^{\mathcal{O}}$ is restricted to nodes.

Finally, we remark that it is a common criticism of finite element error analysis that the norms considered are mathematically convenient but practically irrelevant. This last slide proves that this is patently not the case.

2 Implementation

2.1 Overview

SLIDE 34

Four steps:

- A Proto-Problem,
- Elemental Quantities;
- Assembly;
- Boundary Conditions;

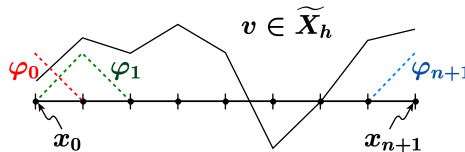
and Quadrature.

2.2 A Proto-Problem

2.2.1 Space and Basis

SLIDE 35

Let $\tilde{X}_h = \{v \in H^1(\Omega) \mid v|_{T_h} \in \mathbb{P}_1(T_h), \forall T_h \in \mathcal{T}_h\}$
 $= \text{span} \{\varphi_0, \dots, \varphi_{n+1}\}.$



We now include φ_0 and φ_{n+1} in our basis since v need not be zero at $x = 0$ and $x = 1$. Any $v \in \tilde{X}_h$ can be represented as

$$v(x) = \sum_{i=0}^{n+1} v(x_i) \varphi_i(x) ;$$

$\dim(\tilde{X}_h) = n + 2$ (two more than $\dim(X_h)$).

2.2.2 Definition

SLIDE 36

“Find” $\tilde{u}_h \in \tilde{X}_h$ such that

$$a(\tilde{u}_h, v) = \ell(v), \quad \forall v \in \tilde{X}_h .$$

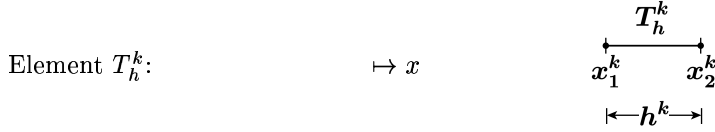
We never actually solve this problem:

it serves only as a convenient pre-processing step.

2.3 Elemental Quantities

2.3.1 Local Definitions

SLIDE 39



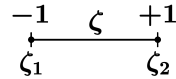
- x_1^k : local node 1 of element T_h^k ;
- x_2^k : local node 2 of element T_h^k ;
- h^k : length of element T_h^k .

Node 1 and node 2 are “local” names for the two (global) nodes (e.g., x_{10} and x_{11}) at the two endpoints. We shall later introduce a mapping between these two different sets of labels. For the present, we restrict attention to each individual element.

2.3.2 Reference Element

SLIDE 40

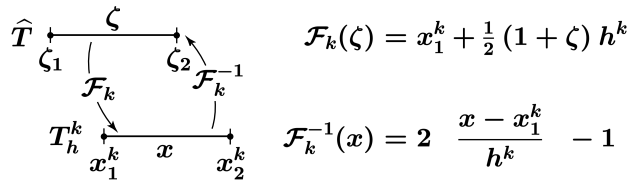
Definition: $\hat{T} = (-1, 1)$



- ζ_1 : reference element node 1;
- ζ_2 : reference element node 2.

SLIDE 41

Relation of \hat{T} to each T_h^k : Affine Mappings



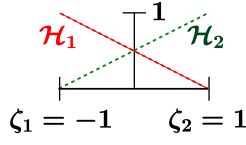
Affine mappings (essentially “linear mappings plus a constant”) have many important properties, one of which is that polynomials of order p map to polynomials of order p . This will be used (implicitly) in what follows.

2.3.3 Reference Element Space, Basis

SLIDE 42

Define space $\widehat{X} = \mathbb{P}_1(\widehat{T})$: all linear polynomials over \widehat{T} ; $\dim(\widehat{X}) = 2$.

Introduce basis for \widehat{X} , $\mathcal{H}_1(\zeta), \mathcal{H}_2(\zeta)$:



$$\left. \begin{array}{l} \mathcal{H}_1(\zeta) = \frac{(1-\zeta)}{2} \\ \mathcal{H}_2(\zeta) = \frac{(1+\zeta)}{2} \end{array} \right\} \text{Lagrangian interpolants}$$

It should be clear that for any $v \in \widehat{X}$, $v(\zeta) = \sum_{i=1}^2 v(\zeta_i) \mathcal{H}_i(\zeta)$.

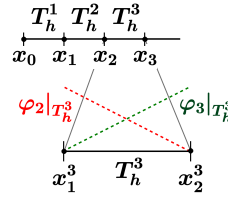
2.3.4 Elemental Matrices

SLIDE 43

$$\widetilde{A}_{h \ i \ j} = a(\varphi_i, \varphi_j) = \int_0^1 \frac{d\varphi_i}{dx} \frac{d\varphi_j}{dx} dx$$

Element T_h^3 (say) contributes

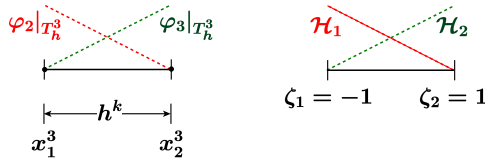
$$\int_{T_h^3} \frac{d\varphi_{2 \text{ or } 3}}{dx} \Big|_{T_h^3} \frac{d\varphi_{2 \text{ or } 3}}{dx} \Big|_{T_h^3} dx$$



Change variables $T_h^3 \rightarrow \widehat{T}$:

N9

SLIDE 44



$$\int_{T_h^3} \frac{d\varphi_{2 \text{ or } 3}}{dx} \frac{d\varphi_{2 \text{ or } 3}}{dx} dx = \int_{-1}^1 \left(\frac{d\mathcal{H}_{1 \text{ or } 2}}{d\zeta} \frac{2}{h^k} \right) \left(\frac{d\mathcal{H}_{1 \text{ or } 2}}{d\zeta} \frac{2}{h^k} \right) \left(d\zeta \frac{h^k}{2} \right)$$

Note 9

More formal mapping

We can do the change of variables in a more explicit way. First, we note that

$$\varphi_{2 \text{ or } 3} \Big|_{T_h^3} (x) = \mathcal{H}_{1 \text{ or } 2}(\mathcal{F}_3^{-1}(x)),$$

which can be easily verified. We also note that

$$\left. \frac{dx}{d\zeta} \right|_{T_h^3} = \frac{d\mathcal{F}_3}{d\zeta} = \frac{h^3}{2}, \quad \left. \frac{d\zeta}{dx} \right|_{T_h^3} = \frac{d\mathcal{F}_3^{-1}}{dx} = \frac{2}{h^3}.$$

Thus

$$\int_{T_h^3} \frac{d\varphi_{2 \text{ or } 3}}{dx} \frac{d\varphi_{2 \text{ or } 3}}{dx} dx = \int_{T_h^3} \frac{d}{dx} \mathcal{H}_{1 \text{ or } 2}(\mathcal{F}_3^{-1}(x)) \frac{d}{dx} \mathcal{H}_{1 \text{ or } 2}(\mathcal{F}_3^{-1}(x)) dx$$

and substituting $\zeta = \mathcal{F}_3^{-1}(x)$ gives

$$\begin{aligned} \int_{T_h^3} \frac{d}{dx} \mathcal{H}_{1 \text{ or } 2}(\mathcal{F}_3^{-1}(x)) \frac{d}{dx} \mathcal{H}_{1 \text{ or } 2}(\mathcal{F}_3^{-1}(x)) dx \\ &= \int_0^1 \frac{d}{d\zeta} \mathcal{H}_{1 \text{ or } 2} \frac{d\zeta}{dx} \frac{d}{d\zeta} \mathcal{H}_{1 \text{ or } 2} \frac{d\zeta}{dx} d\zeta \frac{h^3}{2} \\ &= \frac{h^3}{2} \int_{-1}^1 \frac{d\mathcal{H}_{1 \text{ or } 2}}{d\zeta} \frac{d\mathcal{H}_{1 \text{ or } 2}}{d\zeta} d\zeta. \end{aligned}$$

In higher dimensions with more complicated (“isoparametric”) mappings for curved domains, this more detailed procedure becomes a necessity.

SLIDE 45

Define $\underline{A}^k \in \mathbb{R}^{2 \times 2}$ (e.g., $k = 3$):

E5 E6 N10

$$\frac{2}{h^k} \int_{-1}^1 \frac{d\mathcal{H}_{\alpha(1 \text{ or } 2)}}{d\zeta} \frac{d\mathcal{H}_{\beta(1 \text{ or } 2)}}{d\zeta} d\zeta =$$

$$\frac{2}{h^k} \int_{-1}^1 \frac{d}{d\zeta} \underline{\mathcal{H}} \frac{d}{d\zeta} \underline{\mathcal{H}}^T d\zeta =$$

$$\underline{\mathcal{H}} = \begin{pmatrix} \mathcal{H}_1 \\ \mathcal{H}_2 \end{pmatrix}$$

$$\frac{1}{h^k} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \equiv \underline{A}^k$$

Elemental Stiffness Matrix

\underline{A}^k represents the contribution to $\tilde{\underline{A}}_h$ from element k : of course it still remains to place this element matrix in the right place.

▷ **Exercise 5** Derive the final form given for $A_{\alpha\beta}^k$, $1 \leq \alpha, \beta \leq 2$. ■

▷ **Exercise 6** The *proto-mass matrix* $\tilde{\underline{M}}_h \in \mathbb{R}^{(n+1) \times (n+1)}$ is defined by

$$\tilde{M}_{h \ i \ j} = \int_0^1 \varphi_i \varphi_j dx \quad 0 \leq i, j \leq n+1.$$

(a) Show that $\sum_{i=0}^{n+1} \sum_{j=0}^{n+1} \widetilde{M}_{h i j} = 1$.

(b) By analogy to our procedure for the stiffness matrix above, find the elemental mass matrix $\underline{M}^k \in \mathbb{R}^{2 \times 2}$.

■

Note 10

The dyadic form

The “dyadic” form

$$\begin{aligned} \frac{2}{h^k} \int_{-1}^1 \frac{d}{d\zeta} \underline{H} \frac{d}{d\zeta} \underline{H}^T d\zeta &= \frac{2}{h^k} \int_{-1}^1 \begin{pmatrix} \frac{d\mathcal{H}_1}{d\zeta} \\ \frac{d\mathcal{H}_2}{d\zeta} \end{pmatrix} \begin{pmatrix} \frac{d\mathcal{H}_1}{d\zeta} & \frac{d\mathcal{H}_2}{d\zeta} \end{pmatrix} d\zeta \\ &= \frac{1}{h^k} \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} \end{aligned}$$

makes it clear that the elemental matrices are only rank one, and thus singular (symmetric, but only positive *semi*-definite); the nullspace is $(1 \ 1)^T$. This is not surprising, since in some sense \underline{A}^k corresponds to the pure Neumann problem on a domain which is a single element; we know that the solution to this problem can “float,” as reflected in the $(1 \ 1)^T$ nullspace of \underline{A}^k .

In general, this dyadic form can play an important role in a variety of contexts (e.g., optimization) — it gives the matrix and certain quadratic constraints special properties.

2.3.5 Elemental “Loads”

SLIDE 46

$$\widetilde{F}_{h i} = \ell(\varphi_i) \underset{\text{(say)}}{=} \int_0^1 f \varphi_i dx$$

Element T_h^3 (say) contributes

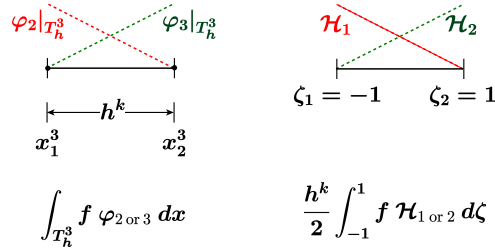


We present here the procedure for $\ell \in L^2(\Omega)$ for which we can write $\ell(v) = \int_0^1 f v dx$ in the usual sense. For delta distributions we need to make (in the end, irrelevant) decisions as to which element (or elements) to which we should

associate the mass. Although not complicated, we restrict our exposition to the more transparent case.

SLIDE 47

Change variables $T_h^3 \rightarrow \hat{T}$:



SLIDE 48

Define $\underline{F}^k \in \mathbb{R}^2$ (e.g., $k = 3$):

$$\begin{aligned}
 F_\alpha^k &= \frac{h^k}{2} \int_{-1}^1 f \mathcal{H}_{\alpha(1 \text{ or } 2)} d\zeta && \text{Elemental Load Vector} \\
 &= \frac{h^k}{2} \int_{-1}^1 f \underline{\mathcal{H}} d\zeta && \underline{\mathcal{H}} = \begin{pmatrix} \mathcal{H}_1 \\ \mathcal{H}_2 \end{pmatrix}.
 \end{aligned}$$

Evaluation (usually) by numerical quadrature.

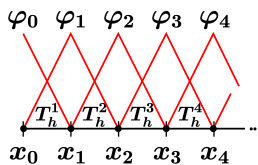
2.4 Assembly

2.4.1 The Idea

The assembly procedure is often denoted the “direct stiffness procedure,” again signaling the roots of finite element practice in the structural community. It is also sometimes referred to as “stamping,” as one stamps in the various elemental contributions. The assembly notion and associated algorithms in fact also apply to discrete (lumped) systems such as electrical circuits, collections of trusses,

SLIDE 49

Recall triangulation and basis functions:



SLIDE 50

$$T_h^3 \text{ contribution to } \tilde{A}_{h \ i \ j} = a(\varphi_i, \varphi_j) = \int_0^1 \frac{d\varphi_i}{dx} \frac{d\varphi_j}{dx} dx$$

$$\int_{T_h^3} \frac{d\varphi_{2 \text{ or } 3}}{dx} \frac{d\varphi_{2 \text{ or } 3}}{dx} dx = \underbrace{\begin{matrix} 2 & 3 \\ \left(\begin{array}{cc} \frac{1}{h^3} & -\frac{1}{h^3} \\ -\frac{1}{h^3} & \frac{1}{h^3} \end{array} \right) \end{matrix}}_{\underline{A}^3}$$

	Column 1 of \underline{A}^3	Column 2 of \underline{A}^3
Row 1 of \underline{A}^3	Adds to \tilde{A}_{22}	Adds to \tilde{A}_{23}
Row 2 of \underline{A}^3	Adds to \tilde{A}_{32}	Adds to \tilde{A}_{33}

SLIDE 51

	C0	C1	C2	C3	C4	
R0	$\left(\begin{array}{cccc} & & & \\ & & & \\ & & \frac{1}{h^3} & -\frac{1}{h^3} \\ & & -\frac{1}{h^3} & \frac{1}{h^3} \\ & & & \\ & & & \end{array} \right)$	$\underline{A}^3 = \left(\begin{array}{cc} \frac{1}{h^3} & -\frac{1}{h^3} \\ -\frac{1}{h^3} & \frac{1}{h^3} \end{array} \right)$				
R1						
R2						
R3						
R4						
⋮						

\tilde{A}_h with T_h^3 accounted for ...

SLIDE 52

$$T_h^4 \text{ contribution to } \tilde{A}_{h \ i \ j} = a(\varphi_i, \varphi_j) = \int_0^1 \frac{d\varphi_i}{dx} \frac{d\varphi_j}{dx} dx$$

$$\int_{T_h^4} \frac{d\varphi_{3 \text{ or } 4}}{dx} \frac{d\varphi_{3 \text{ or } 4}}{dx} dx = \underbrace{\begin{matrix} 3 & 4 \\ \left(\begin{array}{cc} \frac{1}{h^4} & -\frac{1}{h^4} \\ -\frac{1}{h^4} & \frac{1}{h^4} \end{array} \right) \end{matrix}}_{\underline{A}^4}$$

	Column 1 of \underline{A}^4	Column 2 of \underline{A}^4
Row 1 of \underline{A}^4	Adds to \tilde{A}_{33}	Adds to \tilde{A}_{34}
Row 2 of \underline{A}^4	Adds to \tilde{A}_{43}	Adds to \tilde{A}_{44}

SLIDE 53

	C0	C1	C2	C3	C4	...	
R0	$\left(\begin{array}{cccc} & & & \\ & & & \\ & & \frac{1}{h^3} & -\frac{1}{h^3} \\ & & -\frac{1}{h^3} & \frac{1}{h^3} \\ & & \frac{1}{h^3} + \frac{1}{h^4} & -\frac{1}{h^4} \\ & & -\frac{1}{h^4} & \frac{1}{h^4} \end{array} \right)$	$\underline{A}^4 = \left(\begin{array}{cc} \frac{1}{h^4} & -\frac{1}{h^4} \\ -\frac{1}{h^4} & \frac{1}{h^4} \end{array} \right)$					
R1							
R2							
R3							
R4							
⋮							

\tilde{A}_h with T_h^3, T_h^4 accounted for ...

Note we can think of each row i of $\tilde{\underline{A}}_h$ as the contribution to $\delta J_v(u_h)$ due to $v = \varphi_i$: we see that R3 (corresponding to φ_3) of $\tilde{\underline{A}}_h$ takes half the variation due to φ_3 from T_h^3 , and half from T_h^4 — the two elements over which φ_3 is non-zero. In the above example R3 is now complete — and for $h^3 = h^4 = h$ we see that the desired $\frac{1}{h}(-1 \ 2 \ -1)$ has now fully emerged. No other rows are yet complete.

SLIDE 54

By similar arguments:

$$T_h^3 \text{ contribution to } \tilde{F}_h i = \ell(\varphi_i) = \int_0^1 f \varphi_i dx$$

$$\int_{T_h^3} f \varphi_{2 \text{ or } 3} dx = \underbrace{\begin{matrix} 2 & \left(\frac{h^3}{2} \int_{-1}^1 f \mathcal{H}_1 d\zeta \right) \\ 3 & \left(\frac{h^3}{2} \int_{-1}^1 f \mathcal{H}_2 d\zeta \right) \end{matrix}}_{\underline{F}^3}$$

Row 1 of \underline{F}^3	Adds to $\tilde{F}_h 2$
Row 2 of \underline{F}^3	Adds to $\tilde{F}_h 3$

SLIDE 55

$$\begin{matrix} \text{R0} \\ \text{R1} \\ \text{R2} \\ \text{R3} \\ \text{R4} \\ \vdots \end{matrix} \underbrace{\begin{pmatrix} \\ \\ F_1^3 \\ F_2^3 \\ \\ \end{pmatrix}}_{\tilde{\underline{E}}_h \text{ with } T_h^3 \text{ accounted for}} \quad \underline{F}^3 = \begin{pmatrix} F_1^3 \\ F_2^3 \end{pmatrix}$$

SLIDE 56

$$T_h^4 \text{ contribution to } \tilde{F}_h i = \ell(\varphi_i) = \int_0^1 f \varphi_i dx$$

$$\int_{T_h^4} f \varphi_{3 \text{ or } 4} dx = \underbrace{\begin{matrix} 3 & \left(\frac{h^4}{2} \int_{-1}^1 f \mathcal{H}_1 d\zeta \right) \\ 4 & \left(\frac{h^4}{2} \int_{-1}^1 f \mathcal{H}_2 d\zeta \right) \end{matrix}}_{\underline{F}^4}$$

Row 1 of \underline{F}^4	Adds to $\tilde{F}_h 3$
Row 2 of \underline{F}^4	Adds to $\tilde{F}_h 4$

$$\begin{array}{l}
 \text{R0} \\
 \text{R1} \\
 \text{R2} \\
 \text{R3} \\
 \text{R4} \\
 \vdots
 \end{array}
 \left(\begin{array}{c} \\ \\ F_1^3 \\ F_2^3 + F_1^4 \\ F_2^4 \\ \\ \end{array} \right)
 \quad \underline{F}^4 = \begin{pmatrix} F_1^4 \\ F_2^4 \end{pmatrix}$$

$\underline{\tilde{F}}_h$ with T_h^3, T_h^4 accounted for

2.4.2 The Algorithm

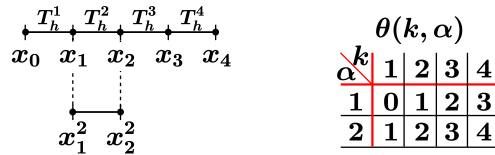
Introduce local-to-global mapping:

$$\theta(k, \alpha): \underbrace{\{1, \dots, K\}}_{\text{element}} \times \underbrace{\{1, 2\}}_{\text{local node number}} \rightarrow \underbrace{\{0, \dots, n+1\}}_{\text{global node number}}$$

such that

$$x_\alpha^k \text{ (local node } \alpha \text{ in element } k) = x_{\theta(k, \alpha)} \text{ (global node } \theta(k, \alpha)).$$

Example: $K = 4$



Note that assigning different numbers to different global nodes — the ordering problem — just corresponds to a different $\theta(k, \alpha)$.

Procedure for $\tilde{\underline{A}}_h$:

- zero $\tilde{\underline{A}}_h$;
- {for $k = 1, \dots, K$
 - {for $\alpha = 1, 2$
 - $i = \theta(k, \alpha)$;
 - {for $\beta = 1, 2$
 - $j = \theta(k, \beta)$;

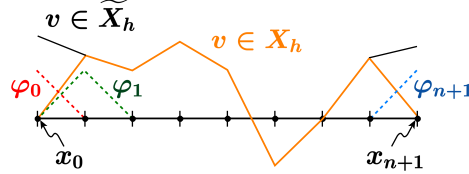
2.5.2 Homogeneous Dirichlet

SLIDE 63

$u_h \in X_h$ such that $a(u_h, v) = \ell(v)$, $\forall v \in X_h$:

$$X_h = \{v \in X \mid v|_{T_h} \in \mathbb{P}_1(T_h), \forall T_h \in \mathcal{T}_h\};$$

$$X = \{v \in H^1(\Omega) \mid v(0) = v(1) = 0\}.$$



SLIDE 64

Explicit Elimination

$X_h \Rightarrow \varphi_0, \varphi_{n+1}$ not admissible variations, so

REMOVE R_0 and R_{n+1} from \tilde{A}_h ;

$\tilde{u}_{h,0} = \tilde{u}_{h,n+1} = 0$, so

REMOVE C_0 and C_{n+1} from \tilde{A}_h .

Recover $\underline{A}_h u_h = \underline{F}_h$

Here R_x and C_y refer to Row x and Column y of \tilde{A}_h . To see why we remove C_0 and C_{n+1} , recall the columnwise interpretation of matrix multiplication: $\tilde{u}_{h,0}$ and $\tilde{u}_{h,n+1}$ “weight” C_0 and C_{n+1} , and since $\tilde{u}_{h,0} = \tilde{u}_{h,n+1} = 0$, these columns may be eliminated.

SLIDE 65

Big-Number Approach

penalty

Place $1/\varepsilon$ ($\varepsilon \ll 1$) on entries $\tilde{A}_{h,0,0}$ and $\tilde{A}_{h,n+1,n+1}$.

Place 0 on entries $\tilde{F}_{h,0}$ and $\tilde{F}_{h,n+1}$.

This replaces R_0 and R_{n+1} with

$$\tilde{u}_{h,0} \cong 0, \tilde{u}_{h,n+1} \cong 0$$

in an “easy,” symmetric way.

This is the easiest to implement, in particular in higher dimensions, in that it requires no modification to the data structure for \underline{A}_h . In practice ε should be very small compared to the entries of R_0 and R_{n+1} , but still not in the “noise” (relative to precision). It might appear that small ε would cause conditioning problems, but this is typically not the case — the “bad modes” are not aggravated.

$$\tilde{u}_{h0} \cong u^D, \tilde{u}_{h_{n+1}} \cong 0.$$

▷ **Exercise 7** Consider the mixed Neumann-Dirichlet problem, $-u_{xx} = f$ in $(0, 1)$, $u(0) = u^D$, $u_x(1) = 0$. How should \tilde{A}_h, \tilde{F}_h be modified to incorporate these conditions using (i) Explicit Elimination, and (ii) the Big-Number Approach? ■

2.6 Quadrature

2.6.1 Question

SLIDE 70

How do we evaluate

$$F_\alpha^k = \frac{h^k}{2} \int_{-1}^1 \underbrace{f\left(x_1^k + \frac{(1+\zeta)}{2} h^k\right)}_{f^k(\zeta)} \mathcal{H}_\alpha(\zeta) d\zeta$$

for general f ?

N11

Note $f^k(\zeta) = f\left(x_1^k + \frac{(1+\zeta)}{2} h^k\right) = f(x = \mathcal{F}_k(\zeta))$, which returns the correct value of $f(x)$ corresponding to $\zeta \in (-1, 1)$ in the reference element.

Note 11

Variable conductivity

The issue of quadrature also arises in the case of variable conductivity (and, in \mathbb{R}^2 , in curved domains). For example, if we wish to solve the problem

$$\begin{aligned} -(\kappa(x) u_x)_x &= f && \text{in } (0, 1) \\ u(0) &= u(1) &= 0, \end{aligned}$$

the weak form is readily found to be

$$\int_0^1 \kappa(x) u_x v_x dx = \int_0^1 f v dx, \quad \forall v \in H_0^1(\Omega);$$

the finite element approximation is thus

$$\underbrace{\int_0^1 \kappa(x) u_{h,x} v_x dx}_{a(u_h, v)} = \int_0^1 f v dx, \quad \forall v \in X_h.$$

This leads to the discrete equations $\underline{A}_h \underline{u}_h = \underline{F}_h$, where

$$A_{h,ij} = a(\varphi_i, \varphi_j) = \int_0^1 \kappa(x) \frac{d\varphi_i}{dx} \frac{d\varphi_j}{dx} dx,$$

and associated elemental matrices

$$A_{\alpha\beta}^k = \frac{2}{h^k} \int_{-1}^1 \kappa \left(x_1^k + \frac{(1+\zeta)}{2} h^k \right) \frac{d\mathcal{H}_\alpha}{d\zeta} \frac{d\mathcal{H}_\beta}{d\zeta} d\zeta .$$

Clearly, for general $\kappa(x)$, we can not hope to integrate these expressions exactly, and thus numerical quadrature is required.

SLIDE 71

Approaches

- “Analytical” Integration
- Symbolic Integration
- Gauss Quadrature ←
- Integration by Interpolation

N12

Note 12

Integration by interpolation

In this approach, we first interpolate $f^k(\zeta)$ as

$$f^k(\zeta) = \sum_{\beta=1}^2 f_\beta^k \mathcal{H}_\beta(\zeta) ,$$

where $f_1^k = f^k(\zeta \rightarrow -1)$ and $f_2^k = f^k(\zeta \rightarrow +1)$; note we indicate the limit rather than the value to recognize that f may be discontinuous.

We then have

$$\begin{aligned} F_\alpha^k &\approx \sum_{\beta=1}^2 \frac{h^k}{2} \int_{-1}^1 \mathcal{H}_\alpha \mathcal{H}_\beta d\zeta f_\beta^k \\ &= \sum_{\beta=1}^2 M_{\alpha\beta}^k f_\beta^k , \end{aligned}$$

where \underline{M}^k is the elemental mass matrix (see Exercise 6).

It can be easily shown that if f is, indeed, continuous, then $\underline{\tilde{F}}_h = \underline{\tilde{M}}_h \underline{\tilde{f}}$, where $\underline{\tilde{f}}$ is the vector of nodal values of f , that is

$$\underline{\tilde{f}} = \begin{pmatrix} f(x_0) \\ \vdots \\ f(x_{n+1}) \end{pmatrix} .$$

From our earlier discussion of the mass matrix $\underline{\tilde{M}}$ (see Exercise 6) we thus obtain

$$\underline{\tilde{F}}_h = \begin{pmatrix} \frac{1}{3} f_0 + \frac{1}{6} f_1 \\ \frac{1}{6} f_0 + \frac{2}{3} f_1 + \frac{1}{6} f_2 \\ \vdots \\ \frac{1}{6} f_n + \frac{1}{3} f_{n+1} \end{pmatrix} .$$

We thus see that the finite element method “distributes” the loads to a few neighboring grid points — not surprising, since the finite element identity, the mass matrix, is not diagonal. (On some occasions one prefers a diagonal identity — the result is a “lumped” mass matrix.)

We note that the above integrations, like all inexact numerical quadratures, constitutes a variational crime — a deviation from the strict variational recipe. Indeed, \tilde{F}_h above (and in this entire section) is really not the \tilde{F}_h of earlier, but rather an *approximation* to the earlier \tilde{F}_h . In the case above, the errors we commit are small in the sense that our earlier error estimates — at least as regards the exponent of h — remain unchanged.

2.6.2 Gauss Quadrature

SLIDE 72

Approximate

$$\begin{aligned} F_\alpha^k &= \frac{h^k}{2} \int_{-1}^1 f^k(\zeta) \mathcal{H}_\alpha(\zeta) d\zeta \\ &\approx \frac{h^k}{2} \sum_{q=1}^{N_q} \rho_q f^k(z_q) \mathcal{H}_\alpha(z_q): \end{aligned}$$

ρ_q : Gauss-Legendre quadrature weights
 z_q : Gauss-Legendre quadrature points.

The Gauss-Legendre ρ_q, z_q are tabulated in any finite element textbook on numerical quadrature, or handbook on numerical methods.

SLIDE 73

The $\rho_q, z_q, q = 1, \dots, N_q$ are chosen so as

to integrate exactly all $g \in \mathbb{P}_{2N_q-1}((-1, 1))$. N13

To conserve “ideal” convergence rates,

require $N_q \geq 1$ ($\geq p$ for \mathbb{P}_p elements).

By “ideal” convergence rates we refer to those for the H^1 and L^2 errors in the absence of any variational crimes (such as quadrature).

Note 13

Gauss quadrature procedures

There are many different approaches to numerical quadrature. Two of the most common are Gauss quadrature and Newton-Cotes formulas: in the former,

both the ρ_q and z_q are free to vary, providing $2N_q$ degrees of freedom; in the latter (standard rectangle, trapezoidal, . . . rules) the z_q are prescribed, and only the ρ_q are free to vary — one thus obtains lower accuracy for fixed N_q . (There are also schemes “in between” these two extremes — in which *some* points z_q are set, but the rest can be optimized.) There are different Gauss quadrature formulas depending on the “weight” in the integral; here the weight is unity, and the formulas are thus known as Gauss-Legendre. (The name derives from the Sturm-Liouville problem which generates the Legendre polynomials; relatedly, the z_q are the zeroes of the N_q^{th} Legendre polynomial.)

We can easily derive the $N_q = 1$ Gauss-Legendre point z_1 and weight ρ_1 . In particular, we wish to choose z_1 and ρ_1 , such that, for $g \in \mathbb{P}_1((-1, 1))$,

$$\int_{-1}^1 g \, dz = \rho_1 g(z_1) .$$

But we can write any $g \in \mathbb{P}_1((-1, 1))$ as $g^0 + g^1 z$; furthermore,

$$\int_{-1}^1 g^0 + g^1 z \, dz = 2g^0 .$$

So clearly we should choose z_1 such that $g^1 z_1 = 0$ — that is, $z_1 = 0$ — and then $\rho_1 = 2$.

To evaluate F_α^k (approximately) with this one point formula, we then have

$$\begin{aligned} F_\alpha^k &\approx h^k f^k(0) \mathcal{H}_\alpha(0) \\ &= \frac{h^k}{2} f\left(\frac{1}{2}(x_1^k + x_2^k)\right) . \end{aligned}$$

Clearly, this is very simple to implement; Gauss quadrature is a very popular approach to evaluation of elemental quantities. Note that it is not always the case that we should use the lowest possible N_q ; sometimes it is of interest to integrate exactly a particular term (e.g., the mass matrix, which involves quadratics), in which case N_q must be larger than (say for linear elements) 1.
