Discretization of the Poisson Problem in $\mathbb{R}^1$: Formulation

April 2, 2003

1 Model Problems

1.1 Dirichlet

1.1.1 Strong Form

SLIDE 1

Domain: $\Omega = (0,1)$.
Find u such that

$$\begin{aligned} -u_{xx} &= f \qquad\qquad \text{in } \Omega \\ u(0) &= u(1) = 0 \end{aligned} \quad ,$$

for given f.

1.1.2 Minimization Statement

SLIDE 2

Define $X \equiv H_0^1(\Omega)$.
Find

$$\boxed{u = \arg\min_{w \in X} J(w)}$$

where

$$J(w) = \frac{1}{2} \int_0^1 w_x^2 \, dx - \int_0^1 f w \, dx \ .$$

This follows from the previous lecture, noting that dA is now dx, and ∇w is now w_x.

1.1.3 Weak Formulation

SLIDE 3

Find $u \in X$ such that

$$\delta J_v(u) = 0 \ , \qquad \forall\, v \in X$$

$$\Updownarrow$$

$$\boxed{\int_0^1 u_x \, v_x \, dx = \int_0^1 f \, v \, dx \ , \qquad \forall\, v \in X \ .}$$

Again, this follows from our earlier lecture with $\nabla u \cdot \nabla v$ now given by $u_x v_x$.

1.1.4 Notation

SLIDE 4

Define
$$a(w,v) = \int_0^1 w_x\, v_x\, dx$$
$$\ell(v) = \int_0^1 f\, v\, dx\ .$$

Minimization: $\boxed{u = \arg \min_{w \in X} \frac{1}{2}\, a(w,w) - \ell(w)}$

Weak: $\boxed{u \in X\!:\ a(u,v) = \ell(v)\ ,\ \forall\, v \in X}$

1.1.5 Generalization

SLIDE 5

For any $\ell(v) \in H^{-1}(\Omega)$,
find $u \in H_0^1(\Omega)$ such that

$$u = \arg \min_{w \in H_0^1(\Omega)} \tfrac{1}{2}\, a(w,w) - \ell(w)\ ; \quad \text{or}$$

$$a(u,v) = \ell(v), \qquad \forall\, v \in H_0^1(\Omega)\ ;$$

for example, $\ell(v) = \langle \delta_{x_0}, v \rangle = v(x_0)$ is admissible.

As indicated earlier, the delta distribution is not admissible if $\Omega \subset \mathbb{R}^2$*, as can be motivated by considering the Green's function.*

1.1.6 Regularity

SLIDE 6

If $\ell \in H^{-1}(\Omega)$,
$$\|u\|_{H^1(\Omega)} \le C\ \|\ell\|_{H^{-1}(\Omega)}\ .$$

If $\ell \in L^2(\Omega)$, $\ell(v) = \displaystyle\int_0^1 f\, v\, dx$
$$\|u\|_{H^2(\Omega)} \le C_0\ \|f\|_{L^2(\Omega)}\ . \qquad \boxed{\text{N1}}$$

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

Note 1 — ***Regularity***

If $\ell(v) = \int_0^1\ f\, v\, dx$, with $f \in L^2(\Omega)$, we immediately obtain from $\|u\|_{H^1(\Omega)} < C\|\ell\|_{H^{-1}(\Omega)}$ that $\|u\|_{H^1(\Omega)} \le C\|f\|_{L^2(\Omega)}$, since the H^{-1} norm is always bounded by the L^2 norm (there is "more" in the denominator). But from the strong form $-u_{xx} = f$ we can see that $|u|_{H^2(\Omega)} \le \|f\|_{L^2(\Omega)}$. It thus follows that C_0 in the above slide is $(1 + C^2)^{1/2}$. This can also be shown by explicit construction of u (see Lecture 2 from earlier in the course).

The fact that u is regular when f is regular (and in $\mathbb{R}^2$, the domain is suitably regular) has very important implications as regards the convergence rate of the finite element method and the construction of *a priori* and *a posteriori* error estimates.

1.2 "Neumann"

1.2.1 Strong Form

SLIDE 7

Domain: $\Omega = (0, 1)$.
Find u such that

$$\begin{aligned} -u_{xx} &= f \qquad \text{in } \Omega \,, \\ u(0) &= 0 \,, \\ u_x(1) &= g \,, \end{aligned}$$

for given f, g .

1.2.2 Minimization Statement

SLIDE 8

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

Find

$$\boxed{u = \arg \min_{w \in X} J(w)}$$

where

$$J(w) = \frac{1}{2} \int_0^1 w_x^2 \, dx - \int_0^1 f \, w \, dx - g \, w(1) \,.$$

This follows from the previous lecture, noting that $\int_{\Gamma^N} g \, v \, dA$ is here just $g \, v(1)$. We can also show this explicitly by integrating by parts to find $\delta J_v(u) = \int_0^1 v \, \{-u_{xx} - f\} \, dx + v(1)\{u_x(1) - g\} = 0$, $\forall \, v \in X$.

1.2.3 Weak Statement

SLIDE 9

Find $u \in X$ such that

$$\delta J_v(u) = 0 \,, \qquad \forall \, v \in X$$

$$\Updownarrow$$

$$\boxed{\int_0^1 u_x \, v_x \, dx = \int_0^1 f \, v \, dx + g \, v(1) \,, \qquad \forall \, v \in X \,.}$$

1.2.4 Notation

SLIDE 10

Define
$$\begin{aligned} a(w,v) &= \int_0^1 w_x\, v_x\, dx \\ \ell(v) &= \int_0^1 f\, v\, dx + g\, v(1)\ . \end{aligned}$$

N2

Minimization:
$$\boxed{u = \arg \min_{w \in X} \frac{1}{2}\, a(w,w) - \ell(w)}$$

Weak:
$$\boxed{u \in X\colon\ a(u,v) = \ell(v)\ ,\ \forall\, v \in X}$$

Note 2 ***Neumann and delta distributions (Optional)***

We note that, in $\mathbb{R}^1$, our Neumann condition looks exactly like a delta distribution forcing at the boundary, $x = 1$. This is fine, since we know the delta distribution is an admissible (bounded) linear functional, that is, is in the space $H^{-1}(\Omega)$, for this one-dimensional problem.

We know that in the interior a delta distribution imposes (weakly) a jump in the derivative (see Note 10 of last lecture). On the boundary, it imposes (weakly) the value of the derivative itself — since there is no "other side" to the jump.

---

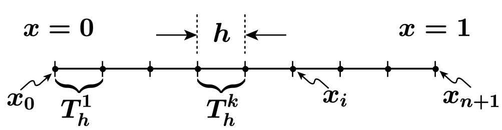
2 Rayleigh-Ritz Approach

2.1 Approximation

2.1.1 Mesh

SLIDE 11

Note our default problem is the Dirichlet problem; we shall explicitly indicate Neumann when we wish to consider that problem (primarily in the exercises).



$$\overline{\Omega} = \bigcup_{k=1}^{K} \overline{T}_h^k \qquad \begin{aligned} T_h^k\ ,\quad & k = 1, \ldots, K = n+1\colon\ \textit{elements} \\ x_i\ ,\quad & i = 0, \ldots, n+1\colon\ \textit{nodes} \end{aligned}$$

N3

Note 3 ***Triangulations $\mathcal{T}_h$***

The above decomposition is known as a *triangulation*, $\mathcal{T}_h$, even though in $\mathbb{R}^1$ our elements are not really triangles (though they are simplices — which are segments in $\mathbb{R}^1$, triangles in $\mathbb{R}^2$, and tetrahedra in $\mathbb{R}^3$). In general, a triangulation $\mathcal{T}_h$ refers to the collection of *elements* (segments, triangles, quadrilaterals, ...) T_h^k, the union of which reconstitutes the original domain Ω. Note the elements are open, so in fact $\overline{\Omega}$ (the closure of Ω) is the sum of the *closure* of the T_h^k. As in finite differences, we also have *nodes* — which will play a central role — but it is the elements that define the approximation.

In general, we may consider non-uniform meshes in which the elements are of different lengths, or "diameters," h^k. In this case the h which appears in $\mathcal{T}_h$ is the maximum diameter over all elements. It is important to remember that we will in fact be concerned with a sequence of triangulations $\mathcal{T}_h$ with $h \to 0$. We say that our sequence of triangulations is *quasi-uniform* if the ratio $h_{\min}/h_{\max}$ over $\mathcal{T}_h$ is bounded from below as $h \to 0$; we shall always assume this to be the case. In higher dimensions we will also define a *regularity* hypothesis related to the shape of the elements.

There is another way to describe elements and triangulations in which T_h shall refer to any particular member of $\mathcal{T}_h$ — that is, the enumeration and k superscript above is left implicit. This is often more convenient for describing various approximations. In terms of this abbreviated notation, we have that

$$\overline{\Omega} = \bigcup_{T_h \in \mathcal{T}_h} \overline{T}_h$$

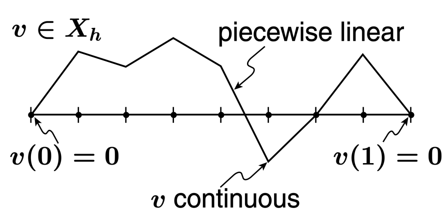
where $T_h \in \mathcal{T}_h$ indicates to take the union over all elements.

2.1.2 Space $X_h \subset X$

SLIDE 12

$$\boxed{X_h = \left\{ v \in X \;\middle|\; v|_{T_h^k} \in \mathbb{P}_1(T_h^k), \quad k = 1, \ldots, K \right\}}$$

Recall that $v|_{T_h^k}$ *means* v restricted to T_h^k. *Thus the above says that a* v *in* X_h *must be in* $X = H_0^1(\Omega)$, *and must be piecewise-linear —* $\mathbb{P}_1(\mathcal{D})$, $\mathcal{D} \subset \Omega$, *is the space of linear polynomials over* $\mathcal{D}$ *— on each element. We can also write* $X_h = \{v \in X \mid v|_{T_h} \in \mathbb{P}_1(T_h),\ \forall\, T_h \in \mathcal{T}_h\}$.



N4

Note 4	***Continuity of v in X_h***

It is clear that if $v \in X_h$, then since $X_h \subset X$ (X_h is a *subspace* of X — any member of X_h is a member of X because $X_h = \{v \in X \,|\, \cdots\}$) $v(0) = v(1) = 0$ — *all* members of v in $H_0^1(\Omega)$ (and hence $X_h \subset X$) vanish at $x = 0$ and $x = 1$. But $X_h \subset X$ also tells us that v must be *continuous*: the (distributional) derivative of the function depicted above is piecewise constant on each element, and hence square integrable, as required by $H^1(\Omega)$; however, if we had jumps in v between elements, the derivative would generate delta distributions at the nodes, which are *not* in $L^2(\Omega)$ (see Note 7 of the last lecture) — a function with jumps is thus *not* in $H^1(\Omega)$. It is important to note that we do not require that our v be in $C^1(\Omega)$, that is, have continuous first derivatives — this is much more difficult to implement numerically.

We remark that there are finite element approximations in which $X_h \not\subset X$ — these are known as *non*conforming approximations, as opposed to the conforming approximations we consider here.

2.1.3 Basis

SLIDE 13

General definition: given a linear space Y,
a set of members $y_j \in Y$, $j = 1, \ldots, M$,
is *a* basis for Y if and only if

$$\forall\, y \in Y,\ \exists \text{ unique } \ \alpha_j \in \mathbb{R} \ \text{ such that}$$

$$\boxed{y = \sum_{j=1}^{M} \alpha_j\, y_j\ ;}$$

dim(ension) $(Y) = M$.

N5 N6 E1 E2

Note 5	***Linear dependence and dimension***

It follows from our definition of a basis that any set of M linearly independent members y_j — members such that

$$\sum_{j=1}^{M} \alpha_j\, y_j = 0 \Rightarrow \alpha_j = 0,\ j = 1, \ldots, M$$

— will serve as a basis. It is also readily demonstrated that, although our choice of basis is not at all unique, the dimension of Y, $\dim(Y)$, *is* unique. For simplicity we will use the basis concept primarily in the context of finite-dimensional spaces such as X_h; but infinite dimensional spaces such as $X = H_0^1(\Omega)$ can also be described in these terms. Note we can express a space Y in terms of any basis as $Y = \text{span}\,\{y_j,\ j = 1,\ldots,M\}$, meaning that any member of Y can be represented as a linear combination of the y_j.

Note 6 ***Orthogonality***

If our space Y is a Hilbert space with inner product $(\cdot,\cdot)_Y$, we can introduce the notion of orthogonality: two members $y_1 \in Y$ and $y_2 \in Y$ are *orthogonal* if

$$(y_1, y_2)_Y = 0 \ .$$

An orthogonal basis is thus a basis for which the y_j are mutually orthogonal, $(y_i, y_j)_Y = 0,\ i \neq j$. If, furthermore, $(y_i, y_i)_Y = \|y_i\|_Y^2 = 1$, the basis is *orthonormal*.

$\triangleright$ ***Exercise 1*** Consider the Hilbert space $\mathbb{R}^2$ "equipped" with usual Euclidean inner product, $([x_1, y_1], [x_2, y_2]) = x_1 x_2 + y_1 y_2$, and hence norm $\|[x, y]\| = (x^2 + y^2)^{1/2}$. Note the pair $[x, y]$ refers to a single member (point) in $\mathbb{R}^2$.

(a) Is $(1, 1), (1, 0)$ a basis for $\mathbb{R}^2$? an orthogonal basis?

(b) If $(1, -1)/\sqrt{2}$ is one of our basis vectors, find a second vector such that we have an orthonormal basis.

■

$\triangleright$ ***Exercise 2*** Consider $Y = \mathbb{P}_2([-1, 1]) = \text{span}\,\{1, x, x^2\}$ equipped with the L^2 inner product, $(y, z)_Y = \int_{-1}^{1} y\, z\, dx$ (here y and z are two members of Y, that is, two polynomials).

(a) Replace x^2 with another basis vector (in fact, polynomial) such that we now have an orthogonal basis.

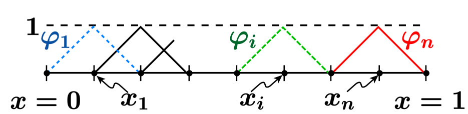
(b) Appropriately normalized, what polynomial system (that is, associated with what famous French mathematician) are you generating by the above "Gram-Schmidt" process.

■

SLIDE 14

Nodal basis for X_h:

$$\varphi_j,\ j = 1,\ldots,n = \dim(X_h)$$



$\boxed{\varphi_i \;\text{ nonzero only on }\; \overline{T}_h^i \bigcup \overline{T}_h^{i+1}}$

$\boxed{\text{N7}}$ $\boxed{\text{N8}}$

Note 7 — ***Counting argument***

We can convince ourselves independent of any basis that $\dim(Y) = n$. First, we note that on any element T_h^k, $v|_{T_h^k} = a^k + b^k x$; since we have $K = n+1$ elements, this gives us $2n+2$ degrees-of-freedom. However we also have 2 boundary conditions (at $x = 0$, $x = 1$) and n interface continuity conditions ($v|_{T_h^i}(x_i) = v|_{T_h^{i+1}}(x_i)$, $i = 1, \ldots, n$), for a total of $n+2$ constraints. Thus, $\dim(Y) = 2n + 2 - (n+2) = n$.

Note 8 — ***Interpretation of basis***

The *nodal* in nodal basis refers to the fact that the basis coefficients are not just "Fourier-like" coefficients, but also have a "physical-space" significance: if v is a member of X_h, we know from the definition of a basis that

$$v = \sum_{i=1}^{n} v_i \, \varphi_i(x) \; ;$$

however, $v(x_j) = \sum_{i=1}^{n} v_i \, \varphi_i(x_j) = v_j$, $j = 1, \ldots, n$, since the φ_i are zero at all nodes except x_i. (Indeed, the φ_i can be uniquely *defined* by the conditions $\varphi_i \in X_h$, $\varphi_i(x_j) = \delta_{ij}$, $i = 1, \ldots, n$; here δ_{ij} is the Kronecker-delta symbol.) Note that there is no "φ_0" or "φ_{n+1}" in the basis since we must have $v(0) = v(1) = 0$ for $v \in X_h$.

Thus $v_i = v(x_i)$, *the value of* v *at* $x = x_i$, the i^{th} node; and $\sum_{i=1}^{n} v_i \, \varphi_i(x)$ "connects" the values of v at the nodes with linear segments on each element. It is then patently clear that we can represent any piecewise-linear continuous function v that vanishes at $x = 0$ and $x = 1$ by the choice $v_i = v(x_i)$, $i = 1, \ldots, n$. Furthermore, the v_i are unique — no choice except $v_i = v(x_i)$ will work. It thus follows that the φ_j are indeed a basis.

There are many other possible choices for basis — we explore a particularly useful one in a later exercise. However, the nodal representation remains the most common, first because of the convenient interpretation as nodal values, and second because of the matrix sparsity induced by the minimal overlap between the φ_j.

2.2 "Projection"

2.2.1 Plan

SLIDE 15

Let

$$\underbrace{u_h \quad (\in X_h)}_{\text{RR/FE Approximation}} = \sum_{j=1}^{n} u_{hj}\, \varphi_j(x) \ ;$$

set $u_{hj} = w_j$ that minimize

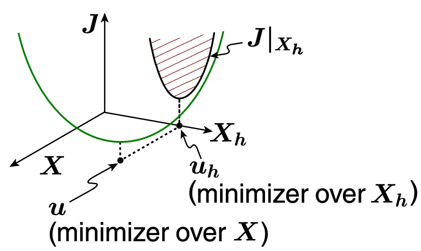
$$J\left(\sum_{j=1}^{n} w_j\, \varphi_j\right) \ .$$

More precisely, what we mean is that u_{hj} *is the minimizer of* $J(\sum_{j=1}^{n} w_j\, \varphi_j)$*, that is,* $\arg\min J(\sum_{j=1}^{n} w_j\, \varphi_j)$*.*

The finite element (FE) approximation is, for this simple problem, a classical Rayleigh-Ritz (RR) approach with a particular choice of space and basis.

SLIDE 16

Geometric Picture:



Since any member of X_h *can be represented as our sum over the* φ_j*, we are finding the minimizer* (u_h) *of* J *over all functions in* X_h*. The choice of basis will thus not affect the minimizer (or minimum), though it will affect the particular coefficients. We later prove that* $J|_{X_h}$ *is a paraboloid — as indicated here — and that by extension* J *over* X *is an infinite-dimensional paraboloid. We see from this picture that as* X_h *grows it absorbs more of* X*, and* u_h *should thus go to* u *as we increase the number of elements; this is indeed the case. Of course,* $J(u_h) \geq J(u)$*, since* $J(u_h)$ *is the minimum of* J *over a subspace* (X_h) *of* X*.*

2.2.2 $J|_{X_h}$

$$\begin{aligned} J\left(\sum_{j=1}^{n} w_j\,\varphi_j\right) &= \frac{1}{2}\,a\left(\sum_{i=1}^{n} w_i\,\varphi_i, \sum_{j=1}^{n} w_j\,\varphi_j\right) - \ell\left(\sum_{i=1}^{n} w_i\,\varphi_i\right) \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i\,a(\varphi_i,\varphi_j)\,w_j - \sum_{i=1}^{n} w_i\,\ell(\varphi_i) \end{aligned}$$

by *bilinearity* and *linearity*. $\boxed{\text{N9}}$

Note 9 ***Application of bilinearity and linearity***

We indicate here the steps evoked above:

$$\begin{aligned} & a\left(\sum_{i=1}^{n} w_i\,\varphi_i, \sum_{j=1}^{n} w_j\,\varphi_j\right) \\ &= a\left(w_1\varphi_1 + \sum_{i=2}^{n} w_i\varphi_i, \sum_{j=1}^{n} w_j\varphi_j\right) \\ &= w_1\,a\left(\varphi_1, \sum_{j=1}^{n} w_j\varphi_j\right) + a\left(\sum_{i=2}^{n} w_i\varphi_i, \sum_{j=1}^{n} w_j\varphi_j\right) \\ &= \sum_{i=1}^{n} w_i\,a\left(\varphi_i, \sum_{j=1}^{n} w_j\varphi_j\right) \\ &= \sum_{i=1}^{n} w_i\left(a\left(\varphi_i, w_1\varphi_1 + \sum_{j=2}^{n} w_j\varphi_j\right)\right) \\ &= \sum_{i=1}^{n} w_i\left(a(\varphi_i,\varphi_1)\,w_1 + a\left(\varphi_i, \sum_{j=2}^{n} w_j\varphi_j\right)\right) \\ &= \sum_{i=1}^{n} w_i \sum_{j=1}^{n} a(\varphi_i,\varphi_j)\,w_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\,a(\varphi_i,\varphi_j)\,w_j\ . \end{aligned}$$

Similar arguments yield the linear term:

$$
\begin{aligned}
\ell\left(\sum_{i=1}^{n} w_i\varphi_i\right) &= \ell(w_1\varphi_1) + \ell\left(\sum_{i=2}^{n} w_i\varphi_i\right)\\
&= w_1\,\ell(\varphi_1) + \ell\left(\sum_{i=2}^{n} w_i\varphi_i\right)\\
&= \sum_{i=1}^{n} w_i\,\ell(\varphi_i).
\end{aligned}
$$

$$
\begin{aligned}
J^R(\underline{w} \in \mathbb{R}^n) &\equiv J\left(\sum_{j=1}^{n} w_j\,\varphi_j\right)\\
&= \frac{1}{2}\,\underline{w}^T\,\underline{A}_h\,\underline{w} - \underline{w}^T\,\underline{F}_h\ .
\end{aligned}
$$

Note that J^R is essentially the same object as $J|_{X_h}$, however $J^R\colon\ \mathbb{R}^n \to \mathbb{R}$ is expressed in terms of the basis coefficients, while $J|_{X_h}\colon\ X_h \to \mathbb{R}$ is expressed in terms of functions in X_h. The minima will be the same: $J^R(u_{hj}) = J|_{X_h}(u_h) = J(u_h)$. Note $\underline{w}$ refers to the vector $(w_1, w_2, \ldots, w_n)^T$; similarly $\underline{u}_h$ shall refer to the vector of basis coefficients (and nodal values) of u_h, $(u_{h1}, u_{h2}, \ldots, u_{hn})^T$.

$$
\begin{aligned}
\underline{F}_h \in \mathbb{R}^n\colon &\quad F_{h\,i} \equiv \ell(\varphi_i)\left(= \int_\Omega f\,\varphi_i\,dx\right)\\
\underline{A}_h \in \mathbb{R}^{n\times n}\colon &\quad A_{h\,ij} \equiv a(\varphi_i,\varphi_j) = \int_\Omega \frac{d\varphi_i}{dx}\,\frac{d\varphi_j}{dx}\,dx \qquad \boxed{\text{E3}}
\end{aligned}
$$

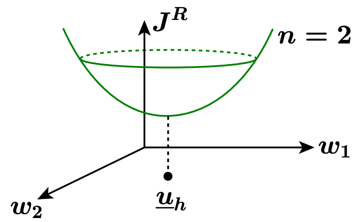
Note this is just a re-expression of our earlier indicial expression in terms of vectors and matrices.

▷ ***Exercise 3*** Show that $\underline{A}_h$ is SPD — symmetric positive-definite. *Hint*: consider what $\underline{v}^T\underline{A}\,\underline{v}$ means in terms of the functions $\sum_{i=1}^{n} v_i\,\varphi_i(x)$; recall that the φ_i are a basis. ■

2.2.3 Minimization

SLIDE 19

$$\underline{u}_h = \arg \min_{\underline{w} \in \mathbb{R}^n} \; J^R(\underline{w})$$



Expand $J^R(\underline{w} = \underline{u}_h + \underline{v})$; require $J^R(\underline{w}) > J^R(\underline{u}_h)$ unless $\underline{v} = 0$.

SLIDE 20

$$\begin{aligned} & J^R(\underline{u}_h + \underline{v}) \\ & \quad = \frac{1}{2}\,(\underline{u}_h + \underline{v})^T\,\underline{A}_h(\underline{u}_h + \underline{v}) - (\underline{u}_h + \underline{v})^T\,\underline{F}_h \\ & \quad = \frac{1}{2}\,\underline{u}_h^T\,\underline{A}_h\,\underline{u}_h - \underline{u}_h^T\,\underline{F}_h \\ & \qquad + \frac{1}{2}\,\underline{v}^T\,\underline{A}_h\,\underline{u}_h + \frac{1}{2}\,\underline{u}_h^T\,\underline{A}_h\,\underline{v} - \underline{v}^T\,\underline{F}_h \\ & \qquad + \frac{1}{2}\,\underline{v}^T\,\underline{A}_h\,\underline{v} \end{aligned}$$

SLIDE 21

$$J^R(\underline{u}_h + \underline{v}) = J^R(\underline{u})$$

$$+\underbrace{(\underline{A}_h\,\underline{u}_h - \underline{F}_h)^T}_{\nabla J^R(\underline{u}_h)}\,\underline{v} \quad \delta J^R_{\underline{v}}(\underline{u}_h) \qquad \mathit{S}\mathrm{PD}$$

$$+\frac{1}{2}\,\underbrace{\underline{v}^T\,\underline{A}_h\,\underline{v}}_{>0,\ \forall\,\underline{v} \neq 0} \qquad \mathrm{S}\mathit{PD}$$

This is essentially a Taylor series about u_h; since J^R is quadratic, this Taylor series terminates with the quadratic term.

SLIDE 22

If (and only if)

$$\delta J^R_{\underline{v}}(\underline{u}_h) = 0, \quad \forall\,\underline{v} \in \mathbb{R}^n$$
$$\Updownarrow$$
$$\nabla J^R(\underline{u}_h) = \underline{A}_h\,\underline{u}_h - \underline{F}_h = \underline{0}$$

then

$$J^R(\underline{w} = \underline{u}_h + \underline{v}) > J^R(\underline{u}_h)\ , \qquad \forall\,\underline{v} \neq 0\ . \qquad \boxed{\text{N10}}$$

If $\underline{A}_h\,\underline{u}_h - \underline{F}_h \neq 0$, we can pick $\underline{v} = -\varepsilon\,(\underline{A}_h\,\underline{u}_h - \underline{F}_h)$; for small enough ε the quadratic terms are negligible, and J^R will thus decrease. This proves the "only *if" — J^R can not be a minimum if $\underline{A}_h\, \underline{u}_h - \underline{F}_h$ is not zero. The "if" is just as easy: if $\underline{A}_h\, \underline{u}_h - \underline{F}_h = 0$, $J^R(\underline{u}_h + \underline{v}) = J^R(\underline{u}_h) + \frac{1}{2}\, \underline{v}^T \underline{A}_h\, \underline{v} > J^R(\underline{u}_h)$ unless $\underline{v} = 0$.*

Note 10 **J^R: a paraboloid**

Since $\underline{A}_h$ is SPD, we know it can be diagonalized as $\underline{Q}^T \underline{A}_h \underline{Q} = \underline{\Lambda}$, where $\underline{Q}$ is the matrix of orthonormal eigenvectors of $\underline{A}$ and $\underline{\Lambda}$ is the diagonal matrix of perforce real, positive eigenvalues, λ_i, $i = 1, \ldots, n$.

We know that $J^R(\underline{u}_h + \underline{v}) = J^R(\underline{u}_h) + \frac{1}{2}\, \underline{v}^T \underline{A}\, \underline{v}$. Expressing (any) $\underline{v}$ as $\underline{Q}\underline{z}$ — a rotation of our axes — we find that

$$\begin{aligned} J^R(\underline{u}_h + \underline{v}) &= J^R(\underline{u}_h) + \frac{1}{2}\, \underline{z}^T \underline{\Lambda}\, \underline{z} \\ &= J^R(\underline{u}_h) + \frac{1}{2} \sum_{i=1}^{n} \frac{z_i^2}{(1/\lambda_i)} \;. \end{aligned}$$

Thus $J^R(\underline{w})$ is a paraboloid with minimum $J^R(\underline{u}_h)$ and (hyper)ellipsiodal cross-sections — $J^R(\underline{w})$ is constant if

$$\frac{1}{2} \sum_{i=1}^{n} \frac{z_i^2}{(1/\lambda_i)} = \text{constant} \;,$$

which is the equation for an ellipsoid in $\mathbb{R}^n$ (with largest or major axis $1/\sqrt{\lambda_{\min}}$ and smallest or minor axis $1/\sqrt{\lambda_{\max}}$). It "follows" that J: $X \to \mathbb{R}$ is an infinite dimensional paraboloid.

2.2.4 Final Result

SLIDE 23

Find $\underline{u}_h \in \mathbb{R}^n$ such that

$$\underbrace{\underline{A}_h}_{a(\varphi_i,\varphi_j)} \;\; \underline{u}_h = \underbrace{\underline{F}_h}_{\ell(\varphi_i)} \quad \Rightarrow \quad u_h(x) = \sum_{j=1}^{N} u_{hj}\, \varphi_j(x) \;.$$

Different bases will give us different $\underline{A}_h$ — with different sparsity, bandedness, and conditioning — and hence different $\underline{u}_h$. For example, if the φ_i are orthonormal in the $a(\cdot,\cdot)$ inner product, $\underline{A}_h$ would be diagonal (not likely in general ...). But $u_h(x)$ depends (at least in infinite precision arithmetic) only on our choice of space — it is basis-independent.

SPD $\Rightarrow$ existence and uniqueness.

3 Galerkin Approach

The Galerkin approach is based on the weak statment, which will exist even when the minimization statement does not. It is thus much more widely applicable, and the cornerstone for general finite element analysis. Indeed, the procedure we describe below will work for any equation and problem and discretization — all that will change in each case is the particular X, X_h, $a(\cdot,\cdot)$, and $\ell(\cdot)$. Furthermore, the general Galerkin procedure is widely used even outside the finite element context.

3.1 Approximation

SLIDE 24

Triangulation $\mathcal{T}_h$;

Space X_h ; and

(Nodal) Basis $X_h = \text{span}\,\{\varphi_1, \ldots, \varphi_n\}$;

as for the Rayleigh-Ritz approach.

3.2 Projection

3.2.1 Plan

SLIDE 25

Let

$$u_h (\in X_h) = \sum_{j=1}^{n} u_{hj}\, \varphi_j(x) \ ;$$

set u_{hj} such that

$$\begin{gathered} \delta J_v(u_h) = 0 \ , \qquad \forall\, v \in X_h \\ \Updownarrow \\ a(u_h, v) = \ell(v) \ , \qquad \forall\, v \in X_h \ . \end{gathered}$$

$\boxed{\text{N11}}$

Note 11 **Interpretation of** $a(u_h, v) = \ell(v)$, $\forall\, v \in X_h$

From the definition of J, we know that

$$\begin{aligned} J(u_h + v) &= \tfrac{1}{2}\, a(u_h + v, u_h + v) - \ell(u_h + v) \\ &= J(u_h) + a(u_h, v) - \ell(v) + \tfrac{1}{2}\, a(v, v) \\ &= J(u_h) + \delta J_v(u_h) + \tfrac{1}{2}\, a(v, v) \ , \end{aligned}$$

where we recognize from the last lecture that $\delta J_v(w) = a(w,v) - \ell(v)$, and hence $\delta J_v(u_h) = a(u_h, v) - \ell(v)$. If we wish $J(u_h)$ to be the minimum of $J(w)$ for all $w \in X_h$, we must have $\delta J_v(u_h) = 0$, $\forall\, v \in X_h$, or $a(u_h, v) = \ell(v)$, $\forall\, v \in X_h$.

We are again minimizing our paraboloid just as in the last section, but this time we start directly from the minimization (or optimality) *condition* rather than from the minimization statement itself. This is shown graphically in the next slide. Note that $\delta J_v(u_h) \neq 0$, $\forall\, v \in X$ ($\delta J_v(u) = 0$, $\forall\, v \in X$): there are directions in X for which $\delta J_v(u_h) \neq 0$, which will lead to lower values of J — unless we are very lucky and the exact solution u is in X_h.

The above arguments again rely on the minimization statement. But we can proceed independently: if $u \in X$ satisfies

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

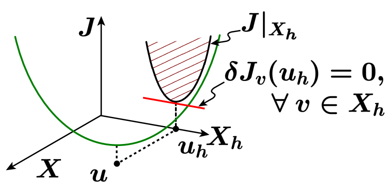
then we simply require $u_h \in X_h$ to satisfy

$$a(u_h,v) = \ell(v), \qquad \forall\, v \in X_h \ .$$

As we shall see in a later Note, this can be interpreted as requiring the equation to be satisfied not at each point, but in an integral sense relative to certain test functions v.

SLIDE 26

Geometric Picture:



3.2.2 Variation

SLIDE 27

Since *any* $v \in X_h$ can be written as

$$v = \sum_{i=1}^{n} v_i\, \varphi_i(x) \ ,$$

$$a(u_h, v) = \ell(v) \ , \qquad \forall\, v \in X_h$$

$$\Updownarrow$$

$$a\left(u_h, \sum_{i=1}^{n} v_i\, \varphi_i\right) = \ell\left(\sum_{i=1}^{n} v_i\, \varphi_i\right) \ , \qquad \forall\, \underline{v} \in \mathbb{R}^n \ .$$

SLIDE 28

But $u_h = \sum_{i=1}^{n} u_{hj}\, \varphi_j$, so

$$a\left(\sum_{j=1}^{n} u_{hj}\,\varphi_j,\ \sum_{i=1}^{n} v_i\,\varphi_i\right) = \ell\left(\sum_{i=1}^{n} v_i\,\varphi_i\right), \qquad \forall\,\underline{v} \in \mathbb{R}^n$$

or, by bilinearity and linearity

$$\underline{v}^T\,\underline{A}_h\,\underline{u}_h = \underline{v}^T\,\underline{F}_h\ , \quad \forall\,\underline{v} \in \mathbb{R}^n\ .$$

We play identical tricks as before, pulling out the v_i and u_{hj} to arrive at

$$\sum_{i=1}^{n}\sum_{j=1}^{n} v_i\,a(\varphi_i,\varphi_j)\,u_{hj} = \sum_{i=1}^{n} v_i\,\ell(\varphi_i), \qquad \forall\,\underline{v} \in \mathbb{R}^n$$

which in matrix form is simply $\underline{v}^T\underline{A}_h\underline{u}_h = \underline{v}^T\underline{F}_h$, $\forall\underline{v} \in \mathbb{R}^n$, with $\underline{A}_h$ and $\underline{F}_h$ as defined earlier.

SLIDE 29

Take $\quad \underline{v} = (1 \;\; 0 \;\; \ldots \;\; 0)^T \quad \Rightarrow \quad \sum_{j=1}^{n} A_{h\,1\,j}\,u_{h\,j} = F_{h\,1}$

$\quad \underline{v} = (0 \;\; 1 \;\; \ldots \;\; 0)^T \quad \Rightarrow \quad \sum_{j=1}^{n} A_{h\,2\,j}\,u_{h\,j} = F_{h\,2}$

$\vdots$

$$\boxed{\underline{v}^T\,\underline{A}_h\,\underline{u}_h = \underline{v}^T\,\underline{F}_h\ ,\ \forall\,\underline{v} \in \mathbb{R}^n \ \Leftrightarrow\ \underline{A}_h\,\underline{u}_h = \underline{F}_h}$$

N12

If we took different $\underline{v}$ test vectors we would get different (even non-symmetric) equations — but the $\underline{u}_h$ would be the same. The test functions chosen here preserve the Galerkin recipe in which the test (v) and trial (u_h) spaces and bases are the same.

Note 12 — ***Weighted residual techniques (WRT)***

Given some general operator $\mathcal{L}$ and associated partial diffential equation $\mathcal{L}u = f$, a *weighted residual technique* looks for a $\hat{u} \in X_1 \subset X$ such that

$$\int_\Omega v\{\underbrace{\mathcal{L}\hat{u} - f}_{\text{residual}}\}\,dA = 0, \qquad \forall\,v \in X_2\ .$$

In particular, we no longer require that $\mathcal{L}u - f = 0$ in a *pointwise* sense, but rather in an *integral sense* relative to test functions v; we expect as X_1 and X_2

become richer $\hat{u}$ should approach u. Many different procedures can be devised based on different choices of X_1, X_2, and their associated bases. In the Galerkin procedure, $X_1 = X_2$.

Note we know from the previous lecture that it is much smarter, for our particular problem in which $\mathcal{L} = -\nabla^2 u$, to write

$$\int_\Omega v\, \mathcal{L}u \, dA = \int_\Omega \nabla u \cdot \nabla v \, dA$$

rather than $\int_\Omega v\, \mathcal{L}u\, dA = \int_\Omega -\nabla^2 u\, v\, dA$, since the former is more general, permits simpler approximation spaces (X_h is only $C^0(\Omega)$, not $C^1(\Omega)$), provides for automatic imposition of natural boundary conditions, ...

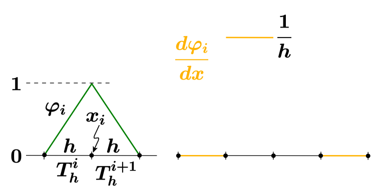
---

4 Discrete Equations

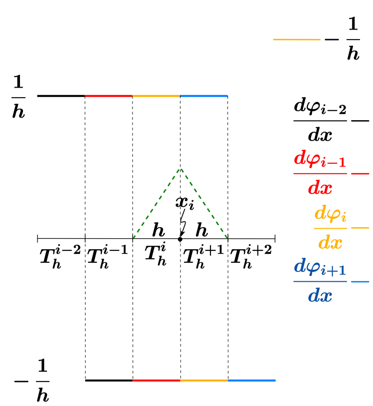
4.1 Matrix Elements: $\underline{A}_h$

4.1.1 φ_i and $d\varphi_i/dx$

SLIDE 30



SLIDE 31



4.1.2 Typical Row

SLIDE 32

$$A_{h\,i\,j} = \int_\Omega \frac{d\varphi_i}{dx}\frac{d\varphi_j}{dx}\, dx = \int_{T_h^i} \frac{d\varphi_i}{dx}\frac{d\varphi_j}{dx}\, dx + \int_{T_h^{i+1}} \frac{d\varphi_i}{dx}\frac{d\varphi_j}{dx}\, dx$$

is nonzero only for $j = i-1,\ i,\ i+1$

$$
\begin{aligned}
A_{h\,i\,i} &= \tfrac{1}{h^2}\,(h) + \tfrac{1}{h^2}\,(h) &&= \tfrac{2}{h} \\
A_{h\,i\,i-1} &= \tfrac{1}{h}\left(-\tfrac{1}{h}\right)(h) &&= -\tfrac{1}{h} \\
A_{h\,i\,i+1} &= \left(-\tfrac{1}{h}\right)\tfrac{1}{h}\,(h) &&= -\tfrac{1}{h}
\end{aligned}
$$

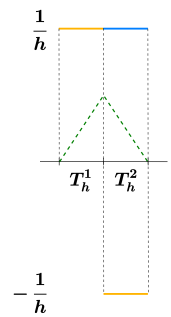
Sparsity naturally arises because the φ_j have very little and localized "support" — are nonzero only over a small patch of elements — and thus interact very little with each other.

4.1.3 Boundary Rows

SLIDE 33

$A_{h\,1\,1} = \frac{2}{h},\ A_{h\,1\,2} = -\frac{1}{h}$,

$A_{h\,n\,n} = \frac{2}{h},\ A_{h\,n\,n-1} = -\frac{1}{h}$.



4.1.4 Properties of $\underline{A}_h$

SLIDE 34

$$
\boxed{\underline{A}_h = \frac{1}{h}\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & 0 & & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}}
$$

$\underline{A}_h$ is SPD; *and*
diagonally dominant; *and*
sparse; *and*
tridiagonal.

The matrix here is the same as that from finite difference approximation except for a factor of h (which will of course appear on the right-hand side as well for consistency). We delay our first comparison with finite differences until the end of the next lecture.

4.2 "Load" Vector Elements: $\underline{F}_h$

SLIDE 35

General case, $\ell(v)$: $F_{h\,i} = \ell(\varphi_i)$

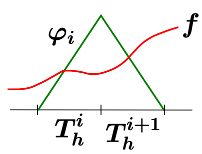
$$
\begin{aligned}
\text{Example:} \quad \ell(v) &= \langle \delta_{x_{i^*}}, v\rangle \\
\ell(\varphi_i) &= \langle \delta_{x_{i^*}}, \varphi_i\rangle \\
&= \varphi_i(x_{i^*}) \\
&= \delta_{i\,i^*}\ .
\end{aligned}
$$

Note i^* here is a particular node (placement at a point which is not a node is not such a good idea); $F_{hi} = \delta_{ii^*}$ means that $\underline{F}_h = \underline{0}$ except entry i^* which is unity.

SLIDE 36

Particular case, $\ell(v) \;=\; \displaystyle\int_\Omega f\, v\, dx$:

$$F_{h\,i} = \int_{T_h^i} f\,\varphi_i\,dx + \int_{T_h^{i+1}} f\,\varphi_i\,dx, \quad i = 1,\ldots,n\ ;$$



Numerical quadrature — "variational crime" — next lecture.

N13

Note 13 **Quadrature**

In fact, numerical quadrature is not "implementation," but really part of the formulation. (Of course if we perform the integrals exactly, that is not the case — but rarely can we do so in practice.) In particular, different quadratures will yield different $\underline{F}_h$ (and perhaps even $\underline{A}_h$ for more complicated problems), and thus different $\underline{u}_h$. In the next lecture we look at the two basic approaches: integration by interpolation; and Gauss quadrature. In general, in contrast to finite differences, we will not obtain $F_{h\,i} = f(x_i)$.

4.3 Summary

SLIDE 37

$\underline{u}_h \in \mathbb{R}^n$ satisfies

$$\frac{1}{h}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & 0 \\ & & \ddots & & \\ & 0 & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}\begin{pmatrix} u_{h\,1} \\ \vdots \\ u_{h\,n}\end{pmatrix} = \begin{pmatrix} F_{h\,i} \\ \vdots \\ F_{h\,n}\end{pmatrix}$$

$\underline{A}_h$ is known as the stiffness *matrix (inherited from structural analysis), or the* system *matrix, or the* global *matrix.*

▷ ***Exercise 4*** Consider the problem

$$\begin{aligned}
-u^{\mathrm{L}}_{xx} &= 0 && 0 < x < \tfrac{1}{2} \ ; \\
-2u^{\mathrm{R}}_{xx} &= 0 && \tfrac{1}{2} < x < 1 \ ; \\
-u^{\mathrm{L}}_{x}(\tfrac{1}{2}) + 1 &= -2u^{\mathrm{R}}_{x}(\tfrac{1}{2}) \ ; \\
u^{\mathrm{L}}(\tfrac{1}{2}) &= u^{\mathrm{R}}(\tfrac{1}{2}) \ ; \\
u^{\mathrm{L}}(0) &= u^{\mathrm{R}}(1) \ = \ 0 \ .
\end{aligned}$$

(a) Find the weak formulation of the above problem. *Hint*: see the examples and exercises of the previous lecture.

(b) Consider a triangulation $\mathcal{T}_h$ with two equi-sized elements, $T^1_h = (0, \frac{1}{2})$, $T^2_h = (\frac{1}{2}, 1)$. Find $\underline{A}_h$, $\underline{F}_h$, and hence u_h.

(c) Explain why $u_h = u$ with only two elements.

■

▷ ***Exercise 5*** Consider linear finite element discretization of the Neumann problem of Section 1.2 on a triangulation $\mathcal{T}_h$ of equi-sized elements T^k_h, $k = 1, \ldots, K = n$; the corresponding $n + 1$ nodes are given by $x_0 = 0, \ldots, x_n = 1$.

(a) Define X_h and the associated nodal basis.

(b) Find the discrete equations $\underline{A}_h\, \underline{u}_h = \underline{F}_h$ analogous to those of Slide 37 for the Dirichlet problem. (Note $\underline{A}_h \in \mathbb{R}^{n \times n}$, $\underline{u}_h \in \mathbb{R}^n$, $\underline{F}_h \in \mathbb{R}^n$.)

(c) Compare the n^{th} equation to what you might expect from finite differences.

■

▷ ***Exercise 6*** Consider (for the Dirichlet problem) approximation by finite elements in which v in each element is quadratic:

$$X_h = \{ v \in X \mid v|_{T_h} \in \mathbb{P}_2(T_h), \ \forall\, T_h \in \mathcal{T}_h \} \ .$$

(a) Find $\dim(X_h)$ by a counting argument. *Hint*: note that $v \in X_h$ are still only $C^0(\Omega)$.

(b) Introduce nodes not only at element boundaries but also at the midpoints of each element. Sketch several nodal basis functions φ_i: recall the $\varphi_i(x)$ are uniquely determined by the conditions that $\varphi_i(x) \in X_h$ and $\varphi_i(x_j) = \delta_{ij}$.

(c) Find the discrete equations for these quadratic elements analogous to those of Slide 37 for linear elements.

■

▷ ***Exercise 7*** For the fourth-order problem of Exercises 7 and 11 of the last lecture, show by a dimension/counting argument that piecewise linear elements do not lead to a viable conforming approximation. *Hint*: does it still suffice to require only $v \in C^0(\Omega)$? ■

5 The Mass Matrix

5.1 Motivation

5.1.1 Definition

SLIDE 38

$\underline{M}_h \in \mathbb{R}^{n \times n}$:

$$M_{h\,i\,j} = \underbrace{\int_\Omega \varphi_i \varphi_j \, dx}_{\text{originating form: } (w,v)_{L^2(\Omega)}} \qquad ;$$

the finite element "identity" (I) operator.

The mass matrix will have the form above — and the stiffness matrix $A_{h\,i\,j}$ will be given by $a(\varphi_i, \varphi_j)$ — for any *choice of basis functions φ_i; however, unless otherwise indicated, we will assume that the φ_i refer to our particular nodal basis functions.*

5.1.2 "Applications"

SLIDE 39

$\underline{M}_h$ appears where the identity appears

- as part of differential operator, $-u_{xx} + Iu = f$; $\boxed{\text{E8}}$

- in eigenvalue problems, $-u_{xx} = \lambda\, I u$;
- in parabolic PDEs, $I \dfrac{\partial u}{\partial t} = \nabla^2 u$;
- in quadrature by interpolation.

We know that in the Galerkin procedure we find the discrete equations simply by replacing u by φ_j and v by φ_i; in each of the above cases the Identity term will give rise to a weak form $(v, u)_{L^2(\Omega)}$ and hence mass matrix. We will see this more clearly in each particular application.

▷ ***Exercise 8*** Consider the "good" Helmholtz problem

$$\begin{aligned} -u_{xx} + \gamma^2 u &= f \qquad \text{in } \Omega = (0,1)\ , \\ u(0) \;=\; u(1) &= 0\ , \end{aligned}$$

with $\gamma^2 \geq 0$ (and real).

(a) Find the minimization statement and weak form associated with this problem: specify X; specify a and ℓ (and hence J); show that a is SPD. *Hint*: multiply the equation by $v \in X$ and integrate the Laplacian term by parts.

(b) Show that the stiffness matrix $\underline{A}_h^{\text{Helmholtz}} = \underline{A}_h^{\text{Laplacian}} + \gamma^2 \underline{M}_h$, where $\underline{A}_h^{\text{Laplacian}}$ is the stiffness matrix for the Laplacian as given in Slide 37.

■

5.2 Properties

5.2.1 General

SLIDE 40

$\underline{M}_h$ is SPD:

$$\begin{aligned} \underline{v}^T \underline{M}\, \underline{v} &= \sum_{i=1}^{n} v_i \sum_{j=1}^{n} v_j \int_0^1 \varphi_i\, \varphi_j\, dx \\ &= \int_0^1 \sum_{i=1}^{n} v_i\, \varphi_i \sum_{j=1}^{n} v_j\, \varphi_j\, dx \\ &= \int_0^1 \underbrace{\left(\sum_{i=1}^{n} v_i\, \varphi_i \right)^2}_{v \in X_h} dx > 0 \quad \underbrace{\text{if } \underline{v} \neq 0}_{\varphi_i \text{ are } basis}\ . \end{aligned}$$

5.2.2 Particular

For linear elements, nodal basis: $\boxed{\text{E9}}$

$$\underline{M}_h = h \begin{pmatrix} \frac{2}{3} & \frac{1}{6} & & & \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & & 0 \\ & & \ddots & & \\ & 0 & & \frac{2}{3} & \frac{1}{6} \\ & & & \frac{1}{6} & \frac{2}{3} \end{pmatrix}$$

sparse, banded, tri-diagonal — "close" to $\underline{I}$.

The mass matrix takes its name (I presume) from its appearance in dynamical equations (or the heat equation) with the mass term. We can also see that if we sum all the entries of $\underline{M}_h$ we get something very close to unity — the volume (or mass) of Ω. We do not get exactly unity only because of the Dirichlet boundary conditions; we revisit this in the next lecture.

▷ ***Exercise 9*** Hierarchical basis functions, in which the finite element coefficients have only a Rayleigh-Ritz role, and no nodal interpretation, are often attractive for adaptive purposes, as, unlike in our nodal system, refinement is achieved without modifying the basis functions already present. Hierarchical basis functions can also achieve very sparse matrix structure in certain "orthogonal" cases, and are also often well-conditioned.

We consider the homogeneous Dirichlet problem on the domain $\Omega = (-1, 1)$, for which the continuous and finite element spaces are given by $X = H_0^1(\Omega) = \{v \in H^1(\Omega) \mid v(-1) = v(1) = 0\}$ and $X_h = \{v \in X \mid v|_{T_h^k} \in \mathbb{P}_1(T_h^k),\ k = 1, \ldots, K\}$, respectively. We choose $K = 2^{L+1}$ elements of equal length $h = 2^{-L}$, for L an integer greater than zero. Note that we are *not* changing our finite element approximation, X_h, but only the way in which we represent X_h in terms of a basis. In particular, as before, $u_h \in X_h$ satisfies $a(u_h, v) = \ell(v)$, $\forall v \in X_h$, for the $a(\cdot,\cdot)$ and $\ell(\cdot)$ corresponding to our problem of interest.

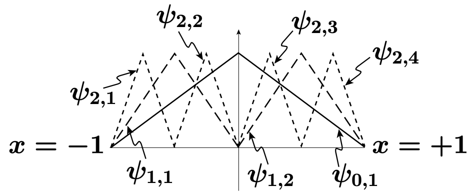
To describe the new basis, we first introduce a "mother" function $\Psi\colon \mathbb{R} \to \mathbb{R}$

$$\Psi(x) = \begin{cases} 1 - |x| & |x| \le 1 \\ 0 & |x| > 1 \end{cases},$$

which is a hat function defined over the entire real line. Now, "wavelet-like," we introduce 2^m shifted and squeezed daughter functions at each of L levels,

$$\psi_{m,j}(x) = \Psi\left(2^m\left(x + 1 - \frac{2j-1}{2^m}\right)\right)\Bigg|_\Omega \qquad \begin{array}{rcl} j &=& 1, \ldots, 2^m \\ m &=& 0, \ldots, L \end{array}.$$

The $\psi_{m,j}(x)$ are centered at $-1 + (2j-1)/2^m$, with half-width 2^{-m}, as shown in the figure. At level $m = 0$ we have the mother function; at level $m = 1$ we have two copies each with half the support; at level m we have 2^m copies each with 2^{-m} the support.



To proceed algebraically, we need to collapse our basis functions to a single index, which we define as

$$\text{ind}\,(m,j) = j + (2^m - 1), \qquad j = 1,\ldots,2^m, \quad m = 0,\ldots,L\ ,$$

and then write

$$\chi_{\text{ind}(m,j)} = \psi_{m,j}, \qquad j = 1,\ldots,2^m, \quad m = 0,\ldots,L\ .$$

Note that $\text{ind}\,(L, 2^L) = 2^{L+1} - 1 \equiv n$, and we thus have $2^{L+1} - 1 = K - 1$ basis functions; note also that the $\chi_i(x)$, $i = 1,\ldots,n$, all satisfy the homogeneous Dirichlet boundary conditions.

(a) Argue that the $\chi_i(x)$, $i = 1,\ldots,n$, form a basis for X_h, that is

$$X_h = \text{span}\,\{\chi_i(x),\ i = 1,\ldots,n\}\ ,$$

or equivalently, for every $w \in X_h$,

$$w(x) = \sum_{i=1}^{n} w_i\, \chi_i(x)\ ,$$

for some unique vector $\underline{w} \in \mathbb{R}^n$. Note $n = K - 1 = \dim(X_h)$, as must be the case. Hint: show that $w(x)$ of above is linear over each element and continuous over Ω; construct a unique correspondence between the nodal values $w(x_j)$ and the w_i by "peeling" off each level of the hierarchy (start with the mother).

(b) Consider the particular problem

$$\begin{aligned} -u_{xx} &= f\ , \\ u(-1) = u(1) &= 0\ . \end{aligned}$$

Apply the Rayleigh-Ritz procedure with the basis functions $\chi_i(x)$, $i = 1,\ldots,n$, to find the discrete equations

$$\underline{A}_h\, \underline{u}_h = \underline{F}_h\ ,$$

where, as always,

$$(A_h)_{ij} = \int_{-1}^{1} \frac{d\chi_i}{dx}\frac{d\chi_j}{dx}\,dx, \qquad 1 \le i,j \le n,$$

$$(F_h)_i = \int_{-1}^{1} \chi_i\, f\, dx, \qquad i = 1,\ldots,n\ ,$$

and $\underline{u}_h \in \mathbb{R}^n$ are the coefficients of the finite element solution

$$u_h(x) = \sum_{i=1}^{n} u_{h\,i}\,\chi_i(x)\ .$$

In particular, give an explicit expression for $\underline{A}_h$, and discuss the structure of this system matrix relative to that associated with the nodal basis. Hint: plot the derivatives $\frac{d\chi_i}{dx}$, which resemble "Haar" functions; consider how these derivatives interact at the same and different levels of the hierarchy.

(c) Repeat Part (b) for the modified problem (see Exercise 8)

$$\begin{aligned} -u_{xx} + u &= f\ , \\ u(-1) = u(1) &= 0\ . \end{aligned}$$

Is the good matrix structure obtained in Part (b) "robust"? You need not give an explicit expression for $\underline{A}_h$ in this case, but you should clearly identify the sparsity structure.

■