The collocation method can be viewed as a special case of a more general method known as the method of weighted residuals (MWR)

Consider again the BVP

$$\begin{cases} -\dfrac{d}{dx}\left[a(x)\dfrac{du}{dx}\right] + b(x)u(x) = f(x), & a(x) > 0 \\ u(0) = \alpha \\ u(1) = \beta \end{cases}$$

Define a residual, or "defect"

$$r(x;\vec{c}) \equiv -\frac{d}{dx}\left[a(x)\frac{d\hat{u}}{dx}\right] + b(x)\hat{u}(x) - f(x)$$

where $\hat{u}(x) \equiv \sum_{l=1}^{N} c_l \phi_l(x)$ ϕ_l = basis functions for some n-dimensional function space.

so that

$$r(x;\vec{c}) = -\sum_{l=1}^{N} c_l \frac{d}{dx}\left[a(x)\frac{d\phi_l}{dx}\right] + \sum_{l=1}^{N} c_l\, b(x)\phi_l(x) - f(x)$$

In addition to this residual we need to handle the boundary conditions $u(0)=\alpha$, $u(1)=\beta$

Thinking along the lines of least squares we might require

$$R \equiv \frac{1}{2}\int_0^1 r(x;\vec{c})^2\,dx$$

to be minimized with respect to $\vec{c}$ subject to the constraints

$$\sum_{\ell=1}^{N} c_\ell \phi_\ell(0) = \alpha$$

and

$$\sum_{\ell=1}^{N} c_\ell \phi_\ell(1) = \beta$$

So let

$$\mathcal{L} \equiv R + \lambda_0\left(\sum_{\ell=1}^{N} c_\ell\phi_\ell(0)-\alpha\right) + \lambda_1\left(\sum_{\ell=1}^{N} c_\ell\phi_\ell(1)-\beta\right)$$

with Lagrange multipliers λ_0 and λ_1.

To minimize, require

$$\frac{\partial\mathcal{L}}{\partial c_\ell} = \int_0^1 r(x;\vec{c})\frac{\partial r}{\partial c_\ell}\,dx + \lambda_0\phi_\ell(0) + \lambda_1\phi_\ell(1) = 0$$

$$\frac{\partial\mathcal{L}}{\partial\lambda_0} = \sum_{\ell=1}^{N} c_\ell\phi_\ell(0)-\alpha = 0$$

$$\frac{\partial\mathcal{L}}{\partial\lambda_1} = \sum_{\ell=1}^{N} c_\ell\phi_\ell(1)-\beta = 0$$

$$0 = \frac{\partial \mathcal{L}}{\partial c_\ell} = \int_0^1 r(x;\vec{c}) \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_\ell}{dx}\right] + b(x)\phi_\ell(x) \right\} dx + \lambda_0 \phi_\ell(0) + \lambda_1 \phi_\ell(1)$$

$$= \int_0^1 \left(\sum_{i=1}^{N} c_i \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_i}{dx}\right] + b(x)\phi_i(x) \right\} - f \right) \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_\ell}{dx}\right] + b(x)\phi_\ell(x) \right\} dx$$

$$+ \lambda_0 \phi_\ell(0) + \lambda_1 \phi_\ell(1)$$

$$\boxed{0 = \sum_{i=1}^{N} c_i \underbrace{\int_0^1 \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_i}{dx}\right] + b(x)\phi_i(x) \right\} \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_\ell}{dx}\right] + b(x)\phi_\ell(x) \right\} dx}_{} - \underbrace{\int_0^1 f(x) \left\{ -\frac{d}{dx}\left[a(x)\frac{d\phi_\ell}{dx}\right] + b(x)\phi_\ell(x) \right\} dx}_{} + \lambda_0 \phi_\ell(0) + \lambda_1 \phi_\ell(1)}$$

plus the b.c. constraints

$$\boxed{\sum_{\ell=1}^{N} c_\ell \phi_\ell(0) = \alpha}$$

$$\boxed{\sum_{\ell=1}^{N} c_\ell \phi_\ell(1) = \beta}$$

<u>Comments</u>

- this is just a big linear system for the unknowns $c_1, c_2, \ldots, c_N$ and λ_0, λ_1.
- the b.c., at least as written, tend to make the problem non-symmetric — more thoughts on b.c. to come...

For the moment not worrying about boundary conditions the least squares approach and the collocation method are special cases of a more general method called the method of weighted residuals . [redacted] , The MWR requires

$$0 = \int_0^1 r(x;\vec{c})\, W_j(x)\, dx \qquad j = 1, 2, \ldots, N$$

(or possibly $j = 2, \ldots, N-1$ if we apply b.c. as in collocation method e.g. $\sum_{\ell=1}^{N} c_\ell \phi_\ell(0) = \alpha$, $\sum_{\ell=1}^{N} c_\ell \phi_\ell(1) = \beta$)

where $W_j(x)$ is a "weighting" function and

$$r(x;\vec{c}) = -\frac{d}{dx}\left[a(x)\frac{d}{dx}\left(\sum_{\ell=1}^{N} c_\ell \phi_\ell(x)\right)\right] + b(x)\sum_{\ell=1}^{N} c_\ell \phi_\ell(x) - f(x)$$

- If $W_j(x) \equiv \delta(x - x_j)$ Dirac Delta Function

$$0 = \int_0^1 r(x;\vec{c})\, \delta(x - x_j)\, dx = r(x_j;\vec{c})$$

i.e. residual required to be zero at specific collocation points x_j.

- If $\quad w_j(x) = -\frac{d}{dx}\left[a(x)\,\frac{d\phi_j}{dx}\right] + b(x)\,\phi_j$

then we have (modulo b.c. issues) the least squares approach.

- If $\quad w_j(x) = \phi_j(x)$ (i.e. the weight function is the basis function)

so $\quad \int_0^1 r(x;\vec{c})\,\phi_j(x)\,dx = 0$

This choice is behind the <u>Galerkin</u> method

— Comments:

- in contrast to the least squares and collocation methods the Galerkin method has the advantage that — once integration by parts on the above integral has been applied — the resulting system of equations does not involve 2^{nd} derivatives of the basis functions (only first derivatives).

$\Rightarrow$ ϕ_ℓ need not be twice differentiable.

BVP $\left[\begin{array}{l} -\frac{d}{dx}\left[a(x)\frac{du}{dx}\right] + b(x)u(x) = f(x) \\ u(0) = \alpha \\ u(1) = \beta \end{array}\right.$

We'll proceed with the Galerkin method [redacted] using the assumed form of the solution approximation

$$\boxed{u(x) \approx U_m(x) \equiv \phi_0(x) + \sum_{\ell=1}^{m} c_\ell \phi_\ell(x)}$$

where $\phi_0(0) = \alpha$ and $\phi_0(1) = \beta$ (e.g. $\phi_0(x) = \alpha + (\beta-\alpha)x$)

and $\phi_\ell(0) = \phi_\ell(1) = 0 \quad \ell = 1, 2, \ldots, m$.

$\phi_\ell \ (\ell = 1, 2, \ldots, m)$ = basis functions in m-dimensional function space. (to be specified)

Comments

- Notice that with these assumptions about ϕ_0 and ϕ_ℓ the b.c. are automatically satisfied by $U_m(x)$.
- A simple choice of monomial basis $\{1, t, t^2, \ldots\}$ will not be suitable here

Again, define the residual

$$r(x;\vec{c}) = -\frac{d}{dx}\left[a(x)\frac{dU_m}{dx}\right] + b(x)U_m(x) - f(x)$$

$$= -\frac{d}{dx}\left[a(x)\frac{d}{dx}\left(\sum_{\ell=1}^{m} c_\ell \phi_\ell(x)\right)\right] + b(x)\sum_{\ell=1}^{m} c_\ell \phi_\ell(x)$$

$$-\frac{d}{dx}\left[a(x)\frac{d\phi_0}{dx}\right] + b(x)\phi_0(x) - f(x) \Bigg\} \text{ this is like a modified r.h.s. function } \tilde{f}.$$

Note: If we define $u(x) = \phi_0(x) + \tilde{u}(x)$ then $\tilde{u}(x)$ satisfies the BVP

$$\begin{cases} -\frac{d}{dx}\left[a(x)\frac{d\tilde{u}}{dx}\right] + b(x)\tilde{u}(x) = \tilde{f}(x) \\ \tilde{u}(0) = 0 \\ \tilde{u}(1) = 0 \end{cases}$$

where $\tilde{f}(x) \equiv f(x) + \frac{d}{dx}\left[a(x)\frac{d\phi_0}{dx}\right] - b(x)\phi_0(x)$

So we are equivalently seeking a solution approximation

$$\tilde{u}(x) \approx \tilde{u}_m(x) \equiv \sum_{\ell=1}^{m} c_\ell \phi_\ell(x) \qquad \boxed{\phi_\ell \in \mathring{H}_m}$$

recall $\mathring{H}_m = \text{span}\{\phi_1, \dots, \phi_m\}$ where the functions $\phi_1, \dots, \phi_m$ satisfy b.c. $\phi_\ell(0) = \phi_\ell(1) = 0$.

more details to follow

where $\phi_\ell(0) = \phi_\ell(1) = 0$ for $\ell = 1, 2, \dots, m$.

Now, the residual $r(x;\vec{c})$ is

$$r(x;\vec{c}) = -\frac{d}{dx}\left[a(x)\frac{d\tilde{u}_m}{dx}\right] + b(x)\tilde{u}_m(x) - \tilde{f}(x)$$

If we equip the finite dimensional space $\mathring{H}_m$ with an inner product $\langle \cdot, \cdot \rangle$ the Galerkin method can be stated as requiring

$$\langle r(x;\vec{c}), \phi_k \rangle = 0 \quad \text{for } k = 1, 2, \dots, m$$

This is the second key principle for FEM

$$\boxed{\text{Choose the approximation so that the residual is orthogonal to the space } \mathring{H}_m}$$

To fix ideas, let us for now consider the space $L^2(\Omega)$

$$L^2([0,1]) = \left\{ f : [0,1] \to \mathbb{R} \;\middle|\; \int_0^1 f^2 \, dx < \infty \right\}$$

with the standard Euclidean inner product over functions

$$\langle f, g \rangle = \int_0^1 f(x) g(x) \, dx$$

The Galerkin method requires for $k = 1, 2, \ldots, m$ that

$$0 = \int_0^1 \underbrace{\left\{ -\frac{d}{dx}\left[a(x) \frac{d}{dx}\left(\sum_{\ell=1}^{m} c_\ell \phi_\ell(x) \right) \right] + b(x) \left(\sum_{\ell=1}^{m} c_\ell \phi_\ell(x) \right) - \tilde{f}(x) \right\}}_{r(x;\vec{c})} \phi_k(x) \, dx$$

This again amounts to a linear system for the unknowns c_ℓ, $\ell = 1, 2, \ldots, m$.

Comment

- For a more detailed discussion of the Galerkin method see Gockenbach ('Understanding and Implementing the Finite Element Method', SIAM) p. 57 (Ch. 3)

Before implementing this linear system we rewrite it using integration by parts:

$$0 = \sum_{\ell=1}^{m} c_\ell \int_0^1 \Big(- \underbrace{\phi_k(x)}_{u} \underbrace{\frac{d}{dx}\Big[a(x)\frac{d\phi_\ell}{dx}\Big]}_{dv} + \phi_k(x)\,b(x)\,\phi_\ell(x) \Big)\,dx$$

$$- \int_0^1 \tilde{f}(x)\,\phi_k(x)\,dx$$

by property of basis function $\phi_\ell(0) = \phi_\ell(1) = 0$

$$= \sum_{\ell=1}^{m} c_\ell \Bigg\{ \cancelto{0}{-\phi_k(x)\,a(x)\frac{d\phi_\ell}{dx}\Big|_0^1} + \int_0^1 a(x)\frac{d\phi_k}{dx}\frac{d\phi_\ell}{dx}\,dx$$

$$+ \int_0^1 b(x)\,\phi_k(x)\,\phi_\ell(x)\,dx \Bigg\} - \int_0^1 \tilde{f}(x)\,\phi_k(x)\,dx$$

$$0 = \sum_{\ell=1}^{m} c_\ell \left\{ \int_0^1 a(x)\frac{d\phi_k}{dx}\frac{d\phi_\ell}{dx} + b(x)\,\phi_k(x)\,\phi_\ell(x)\,dx \right\} - \int_0^1 \tilde{f}(x)\,\phi_k(x)\,dx$$

This is the linear system we'll solve. It takes the form

$$\boxed{A\tilde{c} = b} \qquad \begin{bmatrix} & A_{k\ell} & \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$A_{k\ell} = \int_0^1 \left[a(x)\frac{d\phi_k}{dx}\frac{d\phi_\ell}{dx} + b(x)\,\phi_k(x)\,\phi_\ell(x) \right] dx$$

$$b_k = \int_0^1 \tilde{f}(x)\,\phi_k(x)\,dx$$

Comments

- The form of the equations we actually solve suggest that the basis functions $\phi_e(x)$ need not be twice differentiable — recall this was not the case for collocation or "Least squares" methods.
- In fact, because the derivatives $\frac{d\phi_e}{dx}$ appear inside the integral we need only require that the basis functions $\phi_e(x)$ be <u>piecewise-differentiable</u> in $[0,1]$

This is the third key principle for FEM

Integrate by parts to depress to the maximal extent possible the differentiability requirements of the space $\mathring{H}_m$ (i.e. the functions $\phi_1, \phi_2, \dots, \phi_m$)

The previous ideas developed are linked to what are known as a "weak solution" of a boundary value problem (or the weak form of a BVP).

<u>The weak form of a BVP</u> (see Brockenbach, p.20 →)

We'll generalize the discussion to 2D but will eventually go back to the 1D case...

Consider the BVP

Dirichlet problem
$$\begin{cases} -\nabla\cdot(a\nabla u) + bu = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \qquad (*) \text{ Strong form}$$

$a > 0$ (strictly positive)

Assuming f is continuous and a, b are sufficiently smooth we expect that the solution u and its partial derivatives up to order 2 (eg $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial^2 u}{\partial x^2}, \dots$) are continuous and

So we seek a solution to $(*)$ strong form from the space

$$C^2_D(\bar{\Omega}) = \left\{ u \in C^2(\bar{\Omega}) \mid u = 0 \text{ on } \partial\Omega \right\}$$

= set of all continuous functions defined on $\bar{\Omega} = \Omega \cup \partial\Omega$ (closure of Ω) with the property that u and all its partial derivatives of up to order 2 are continuous on $\bar{\Omega}$ and u satisfies Dirichlet b.c. on $\partial\Omega$.

So suppose u is a solution to ⊛strong form.

Then for any function v defined on Ω with suitable properties with respect to integration on Ω we have

$$-\nabla\cdot(a\nabla u)v + buv = fv \quad \text{in } \Omega$$

So that

$$\boxed{-\int_{\Omega}\nabla\cdot(a\nabla u)v\,d\vec{x} + \int_{\Omega}buv\,d\vec{x} = \int_{\Omega}fv} \qquad ⊛_{INT}$$

In this context v is known as a test function. The idea of the <u>weak form</u> of the BVP is to require that this integral equality hold for all test functions v in a sufficiently broad set of functions.

$C^2_D(\bar{\Omega})$ is one possible set of functions but we'll end up using ones that are less restrictive.

Q: How does satisfying $⊛_{INT}$ for all "test functions" assure us that u satisfies ⊛strong form?

A: Suppose $\vec{x}_0 = (x_0,y_0) \in \Omega$ and $\delta > 0$ small enough so that the open ball of radius δ centered at $\vec{x}_0$, i.e.

$$B_\delta(\vec{x}_0) = \{\vec{x}\in\mathbb{R}^2 \mid \|\vec{x}-\vec{x}_0\| < \delta\}$$

is contained in Ω (i.e. $B_\delta(\vec{x}_0) \subset \Omega$)

Now consider any function $v \in C_D^2(\Omega)$ with the properties

(1) $v(\vec{x}) > 0$ for all $\vec{x} \in B_\delta(\vec{x}_0)$

(2) $v(\vec{x}) = 0$ for all $\vec{x} \notin B_\delta(\vec{x}_0)$

(3) $$\int_\Omega v(\vec{x})\,d\vec{x} = \int_{B_\delta} v(\vec{x})\,d\vec{x} = 1$$

(See Hockenbach for an argument of the existence of such functions) p.23

Then

$$\xrightarrow{\delta \to 0}$$

- $$\int_\Omega f v\,d\vec{x} = \int_{B_\delta} f v\,d\vec{x} \approx f(\vec{x}_0)\int_{B_\delta} v\,d\vec{x} = f(\vec{x}_0)$$

- $$\int_\Omega b u v\,d\vec{x} = \int_{B_\delta} b u v\,d\vec{x} \approx b(\vec{x}_0)u(\vec{x}_0)\int_{B_\delta} v\,d\vec{x} = b(\vec{x}_0)u(\vec{x}_0)$$

- $$-\int_\Omega \nabla\cdot(a\nabla u)v\,dx = -\int_{B_\delta} \nabla\cdot(a\nabla u)v\,dx \approx \ldots = -\nabla\cdot(a\nabla u)\Big|_{\vec{x}_0}.$$

So at each point $\vec{x}_0 \in \Omega$ if $\circledast_{INT}$ holds we have

$$-\nabla\cdot(a\nabla u) + bu = f \quad \text{in } \Omega$$
$$u = 0 \quad \text{on } \partial\Omega$$

That is, the strong form, $\circledast_{\text{strong form}}$, holds. Converse is also true $\circledast_{\text{strong}} \to \circledast_{INT}$

Therefore $\circledast_{INT}$ holds for some $u \in C^2_D(\bar{\Omega})$ and all $v \in C^2_D(\bar{\Omega})$ if and only if $\circledast_{\text{strong form}}$ holds.

(~~note: if $\circledast_{INT}$ holds for all $v \in C^2_D(\bar{\Omega})$ then it holds for the ones with properties (1), (2) and (3)~~).

But now observe that for all $u, v \in C^2_D(\bar{\Omega})$ $\circledast_{INT}$ can be rewritten using the divergence theorem

$$\cdots \int_\Omega \nabla\cdot(a\nabla u)v\,d\vec{x} = \int_\Omega \Big[\nabla\cdot\big(va\nabla u\big) - a\nabla u\cdot\nabla v\Big]\,d\vec{x}$$

$$= \int_{\partial\Omega} \cancelto{0}{v a\nabla u\cdot\hat{n}\,dS} - \int_\Omega a\nabla u\cdot\nabla v\,d\vec{x}$$

since $v \in C^2_D(\bar{\Omega})$
and so $v=0$ on $\partial\Omega$

So the problem $\circledast_{INT}$ becomes the problem

$$\boxed{\begin{array}{l}\text{find } u \in C^2_D(\bar{\Omega}) \text{ such that} \\[6pt] \displaystyle\int_\Omega \big(a\nabla u\cdot\nabla v + buv\big)\,d\vec{x} = \int_\Omega f v\,d\vec{x} \\[6pt] \text{for all } v \in C^2_D(\bar{\Omega}) \qquad \circledast_{\text{weak form}}\end{array}}$$

So u satisfies $\circledast_{\text{strong form}}$ if and only if it satisfies $\circledast_{\text{weak form}}$.

While the original PDE (BVP) seems to suggest that the solution u be twice differentiable, a, [redacted] b & f sufficiently smooth as well note that these conditions would seem to be too strong for what is needed to evaluate the integrals in ⊛ weak form. The weak form really only requires...

$f\,v$ be integrable

$b\,u\,v$ be integrable

$a\nabla u\cdot\nabla v$ be integrable (so [redacted] u, v can be piecewise differentiable)

So, if we work with the <u>weak form</u> of the BVP it is natural to make the weakest possible assumptions about the functions involved.

This leads to the use of <u>Sobolev spaces</u> in this context. The space of square-integrable functions is a good place to start. Note, if [redacted] $v=f$ then we need $\int_\Omega v^2\,d\vec{x}$ to be integrable

if [redacted] $v=u$ then we [redacted] need (for sufficiently smooth a,b)

$$\int\left(\frac{\partial u}{\partial x}\right)^2 d\vec{x} \qquad \int\left(\frac{\partial v}{\partial x}\right)^2 d\vec{x}$$ to be integrable

etc.

Let

$$\boxed{L^2(\Omega) = \left\{ v : \Omega \to \mathbb{R} \;\middle|\; \int_\Omega v^2 d\vec{x} < \infty \right\}}$$

A Hilbert Space (complete inner product space)

So we'll require $f \in L^2(\Omega)$. Further, we'll require that

$$u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \in L^2(\Omega)$$

as well as

$$v, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y} \in L^2(\Omega)$$

These conditions define the Sobolev Space $H^1(\Omega)$

$$\boxed{H^1(\Omega) = \left\{ v \in L^2(\Omega) \;\middle|\; \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y} \in L^2(\Omega) \right\}}$$

(i.e. the set of functions in $L^2(\Omega)$ whose first partial derivatives are also in $L^2(\Omega)$).

In the present context, we'll also define the Sobolev Space $H^1_0(\Omega)$ as

$$\boxed{H^1_0(\Omega) = \left\{ v \in H^1(\Omega) \;\middle|\; v = 0 \text{ on } \partial\Omega \right\}}$$

The spaces L^2, H^1 and H^1_0 are complete
(every Cauchy sequence converges to an element in the space)

Gockenbach p.39

For more details on these function spaces see...

- Gockenbach; p. 27, 35-44, 27-29
- J.P. Keener; "Principles of Applied Mathematics: Transformation + Approximation" (Addison-Wesley); p. 66
- Iserles; p. 439-441

So the "weakened" version of the weak form of the BVP is now stated as

Find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \left(a\nabla u\cdot\nabla v + buv\right) d\vec{x} = \int_\Omega f v\, d\vec{x}$$

for all $v \in H_0^1(\Omega)$

(*) weak form

(see Gockenbach, p. 27)

So let's return now to 1D [redacted]

Consider

$$\begin{cases} -\dfrac{d}{dx}\left[a(x)\dfrac{du}{dx}\right] + b(x)u(x) = f(x) \\ u(0) = 0 \\ u(1) = 0 \end{cases}$$

(we'll just deal with the case of homogeneous Dirichlet conditions but recall the connection between "u" and "$\tilde{u}$")

The weak form of this BVP is ... find $u \in H_0^1([0,1])$

such that
$$\int_0^1 \left(a\frac{du}{dx}\frac{dv}{dx} + buv\right)dx = \int_0^1 fv\,dx$$

for all $v \in H_0^1([0,1])$

__Comments__

- ~~Recall that this is basically the statement~~

  $$\langle r(u), v\rangle = 0 \quad \text{for all } v \in H_0^1$$

  where
  $$r(u) = -\frac{d}{dx}\left[a\frac{du}{dx}\right] + bu - f \quad : \text{ residual or 'defect'}$$

  but after integration by parts

- H_0^1 is an example of the type of space Iserles refers to with $\mathring{H}$ (see p.174).

We return to our approximation (drop tildes)

$$u(x) \approx u_m(x) = \sum_{\ell=1}^{m} c_\ell \phi_\ell(x)$$

where $\phi_\ell \in \mathring{H}_m$ (finite dimensional space to be defined)

and where the Galerkin method requires (notes, p. 230)

$$\boxed{\begin{aligned} 0 = \sum_{\ell=1}^{m} c_\ell \left\{ \int_0^1 a(x) \frac{d\phi_k}{dx} \frac{d\phi_\ell}{dx} + b(x)\,\phi_k(x)\,\phi_\ell(x)\,dx \right\} \\ - \int_0^1 f(x)\,\phi_k(x)\,dx \\ \text{for } k = 1,2,\dots,m \qquad \boxed{\phi_\ell \in \mathring{H}_m \quad \begin{matrix}\phi_\ell(0)=\phi_\ell(1)=0 \\ \ell=1,2,\dots,m\end{matrix}} \end{aligned}}$$

The next task is to address the choice of the space $\mathring{H}_m$ and the basis functions $\phi_\ell(x)$.

Some choices:
- functions with global support - spectral methods
  (e.g. sin/cos, ...) [high accuracy with few points (small m), dense matrices]
- Functions with local support · F.E.M.
  [sparse matrices, high accuracy requires larger m]

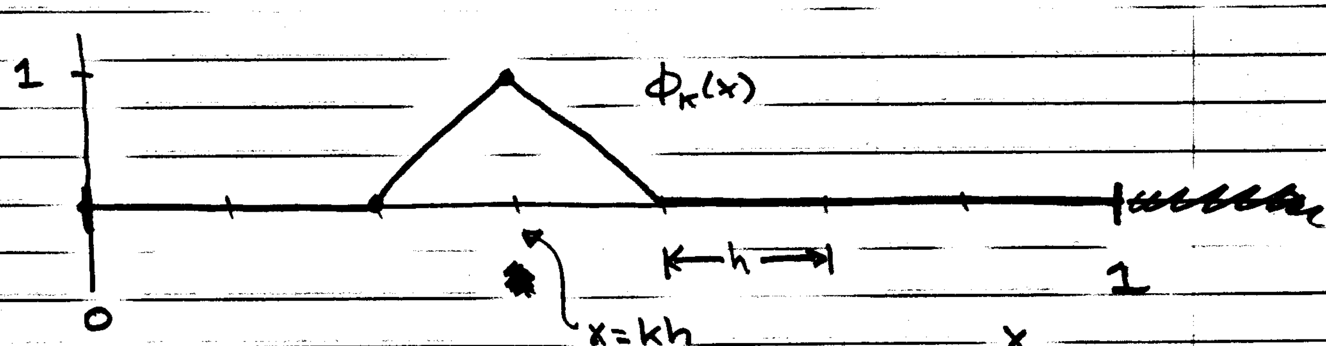
Choice of $\mathring{H}_m$ and basis functions $\phi_\ell(x)$:

- to ease the computations necessary to set up and solve the linear system for c_ℓ $(\ell=1,\dots,m)$ defined on the previous page, we adopt our next key principle of the FEM

$$\boxed{\text{Choose each } \phi_K \text{ so that it vanishes along most of the interval } (0,1) \text{ thereby ensuring that } \phi_K \phi_\ell = 0 \text{ for most choices of } k,\ell = 1,2,\dots,m.}$$

- Recall that the weak form required only that the test functions be piecewise differentiable. We'll choose the basis functions with this property as well.

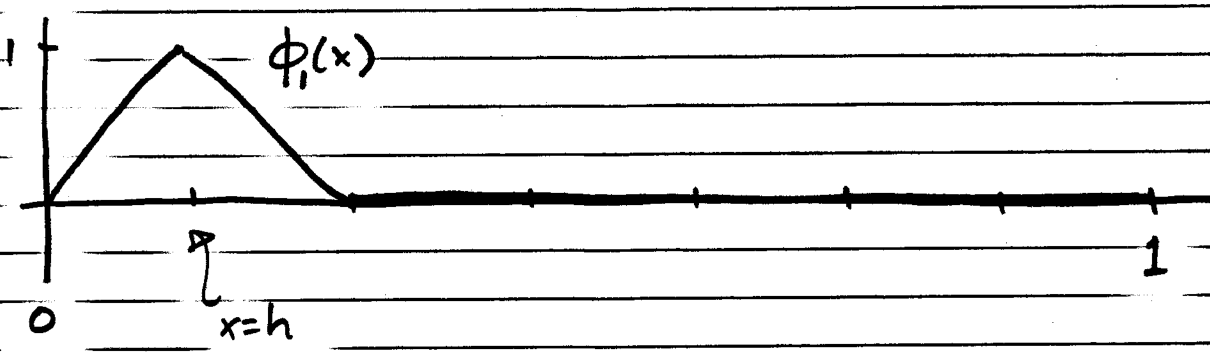
  - our choice here will be the "hat function"

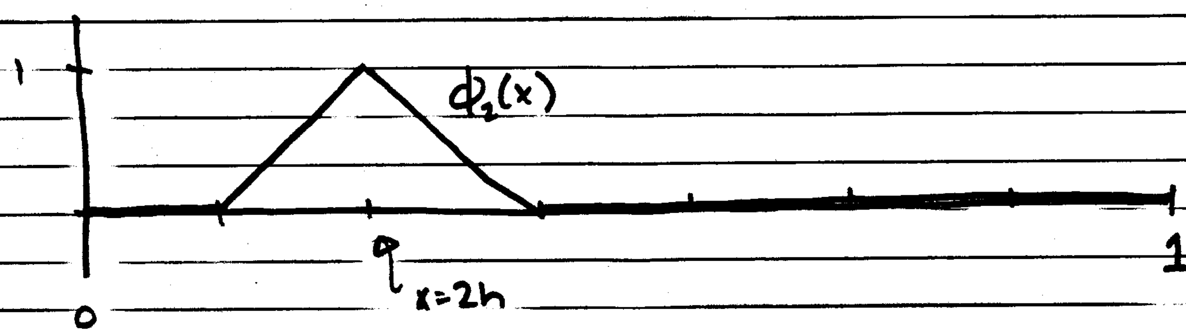


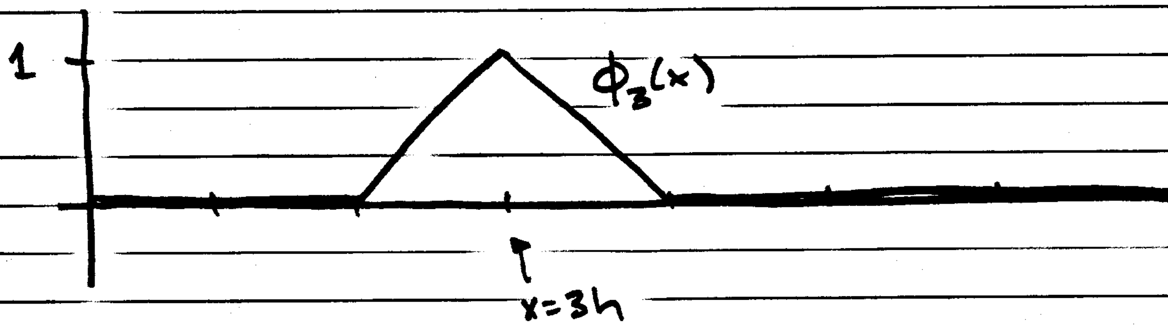
$h = \frac{1}{m+1}$

$(j=0)$ $(j=1)$ $(j=2)$ . . . $(j=m)$ $(j=m+1)$

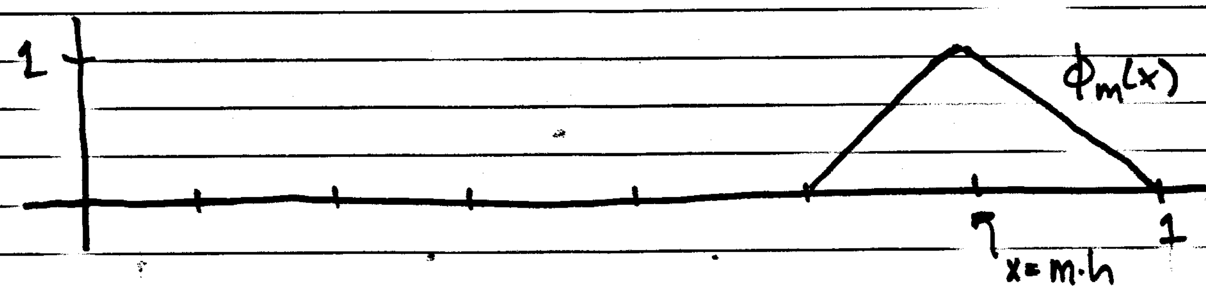
That is,







etc.



• (but see also discussion of B-splines ... e.g. Heath, p. 330, 435)

These can be expressed analytically as

$$\boxed{\phi_k(x) = \begin{cases} 1 - k + \dfrac{x}{h} & (k-1)h \le x \le kh \\ 1 + k - \dfrac{x}{h} & kh \le x \le (k+1)h \\ 0 & \text{otherwise} \\ & (\text{i.e. } |x - kh| \ge h) \end{cases} \qquad h \equiv \frac{1}{m+1} \qquad k = 1, 2, \ldots, m}$$

<u>Comments</u>

- This choice of basis functions means that

$$\left.\begin{aligned} \phi_k \phi_\ell &= 0 \\ \frac{d\phi_k}{dx}\frac{d\phi_\ell}{dx} &= 0 \end{aligned}\right\} \text{ unless } |k - \ell| \le 1$$

  which means that the linear system for c_ℓ will be tri-diagonal.

- Depending on $a(x)$, $b(x)$ and $f(x)$ some of the required integrals can be done by hand in advance of any numerical computing. Otherwise some type of quadrature will be necessary to evaluate the integrals.

- $\phi_k \in L^2((0,1))$, $H^1((0,1))$, $H^1_0((0,1))$

  all ϕ_k satisfy $\phi(0) = \phi(1) = 0$