Lecture No. 2

Method of Weighted Residuals

- ***The basic concept of the method of weighted residuals is to drive a residual error to zero through a set of orthogonality conditions.***

$$L\,(u) = p\,(x) \;\epsilon\, \mathrm{V}$$

$$S\,(u) = g\,(x) \;\epsilon\, \Gamma$$

- We let

$$u_{app} = u_B + \sum_{i=1}^{N} \alpha_i \phi_i(x)$$

where

u_B = a function that satisfies all the boundary conditions

α_i = unknown coefficients which we must solve for

ϕ_i = known functions from a complete sequence

- To satisfy admissibility we must satisfy functional continuity requirements, i.e. u_{app} must be sufficiently differentiable, ***in addition*** to satisfying all the boundary conditions:

$$S(u_B) = g(x)$$

$$S(\phi_i) = 0 \quad i = 1, \dots, N$$

  Therefore, the b.c.'s must be satisfied ***independently*** of the parameters α_i.

- The problem with the method of weighted residuals is that it may be difficult to find functions which satisfy the above boundary conditions requirements.

- We now define an interior domain “residual” or error:

$$\mathcal{E}_I = L(u_{app}) - p(x)$$

- In addition we require the interior error to be orthogonal to a set of linearly independent weighting functions

$$< \mathcal{E}_I, w_j > = 0 \quad j = 1, 2, \dots, N$$

  - We note that w_j must be linearly independent functions.
  - If u_{app} satisfies admissibility requirements, ϕ_i come from a complete sequence and w_j are linearly independent, then the method will work.

  - Substituting for $\mathcal{E}_I$

$$< L(u_{app}) - p(x), w_j > = 0 \quad j = 1, 2, \dots, N$$

  - Now substituting for u_{app}

$$< L\left(u_B + \sum_{i=1}^{N} \alpha_i \phi_i(x)\right) - p(x), w_j > = 0 \quad j = 1, 2, \dots, N$$

- Expanding this equation

$$\sum_{i=1}^{N} < L(\phi_i(x)), w_j > \alpha_i \; =< -L(u_B) + p(x), w_j > \qquad j = 1, 2, \ldots, N$$

- Writing this equation in matrix form for N=3

$$\begin{bmatrix} < L(\phi_1(x)), w_1 > & < L(\phi_2(x)), w_1 > & < L(\phi_3(x)), w_1 > \\ < L(\phi_1(x)), w_2 > & < L(\phi_2(x)), w_2 > & < L(\phi_3(x)), w_2 > \\ < L(\phi_1(x)), w_3 > & < L(\phi_2(x)), w_3 > & < L(\phi_3(x)), w_3 > \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$$= \begin{bmatrix} < -L(u_B) + p(x), w_1 > \\ < -L(u_B) + p(x), w_2 > \\ < -L(u_B) + p(x), w_3 > \end{bmatrix}$$

Solution Procedure

- Construction of an approximate solution (follow the rules)
- Reduction to system of algebraic equations through a set of orthogonality conditions (orthogonality of the residual error on the interior of the domain and a set of linearly independent weighting functions)
- Solve a set of simultaneous algebraic conditions

Selection of Weighting Functions

1. Collocation Method (Point Collocation)

- Constrain the error only at a set of selected points.
- We define the approximating function as:

$$u_{app} = u_B + \sum_{k=1}^{N} \alpha_k \, \phi_k$$

$$\Rightarrow$$

$$\varepsilon_I = L\left(u_{app}\right) - p(x)$$

$$\varepsilon_I = L(u_B) - p(x) + \sum_{k=1}^{N} \alpha_k \, L(\phi_k)$$

- The parameters α_k are determined by enforcing the condition $\varepsilon_I = 0$ at N points within the domain. Thus we select as the weighting function the dirac delta function:

$$w_j = \delta\left(x - x_j\right) \qquad j = 1, N$$

- Substituting

$$< \varepsilon_I, w_j > = \int_{x_1}^{x_2} \varepsilon_I(x) w_j(x) dx \qquad j = 1, N$$

$$= \int_{x_1}^{x_2} \varepsilon_I(x) \delta\left(x - x_j\right) dx \qquad j = 1, N$$

$$= \varepsilon\left(x_j\right) = 0 \qquad j = 1, N$$

- Therefore we set the residual error equal to zero at a set of collocation points.

Example of Point Collocation

d.e. $L(u) - p = \frac{d^2u}{dx^2} + u + x = 0$

b.c.'s $u = 0$ at $x = 0$

$u = 0$ at $x = 1$

- The approximate solution is defined as

$$u_{app} = u_B + \sum_{i=1}^{N} \alpha_i \phi_i(x)$$

- However the boundary component is $u_B \equiv 0$ due to the homogeneous specified b.c.'s

- Let's approximate the function as:

$$u_{app} = \alpha_1 x(1-x) + \alpha_2\, x\, x(1-x) + \alpha_3\, x^2\, x(1-x) + \cdots$$

- This function for u_{app} is infinitely differentiable and satisfies the b.c.'s for arbitrary α_i and is therefore admissible. Furthermore the ϕ_i come from a complete sequence.
- Note that:

$$\phi_1 = x\,(1 - x)$$

$$\phi_2 = x\,(1 - x)\,x$$

$$\phi_3 = x\,(1 - x)\,x^2$$

- Let's only use 2 terms in the approximation:

$$u_{app} = \alpha_1\, x\,(1 - x) + \alpha_2\; x^2\,(1 - x)$$

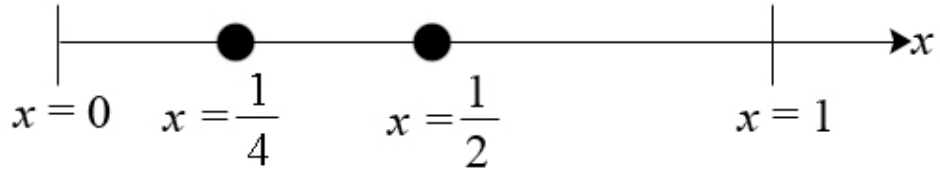
- The error is (substituting u_{app} into the d.e.):

$$\mathcal{E}_I = L(u_{app}) - p$$

$$\mathcal{E}_I = x + (-2 + x - x^2)\,\alpha_1 + (2 - 6x + x^2 - x^3)\,\alpha_2$$

- Now select 2 collocation points (since there are only 2 unknowns):

$$x_1 = \frac{1}{4} \quad \text{and} \quad x_2 = \frac{1}{2}$$



- Enforcing the constraint that the residual equals zero at the collocations points

$$\mathcal{E}_I\,(x_j) = 0 \quad j = 1, 2$$

- Leads to the system of simultaneous equations

$$\begin{bmatrix} \frac{29}{16} & -\frac{35}{64} \\ \frac{7}{4} & \frac{7}{8} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}$$

- Solving the system of algebraic equations

$$\alpha_1 = \frac{6}{31} \text{ and } \alpha_2 = \frac{40}{217}$$

- Thus

$$u_{app} = \frac{x(x-1)}{217}(42 + 40x)$$

Notes on Point Collocation

- The computational effort required in the collocation procedure is minimal.
- The procedure does not produce symmetrical coefficient matrices nor does it produce positive definite matrices. These are both desirable properties. We also note that symmetry has nothing to do with ϕ_i's selected!
- Setting the residual error to zero at discrete points does not mean that you have zero error at those points. The error will only go to zero as you take more and more functions in the approximating sequence, u_{app}. The residual is the difference between the differential operator operating on the approximating functions, which have been truncated, and the function $p(x)$.

2. Least Squares

- The least squares method is implemented by taking the inner product of the error by itself and requiring this quantity to be a minimum. Thus we define u_{app} as:

$$u_{app} = u_B + \sum_{i=1}^{N} \alpha_i \phi_i$$

- And define the residual error as:

$$\mathcal{E}_I = L\,(u_{app}) - p(x)$$

- Now we let:

$$F = < \mathcal{E}_I\,, \mathcal{E}_I > = < L\,(u_{app}) - p, L\,(u_{app}) - p(x) >$$

- We note that any integrated measure of a function must be a quadratic (otherwise +'s and –'s cancel). Therefore a measure of the error is the square of the error.
- Our objective is to minimize F. Driving F to zero, will drive $\mathcal{E}_I$ to zero.

- Since the α_i's are the unknowns and F is a function of α_i, we minimize F with respect to each coefficient α_j

$$\frac{\partial F}{\partial \alpha_j} = 0 \quad j = 1, 2, \ldots N$$

  - Substituting for F

$$\frac{\partial}{\partial \alpha_j} \int \varepsilon_I^2 dx = 0$$

$$\Rightarrow$$

$$\int 2\varepsilon_I \frac{\partial \varepsilon_I}{\partial \alpha_j} dx = 0$$

$$\Rightarrow$$

$$< \varepsilon_I, \frac{\partial \varepsilon_I}{\partial \alpha_j} > = 0$$

- However we recall that:

$$\varepsilon_I = L(u_B) - p + \sum_{i=1}^{N} \alpha_i L(\phi_i)$$

- Assuming a linear operator

$$\frac{\partial \varepsilon_I}{\partial \alpha_j} = L(\phi_j)$$

- This results in:

$$< \varepsilon_I, L(\phi_j) > = 0$$

- Thus the weighting (test) functions are now the trial functions pushed through the differential operator. Therefore:

$$w_j = L(\phi_j)$$

- Substituting for ε_I:

$$< L(u_B) - \ p + \sum_{i=1}^{N} \alpha_i L(\phi_i), L(\phi_j) > = 0 \qquad j = 1, \dots, N$$

$$\Rightarrow$$

$$\sum_{i=1}^{N} \alpha_i < L(\phi_i), L(\phi_j) > = - < L(u_B) - p, L(\phi_j) > , \qquad j = 1, \dots, N$$

- This leads to a set of simultaneous equations:

$$\sum_{i=1}^{N} \alpha_i b_{ji} = - c_j \qquad j = 1, \dots, N$$

where $b_{ji} =< L(\phi_i), L(\phi_j) >$ defines the coefficient matrix $\underline{b}$.

- Thus:

$$\underline{\boldsymbol{b}}\ \underline{\alpha} = -\underline{c}$$

- Again we have developed a system of simultaneous algebraic equations from a differential equation.

Example of the Least Squares Method

d.e.

$$L(u) - p = \frac{d^2u}{dx^2} + u + x = 0$$

$$L(u) = \frac{d^2u}{dx^2} + u$$

$$p(x) = -x$$

b.c.'s

$$u = 0 \text{ at } x = 0 \text{ and } x = 1$$

- Recall from the previous example, we set up an approximating function which was admissible *(satisfied b.c.'s and had sufficient degree of functional continuity)* and from a complete sequence *(in this case we selected a sequence of polynomials)*:

$$u_{app} = u_B + \sum_{i=1}^{2} \alpha_i \phi_i$$

where

$$u_B = 0$$
$$\phi_1 = x(1-x)$$
$$\phi_2 = x^2(1-x)$$

- Thus

$$u_{app} = \alpha_1[x(1-x)] + \alpha_2[x^2(1-x)]$$

- Computing the components of the matrix

$$b_{11} = < L(\phi_1), L(\phi_1) >$$

$$\phi_1 = x - x^2$$

$$\Rightarrow$$

$$L(\phi_1) = \frac{d^2\phi_1}{dx^2} + \phi_1 = -2 + x - x^2$$

$$\Rightarrow$$

$$L(\phi_1)L(\phi_1) = 4 - 4x + 5x^2 - 2x^3 + x^4$$

$$\Rightarrow$$

$$< L(\phi_1), L(\phi_1) > = \int_0^1 (4 - 4x + 5x^2 - 2x^3 + x^4)dx = 3.36667$$

- We proceed to find b_{22}:

$$b_{22} = < L(\phi_2), L(\phi_2) >$$

$$\phi_2 = x^2 - x^3$$

$$L(\phi_2) = \frac{d^2\phi_2}{dx^2} + \phi_2 = 2 - 6x + x^2 - x^3$$

$$L(\phi_2)L(\phi_2) = 4 - 24x + 40x^2 - 16x^3 + 13x^4 - 2x^5 + x^6$$

$$< L(\phi_2), L(\phi_2) > = \int_0^1 (4 - 24x + 40x^2 - 16x^3 + 13x^4 - 2x^5 + x^6)\, dx = 3.7428$$

- Similarly:

$$< L(\phi_1), L(\phi_2) > = < L(\phi_2), L(\phi_1) > = 1.6833$$

- Computing the coefficients of the vector $\underline{c}$, leads to the following system of equations:

$$\begin{bmatrix} 3.36667 & 1.68333 \\ 1.68333 & 3.7428 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.91667 \\ 6.55000 \end{bmatrix}$$

- The final step involves solving this system of equations.

Notes on the Least Squares Method

- The matrix produced is always symmetrical regardless of the operator L or the functions ϕ_2.

$$< L(\phi_i), L(\phi_j) > = < L(\phi_j), L(\phi_i) >$$

$$\Rightarrow$$

$$b_{ji} = b_{ij}$$

- Diagonal entries to the system matrix are always positive. Thus

  $< L(\phi_i), L(\phi_j) > \geq 0$ which leads to a positive definite matrix.

3. Galerkin Method

- The Galerkin method consists of taking the inner product of the error and trial functions themselves. Thus:

$$w_j = \phi_j$$

$$\Rightarrow$$

$$\langle \mathcal{E}_I, \phi_j \rangle = 0 \qquad j = 1, N$$

- Hence

$$< (L(u_B) - p), \phi_j > + \sum_{i=1}^{N} \alpha_i < L(\phi_i), \phi_j > = 0 \qquad j = 1, N$$

$$\Rightarrow$$

$$\sum_{i=1}^{N} b_{ij} \alpha_i = c_j \qquad j = 1, N$$

where

$$b_{ij} = < L(\phi_i), \phi_j >$$

- The Galerkin method is computationally simpler than the least squares method. We no longer push ϕ_j through the operator L.
- We note that the matrix produced is still symmetrical if the operator L is self adjoint. The operator L is self adjoint if in the transformation

  $< L(u), v > = < u,\ L^*(v) > + \ldots$

  we have $L = L^*$.

  Self adjointness of an operator is analogous to symmetry of a matrix.
- We note that the matrix is still positive definite if the operator L is positive definite (*and self adjoint)*. For a positive definite operator $L(u)$ we have $< L(u), u > > 0$ for any u. We will discuss the above matters in more detail later.

Example of the Galerkin Method

d.e $L(u) - p = \frac{d^2u}{dx^2} + u + x = 0$

b.c. $u = 0$ at $x = 0$ and $x = 1$

- The approximating function:

$$u_{app} = \alpha_1[x(1-x)] + \alpha_2[x^2(1-x)]$$

and

$$u_B = 0 \quad \phi_1 = x - x^2, \phi_2 = x^2 - x^3$$

- This allows us to compute the components of the system matrix b_{ij}

$$b_{11} = < L(\phi_1), \phi_1 > = \int_0^1 (-2 + x - x^2)(x - x^2)dx$$

$$b_{22} = < L(\phi_2), \phi_2 > = \int_0^1 (2 - 6x + x^2 - x^3)(x^2 - x^3)dx$$

$$b_{12} = < L(\phi_2), \phi_1 > = \int_0^1 (2 - 6x + x^2 - x^3)(x - x^2)dx$$

$$b_{21} = < L(\phi_1), \phi_2 > = \int_0^1 (-2 + x - x^2)(x^2 - x^3)dx$$

- Also computing the components c_j :

$$\begin{bmatrix} \frac{3}{10} & \frac{3}{20} \\ \frac{3}{20} & \frac{13}{105} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{12} \\ \frac{1}{20} \end{bmatrix}$$

- We now solve this system of simultaneous equations.
- ***Thus for the Galerkin's method the test (weighting) functions are the same as the trial functions!***

Alternative notation for the Galerkin method

$$w_j = \phi_j$$

$$< \mathcal{E}_I, \phi_j > = 0 \qquad j = 1, 2, \ldots, N$$

$$\Rightarrow$$

$$\int_V (L(u) - p)\phi_j dV = 0 \qquad j = 1, 2, \ldots, N$$

- Now let's define:

$$\delta u = \delta\alpha_1\phi_1 + \delta\alpha_2\phi_2 + \cdots + \delta\alpha_N\phi_N$$

- This is a *notational* change (often used with the ***Galerkin*** method). $\delta\alpha_1, \delta\alpha_2 \ldots \delta\alpha_N$ are arbitrary coefficients. Since ϕ_j are linearly independent functions, the error statement may now be written as:

$$\int_V (L(u) - p)\delta u dV = 0$$

for arbitrary $\delta\alpha_j$

- The previous expression produces N equations since ϕ_j are linearly independent functions. Since the coefficients $\delta\alpha_j$ are arbitrary, we can select them such that:

$$\delta u_1 = 1, \delta u_2 = 0, \delta u_3 = 0$$
$$\delta u_1 = 0, \delta u_2 = 1, \delta u_3 = 0$$

- This leads us directly back to our first statement:

$$< \mathcal{E}_I, \phi_j > = 0 \qquad j = 1, N$$

4. Subdomain Method

- Divide the domain V and N smaller domains and select the weighting functions as:

$$w_i = \begin{cases} 1 & x \ in \ V_i \\ 0 & x \ not \ in \ V_i \end{cases}$$

Thus this method integrates the residual error over each subdomain and sets it equal to zero.

5. Method of Moments

Select

$$w_i = x^{i-1}, i = 1, \dots, N$$

- Thus the method applies the series $1, x, x^2, x^3, \dots, x^{N-1}$ as weighting functions. Thus we compute higher order moments of the residual and force them to zero.

6. Least Squares Collocation

- For conventional Collocation, the number of collocation points equals the number of unknown α_i's.
- We extend the method to treat more collocation points than unknowns. Therefore the error is evaluated $M > N$ points:

$$\mathcal{E}_I = L(u_{app}) - p(x) \text{ at } M > N \text{ points}$$

- Now sum $\mathcal{E}_I^2$ at these M different points:

$$F = < \{L(u) - p\}^2, \delta(x - x_m) > \qquad m = 1, \ldots, M$$