Contents

1 The Finite Element Method

1.1 Two-point boundary value problem

Definition. ***(problem setup)*** *Linear two-point boundary value problem*

$$-\frac{d}{dx}\left[a(x)\frac{du}{dx}\right] + b(x)u = f \qquad 0 \le x \le 1$$

where a,b, and f are given functions, a is differentiable and $a(x) > 0$, $b \ge 0$, $0 < x < 1$. Assume dirichlet boundary condition

$$u(0) = \alpha \qquad u(1) = \beta$$

Definition. ***(4 principles of FEM)***

1. ***(Approximate the solution in a finite-dimensional space)****, i.e. approximate u by a linear combination of functions in $\mathring{\mathbb{H}}_m = span\{\varphi_1, \cdots, \varphi_m\}$ where $\varphi_0, \cdots, \varphi_m$ are linearly independent functions satisfying zero boundary condition. So we have an approximate solution u_m given by*

$$u_m(x) = \varphi_0(x) + \sum_{l=1}^{m} \gamma_l \varphi_l(x) \qquad 0 \le x \le 1$$

2. ***(Choose the approximation so the defect is orthogonal to the space $\mathring{\mathbb{H}}_m$)*** *Consider defect as the error of u_m as approximation to u,*

$$d_m(x) = -\frac{d}{dx}\left[a(x)\frac{du_m(x)}{dx}\right] + b(x)u_m(x) - f(x) \qquad 0 < x < 1$$

Alternatively, we seek u_m such that d_m is orthogonal to all basis elements of $\mathring{\mathbb{H}}_m$, in other words, satisfy the ***Galerkin equations***

$$\langle d_m, \varphi_k \rangle \qquad k = 1, 2, \cdots, m$$

3. ***(Integrate by parts to depress the maximal extent possible the differentiability requirements of the space $\mathring{\mathbb{H}}_m$)*** *We can expand Gerlerkin equations by substituting u_m as a linear combination of basis to obtain a linear system of m equations in m unknown $\gamma_1, \cdots, \gamma_m$*

$$\sum_{l=1}^{m} \gamma_l \left[\left\langle -(a\varphi_l')', \varphi_k \right\rangle + \langle b\varphi_l, \varphi_k \rangle\right] = \langle f, \varphi_k \rangle - \left[\left\langle -(a\varphi_0')', \varphi_k \right\rangle + \langle b\varphi_0, \varphi_k \rangle\right]$$

for $k = 1, 2, \cdots, m$. Given standard Euclidean inner product over funtion spaces

$$\langle v, w \rangle = \int_0^1 v(\tau)w(\tau)d\tau$$

We compute integration by parts to reduce differentiability requirements for φs

$$\begin{aligned}\left\langle -(a\varphi_l')', \varphi_k \right\rangle &= \int_0^1 \left\{-(a(\tau)\varphi_l(\tau)')'\right\} \varphi_k(\tau)d\tau \\ &= \left[-a(\tau)\varphi_l'(\tau)\varphi_k(\tau)\right]_0^1 + \int_0^1 a(\tau)\varphi_l'(\tau)\varphi_k'(\tau)d\tau \\ &= \int_0^1 a(\tau)\varphi_l'(\tau)\varphi_k'(\tau)d\tau = \left\langle a\varphi_l', \varphi_k' \right\rangle \qquad l =, 0, 1, \cdots, m\end{aligned}$$

The linear system becomes

$$\sum_{l=1}^{m} a_{k,l}\gamma_l = \int_0^1 f(\tau)\varphi_k(\tau)d\tau - a_{k,0} \qquad k = 1, 2, \cdots, m$$

where

$$a_{k,l} = \int_0^1 \left[a(\tau)\varphi_l'(\tau)\varphi_k'(\tau) + b(\tau)\varphi_l(\tau)\varphi_k(\tau)\right] d\tau$$

With the help of integration by parts, and knowledge that value of integral is independent of values that integrand assumes on a finite set (at knots), we can choose basis functions that are piecewise once-differentiable.

4. ***(Choose each function*** φ_k ***so that it vanishes along most of*** $(0,1)$***, thereby ensuring that*** $\varphi_k\varphi_l = 0$ ***for most choices of*** $k, l = 1, 2, \cdots, m$***)*** *Ideally, we want to choose* $\mathring{\mathbb{H}}_m$ *to be*

   (a) *Dense in infintie-dimensional space inhabited by the weak solution, i.e.* $\|u_m - u\|$ *is as small as possible. (spectral method) and*

   (b) *Choose* $\varphi_1, \cdots, \varphi_m$ *such that evaluating* $\mathcal{O}(m^2)$ *integrals to solve the linear system is as fast as possible.*

   To satisfy (2), we want each function φ_k *to be supported on a relatively small set* $\mathbb{E}_k \subset (0,1)$ *and* $\mathbb{E}_k \cap \mathbb{E}_l = \emptyset$ *for as many* k, l *as possible. One example is the chapeau (hat) function*

$$\varphi_k(x) = \begin{cases} 1 - k + \frac{x}{h} & (k-1)h \leq x \leq kh \\ 1 + k - \frac{x}{h} & kh \leq x \leq (k+1)h \\ 0 & |x - kh| \geq h \end{cases}$$

   which reduces the number of integral evaluation from $\mathcal{O}(m^2)$ *to* $\mathcal{O}(m)$

Proposition. ***(A is positive definite)***

Proof. Idea is to show A is a gram matrix (i.e. $A_{k,l} = \langle \varphi_k, \varphi_l \rangle$ where $\langle \cdot, \cdot \rangle$ is an inner product in $\mathring{\mathbb{H}}_m$). This is easy by verifying axioms of inner product. We claim that any gram matrix is symmetric positive definite. Let $\gamma \in \mathbb{R}^m$ where $\gamma \neq 0$, then

$$\langle A\gamma, \gamma \rangle = \sum_k \gamma_k (A\gamma)_k = \sum_k \gamma_k \sum_l \gamma_l \langle \varphi_k, \varphi_l \rangle = \left\langle \sum_k \gamma_k \varphi_k, \sum_l \gamma_l \varphi_l \right\rangle = \langle u, u \rangle > 0$$

$u = \sum_k \gamma_k \varphi_k \in \mathring{\mathbb{H}}_m \neq 0$, so inequality follows from positive definiteness of inner product over $\mathring{\mathbb{H}}_m$ □

Definition. ***(Weak Solutions)*** *Consider* $\mathcal{L}\boldsymbol{u} = \boldsymbol{f}$ *where* $\mathcal{L}$ *is a differential operator. The classical solution is given by function* $\boldsymbol{u}$ *that satisfies the equation exactly. In constrast, we can consider* $\boldsymbol{v} \in \mathring{\mathbb{H}}$*, an infinite-dimensional linear space, as a* ***weak solution*** *if the Galerkin equations are satisfied, i.e.*

$$\langle \mathcal{L}\boldsymbol{v} - \boldsymbol{f}, \mathbf{w} \rangle = 0 \qquad \forall\, \mathbf{w} \in \mathring{\mathbb{H}}$$

Weak and exact solution are same for ODEs, since Lipschitz condition ensures existence of unique solution. But this is not the case for boundary value problems.

Definition. ***(Variational formulation of two point BVP)*** *Given a functional* $\mathcal{J} : \mathbb{H} \to \mathbb{R}$*, where* $\mathbb{H}$ *is some functional space. Wish to find a function* $u \in \mathbb{H}$ *such that*

$$\mathcal{J}(u) = \min_{v \in \mathbb{H}} \mathcal{J}(v)$$

Consider $a, b,$ *and* f *given on* $(0,1)$ *where* $a(x) > 1$ *and* $b(x) \geq 1$ *and any* $v \in \mathbb{H}$ *satisfies*

$$\int_0^1 v^2(\tau) d\tau, \int_0^1 \left[v'(\tau)\right]^2 d\tau < \infty$$

We let

$$\mathcal{J}(v) = \int_0^1 \left\{a(\tau)\left[v'(\tau)\right]^2 + b(\tau)\left[v(\tau)\right]^2 - 2f(\tau)v(\tau)\right\} d\tau \qquad v \in \mathbb{H}$$

We can show that $u \in H$ *which minimizes* J *if and only if* u *is the weak solution of the differential equation given at start of the section. We call the two value boundary value problem the Euler-Lagrange equation of* $\mathcal{J}$

Definition. ***(Ritz Method)*** *Let $\mathring{\mathbb{H}} = span(\varphi_1, \cdots, \varphi_m)$ and choose arbitrary $\varphi_0 \in \mathbb{H}$. Therefore $\mathring{\mathbb{H}}$ is a m-dimensional linear space. We seek a minimum of $\mathcal{J}$ in the affine space $\varphi_0 + \mathring{\mathbb{H}}$. In other word, we seem vector $\boldsymbol{\gamma}$ that minimizes*

$$\mathcal{J}_m(\boldsymbol{\delta}) = \mathcal{J}\left(\varphi_0 + \sum_{l=1}^{m} \delta_l \varphi_l\right) \qquad \boldsymbol{\delta} \in \mathbb{R}^m$$

To find the minimum of sJ, We let gradient of $\mathcal{J}_m$ be zero,

$$\frac{1}{2}\frac{\partial \mathcal{J}(\boldsymbol{\delta})}{\delta_k} = \sum_{l=1}^{m} \int_0^1 \left(a\varphi_l'\varphi_k' + b\varphi_l\varphi_k\right) d\tau + \int_0^1 \left(a\varphi_0'\varphi_k' + b\varphi_0\varphi_k\right) d\tau - \int_0^1 f\varphi_k d\tau$$

Note letting $\partial\mathcal{J}_m/\partial\delta_k$ for $k = 1, 2, \cdots, m$ recovers the Galerkin equations! We can also show that the hessian matrix $(\partial^2\mathcal{J}_m/\partial\partial\delta_k\partial\delta_j)_{k,j=1}^m$ be nonnegative definite. Although we get the same system of equations, we are able to ignore natural boundary conditions.

Definition. ***(Ritz-Galerkin method)***

1. *Approximate solution in finite dimensional space $\varphi_0 + \mathring{\mathbb{H}}_m \subset \mathbb{H}$*
2. *Retain only essential boundary conditions*
3. *Choose the approximate so that the defect is orthogonal to $\mathring{\mathbb{H}}_m$, or alternatively, so that a variational problem is minimized in $\mathring{\mathbb{H}}_m$*
4. *Integrate by parts to depress the maximal extent possible the differentiability requirements of the space $\mathring{\mathbb{H}}_m$*
5. *Choose each function in a basis of $\mathring{\mathbb{H}}_m$ in such a way that it vanishes along much of the spatial domain of interest, thereby ensuring that the intersectino between the supports of most of the basis function is empty*

1.2 Synopsis of FEM theory

Definition. ***(problem setup)*** *Given boundary value problem*

$$\mathcal{L}u = f \qquad \boldsymbol{x} \in \Omega$$

where $u = u(\boldsymbol{x})$ and $f = f(\boldsymbol{x})$ is bounded and $\Omega \subset \mathbb{R}^d$ is an open, bounded, connected set with sufficiently smooth boundary. $\mathcal{L}$ is a linear differentiable operator

$$\mathcal{L} = \sum_{|\alpha|<2\nu} c_\alpha(\boldsymbol{x}) D^\alpha$$

Definition. ***(Properties of linear differential operator)*** *Given a linear differentiable linear operator $\mathcal{L}$ and a bilinear form $\tilde{a}(\cdot,\cdot)$ such that $\tilde{a}(v, w) = \langle \mathcal{L}v, w\rangle$ for sufficiently smooth v and w. Then $\mathcal{L}$ is*

1. ***(self-adjoint)*** *if $\tilde{a}(v, w) = \tilde{a}(w, v)$ for all $v, w \in \mathring{\mathbb{H}}$*
2. ***(elliptic)*** *if $\tilde{a}(v, v) > 0$ for all $v \in \mathring{\mathbb{H}} \neq 0$*
3. ***(positive definite)*** *if it is self-adjoint and elliptic*

As an example, the following is a positive definite operator (genearlization of 2-point BVP in 9.1 and negative of Laplace operator $-\nabla^2$)

$$\mathcal{L} = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i} \sum_{j=1}^{d} b_{i,j}(\boldsymbol{x}) \frac{\partial}{\partial x_j}$$

where $(b_{i,j}(\boldsymbol{x}))$ is symmetric positive definite for all $\boldsymbol{x} \in \Omega$

Theorem. ***(Variational formulation of linear differentiable operator)*** *Provided that the operator $\mathcal{L}$ is positive definite. $\mathcal{L}u = f$ is the Euler-Lagrange equation of the variational problem*

$$\mathcal{J}(v) = \tilde{a}(v, v) - 2\langle f, v\rangle \qquad v \in \mathbb{H}$$

*The weak solution of $\mathcal{L}u = f$ is therefore, the **unique** minimum of $\mathcal{J}$ in $\mathbb{H}$ (up to Lebesque zero)*

Definition. ***(Methods for solving $\mathcal{L}u = f$)*** *Choose $\varphi_0 \in \mathbb{H}$ and let $\varphi_1, \cdots, \varphi_m$ span $\mathring{\mathbb{H}}_m$.*

1. ***(Ritz method for $\mathcal{L}u = f$)*** *Seek $\boldsymbol{\gamma} \in \mathbb{R}^m$ that will minimize*

$$\mathcal{J}_m(\boldsymbol{\delta}) = \mathcal{J}(\varphi_0 + \sum_{l=1}^{m} \delta_l \varphi_l) \qquad \boldsymbol{\delta} \in \mathbb{R}^m$$

   Setting $\partial \mathcal{J}_m / \partial \delta_l$ to zero for $l = 1, \cdots, m$ results in a linear system,

$$a_{k,l} = \tilde{a}(\varphi_k, \varphi_l)$$

2. ***(Galerkin method)*** *Seek $\boldsymbol{\gamma} \in \mathbb{R}^m$ that will satisfy*

$$\tilde{a}\left(\varphi_0 + \sum_{l=1}^{m} \gamma_l \varphi_l, \varphi_k\right) - \langle f, \varphi_k\rangle = 0 \qquad k = 1, 2, \cdots, m$$

Definition. ***(Special case of Sobolev norm)*** *Given any $v \in \mathbb{H}$, let*

$$\|v\|_H = \left(\|v\|^2 + [\tilde{a}(v, v)]^{1/2}\right)$$

be a norm in $\mathbb{H}$. $\tilde{a}$ is

1. ***(bounded)*** *if there exists $\delta > 0$ such that $|\tilde{a}(v, w)| \leq \delta \|v\|_H \|w\|_H$ for all $v, w \in H$*
2. ***(coercive)*** *if there exists $\kappa > 0$ such that $\tilde{a}(v, v) \geq \kappa \|v\|_H^2$ for every $v \in \mathbb{H}$*

Theorem. ***(Lax-Milgram theorem)*** *Let $\mathcal{L}$ be linear, bounded, and coercive and let $\mathbb{V}$ be a closed linear subspace of $\mathring{\mathbb{H}}$. There exists a unique $\tilde{u} \in \varphi_0 + \mathbb{V}$ such that*

$$\tilde{a}(\tilde{u}, v) - \langle f, v\rangle = 0 \qquad v \in \mathbb{V}$$

and the error is bounded

$$\|\tilde{u} - u\|_H \leq \frac{\delta}{\kappa} \inf\{\|v - u\|_H \mid v \in \varphi_0 + \mathbb{V}\}$$

where $\varphi_0 \in \mathbb{H}$ is arbitrary and u is a weak solution of $\mathcal{L}u = f$ in $\mathbb{H}$

1. *(observation) constant δ/κ is independent of choice of finite dimensional $\mathring{\mathbb{H}}$ or of the norm $\|\cdot\|$*
2. *(observation) although the infimum on right hand side is unknown (since u is not known), the distance of u to $\varphi_0 + \mathring{\mathbb{H}}_m$ can be bounded in terms of distance of an arbitrary member $w \in \varphi_0 + \mathring{\mathbb{H}}$ from $\varphi_0 + \mathring{\mathbb{H}}_m$*
3. *(remark) error estimation in Galerkin method can be replaced by an appoximation problem: given a function $w \in \mathring{\mathbb{H}}$ find the distance $\inf_{v \in \mathring{\mathbb{H}}_m} \|w - v\|_H$*

2 Spectral Method

3 Gaussian elimination for sparse linear equations

4 Classical iterative methods for sparse linear equations

5 Multigrid techniques

6 Conjugate Gradient