# Finite Element Methods in One Dimension: Part 2.

$$\mathcal{L}u(x) \equiv -\frac{d}{dx}\left(p(x)\frac{du}{dx}\right) + q(x)u(x) = f(x) \tag{1}$$

$$x(0) = 0, \quad x(1) = 0 \tag{2}$$

As before, divide the interval $[0,1]$ into subintervals and approximate $u(x)$ by a *continuous piecewise linear* function:

$$\tilde{u}(x) = \sum_{j=1}^{n-1} c_j \varphi_j(x), \tag{3}$$

where $\varphi_1(x), \ldots \varphi_{n-1}(x)$ are the hat functions that form a basis for the set of continuous piecewise linear functions with value 0 at the endpoints.

Previously, we determined the coefficients $c_1, \ldots, c_{n-1}$ via the *Galerkin* conditions:

$$\langle \mathcal{L}\tilde{u} - f, \varphi_i \rangle = 0, \quad i = 1, \ldots, n-1; \tag{4}$$

that is, the residual was forced to be orthogonal to each basis function.

One could choose the coefficients in other ways. For example, one might choose $c_1, \ldots, c_{n-1}$ to minimize the $L^2$-norm of the residual: $\langle \mathcal{L}\tilde{u} - f, \mathcal{L}\tilde{u} - f \rangle^{1/2}$. This leads to a *least squares approximation.*

Another idea, when the operator $\mathcal{L}$ is *self-adjoint* (i.e., $\langle \mathcal{L}v, w \rangle = \langle v, \mathcal{L}w \rangle$ for all $v$ and $w$) and *positive definite* (i.e., $\langle \mathcal{L}v, v \rangle > 0$ for all $v \neq 0$), is to minimize the *energy norm* of the error:

$$\langle \mathcal{L}\tilde{u} - f, \tilde{u} - \mathcal{L}^{-1}f \rangle = \langle \mathcal{L}\tilde{u}, \tilde{u} \rangle - 2\langle f, \tilde{u} \rangle + \text{ constant } . \tag{5}$$

Note that while we cannot compute $\mathcal{L}^{-1}f$ in (5) (since this is the solution that we are seeking), we can compute the expression on the right-hand side. For problem (1) this is

$$\int_0^1 \left( -\frac{d}{dx}\left(p(x)\tilde{u}'(x)\right) + q(x)\tilde{u}(x) \right) \tilde{u}(x)\, dx - 2\int_0^1 f(x)\tilde{u}(x)\, dx.$$

After integrating by parts and using the fact that $\tilde{u}(0) = \tilde{u}(1) = 0$, this becomes

$$\int_0^1 \left( p(x)\left(\tilde{u}'(x)\right)^2 + q(x)\left(\tilde{u}(x)\right)^2 \right) dx - 2\int_0^1 f(x)\tilde{u}(x)\, dx. \tag{6}$$

By choosing $c_1, \ldots, c_{n-1}$ to minimize the expression in (6), we obtain the *Ritz* approximation.

**Claim:** If $\mathcal{L}$ is self-adjoint and positive definite, then the Galerkin approximation is the same as the Ritz approximation.

*Proof:* Suppose $\tilde{u}$ minimizes

$$\mathcal{I}(v) \equiv \langle \mathcal{L}v - f, v - \mathcal{L}^{-1} f \rangle$$

over all functions $v$ in the trial space (i.e., linear combinations of $\varphi_1, \ldots, \varphi_{n-1}$). For any function $v$ in this space and any number $\epsilon$, we have

$$\begin{aligned} \mathcal{I}(\tilde{u} + \epsilon v) &= \langle \mathcal{L}\tilde{u} - f + \epsilon \mathcal{L}v, \tilde{u} - \mathcal{L}^{-1} f + \epsilon v \rangle \\ &= \mathcal{I}(\tilde{u}) + 2\epsilon \langle \mathcal{L}\tilde{u} - f, v \rangle + \epsilon^2 \langle \mathcal{L}v, v \rangle. \end{aligned} \tag{7}$$

The only way that this can be greater than or equal to $\mathcal{I}(\tilde{u})$ for all $\epsilon$ is for the coefficient of $\epsilon$ to be zero; i.e., $\langle \mathcal{L}\tilde{u} - f, v \rangle = 0$ for all $v$ in the trial space.

Conversely, if $\langle \mathcal{L}\tilde{u} - f, v \rangle = 0$ for all $v$ in the trial space, then expression (7) is always greater than or equal to $\mathcal{I}(\tilde{u})$, so $\tilde{u}$ minimizes $\mathcal{I}(v)$. □

This result is important because it means that of all functions in the trial space, the finite element approximation is *best* as far as minimizing the energy norm of the error. Now one can derive results about the global error (in energy norm) just by determining how well an arbitrary function $u(x)$ can be approximated by a linear combination of $\varphi_1(x), \ldots, \varphi_{n-1}(x)$; e.g., by a continuous piecewise linear function. Other trial spaces are possible too, such as continuous piecewise quadratics or Hermite cubics.

The following theorem about the piecewise polynomial *interpolant* of a function can be used to bound the error in the finite element approximation, since, in energy norm, the finite element approximation is a better approximation to $u$ than its piecewise polynomial interpolant. It also can be shown that in other norms the order of accuracy of the finite element approximation is the same as that of the piecewise polynomial interpolant.

**Theorem.** Let $S$ contain all continuous piecewise polynomials of degree $k-1$ or less, and let $u_I$ be the interpolant of $u$ in $S$. Then

$$\|u - u_I\|_0 \leq C_0 \; h^k \; \|u^{(k)}\|_0,$$

where $h = \max_i (x_{i+1} - x_i)$ and $\|v\|_0 = [\int_0^1 (v(x))^2 \; dx]^{1/2}$.

For piecewise linear functions, $k = 2$, and the theorem tells us that the error in the piecewise linear interpolant is of order $O(h^2)$. This is the order of the error in the piecewise linear finite element approximation as well.