

The results in these two examples are of roughly the same quality, showing that the Galerkin method can be as effective as the Fourier method. The most significant difference in these two methods is the need to form the matrix \mathbf{K} in the second example and solve the $N \times N$ system $\mathbf{K}\mathbf{u} = \mathbf{f}$. This is very time-consuming compared to the computations required in the first example (where the system of linear equations is diagonal). This question of efficiency is particularly important when we have two or three spatial dimensions; in a problem of realistic size, it may take impossibly long to solve the resulting linear system if the coefficient matrix is *dense*. A dense matrix is a matrix in which most or all of the entries are nonzero. A *sparse matrix*, on the other hand, has mostly zero entries.

The finite element method is simply the Galerkin method with a special choice for the subspace and its basis; the basis leads to a sparse coefficient matrix. The ultimate sparse, nonsingular matrix is a diagonal matrix. Obtaining a diagonal matrix requires that the basis for the approximating subspace be chosen to be orthogonal with respect to the energy inner product. As mentioned earlier, it is too difficult, for a problem with variable coefficients, to find an orthogonal basis. The finite element method uses a basis in which *most* pairs of functions are orthogonal; the resulting matrix is not diagonal, but it is quite sparse.

Exercises

1. Determine whether the bilinear form

$$a(u, v) = \int_0^\ell \frac{du}{dx}(x) \frac{dv}{dx}(x) dx$$

defines an inner product on each of the following subspaces of $C^2[0, \ell]$. If it does not, show why.

- (a) $\{v \in C^2[0, \ell] : v(\ell) = 0\}$
 - (b) $\{v \in C^2[0, \ell] : v(0) = v(\ell)\}$
 - (c) $C^2[0, \ell]$ (the entire space)
2. Show that if F_N is the subspace defined in Example 5.18, and the Galerkin method is applied to the weak form of

$$\begin{aligned} -k \frac{d^2 u}{dx^2} &= f(x), \quad 0 < x < \ell, \\ u(0) &= 0, \\ u(\ell) &= 0, \end{aligned}$$

with F_N as the approximating subspace, the result will always be the partial Fourier sine series (with N terms) of the exact solution u .

3. Define S to be the set of all polynomials of the form $ax + bx^2$, considered as functions defined on the interval $[0, 1]$.
 - (a) Explain why S is a subspace of $C^2[0, 1]$.

(b) Explain why the bilinear form

$$a(u, v) = \int_0^1 \frac{du}{dx}(x) \frac{dv}{dx}(x) dx$$

defines an inner product on S .

(c) Compute the best approximation from S , in the energy norm, to $f(x) = e^x$.

(d) Explain why the best approximation from P_2 , in the energy norm, to $f(x) = e^x$ is not unique. (The subspace P_2 is defined in Exercise 3.1.3d.)

4. Repeat Exercise 3 with

$$S = \{v \in C^2[0, 1] : v(x) = ax + bx^2 + cx^3 \text{ for some } a, b, c \in \mathbf{R}\}.$$

5. Define

$$V_2 = \text{span} \left\{ x(1-x), x \left(\frac{1}{2} - x \right) (1-x) \right\}$$

and regard V_2 as a subspace of $C^2_D[0, 1]$. Apply the Galerkin method, using V_2 as the approximating subspace, to estimate the solution of

$$-\frac{d}{dx} \left((1+x) \frac{du}{dx} \right) = x, \quad 0 < x < 1,$$

$$u(0) = 0,$$

$$u(1) = 0.$$

Find the exact solution and graph the exact and approximate solutions together.

6. Repeat Exercise 5 using the subspace

$$V_3 = \text{span} \left\{ x(1-x), x \left(\frac{1}{2} - x \right) (1-x), x \left(\frac{1}{3} - x \right) \left(\frac{2}{3} - x \right) (1-x) \right\}.$$

7. The standard Gaussian elimination algorithm for solving $Ax = b$ requires

$$O\left(\frac{2}{3}n^3\right)$$

arithmetic operations when $A \in \mathbf{R}^{n \times n}$. (The exact number is a cubic polynomial in n , and $2n^3/3$ is the leading term. When n is large, the lower degree terms are negligible in comparison to the leading term.)

(a) Suppose that, on a certain computer, it takes 1 second to solve a 100×100 dense linear system by Gaussian elimination. How long will it take to solve a 1000×1000 system? A 10000×10000 system?