1 Finite element method

For the simple, one dimensional case, the exact solution to the Helmholtz problem is know. Finding the exact solution for more complex and higher dimensional problems turns out to be hard and impractical. Often, an approximation to the exact solution suffices for engineering purposes. A widly used method to find such approximations is know as the *Galerkin finite element method*.

Using this method, an approximation to the real solution is found by transforming the problem in a system linear equations. This results in a sparse linear system for which many solving techniques have been studied []. Although the linear systems resulting from this method are sparse, for an accurate aproximation of solution to higher dimensional or heavily oscillating problems a very large sytem can be needed.

1.1 One dimensional Helmholtz problem

As an example, we will consider a simple one dimensional wave problem. Suppose we have the following conditions for $u(x)$ on $[0, 1]$:

$$-\frac{d^2u}{dx^2} - k^2u = f, \quad \text{on } \Omega = (0, 1), \tag{1}$$

$$u(0) = 0, \tag{2}$$

$$\frac{du}{dx}(1) - iku(1) = 0. \tag{3}$$

It can be shown that the exact solution to is given by

$$u(x) = \frac{e^{ikx}}{k}\int_0^x \sin(ks)f(s)ds + \frac{\sin(kx)}{k}\int_x^1 e^{iks}f(s)ds, \tag{4}$$

which is periodic with wavelenth $\lambda = \frac{2\pi}{k}$.

While this problem has a simple solution, exact solutions to more complex and higher order problems are difficult, nigh impossible to find in a simmilar matter. To find a good approximation, we have to restrict the problem to one that can be solved with a computer.

1.2 Galerking finite element method

Since the original problem is infinite dimensional, we simplify the problem by restricting the approximation of the real solution to a finite dimensional search space. Suppose we have a bounded domain $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$. We define the function space $L_2(\Omega)$ of square integrable function on Ω by saying that $f \in L_2(\Omega)$ if

$$\|f\| := \left(\int_\Omega |f(x)|^2 d\Omega\right)^{\frac{1}{2}} < \infty. \tag{5}$$

Furthermore, we say $f \in H^1_{(0}(\Omega)$ if $f(0) = 0$ and

$$\|\nabla f\|^2 + \|f\|^2 < \infty. \tag{6}$$

For solving the above system with a finite element method, we will first rewrite the problem in its weak form. We multiply both sides with a test function $v \in H^1_{(0}(\Omega)$ and integrate afterwards to obtain

$$-\int_0^1 u''(x)v(x) - k^2u(x)v(x)dx = \int_0^1 f(x)v(x)dx. \tag{7}$$

By substituting boundary condition (3) and taking the fact that $v(0) = (0)$ into consideration, we obtain

$$\int_0^1 u'(x)v'(x)dx - k^2\int_0^1 u(x)v(x)dx - iku(1)v(1) = \int_0^1 f(x)v(x)dx. \tag{8}$$

Instead of solving the above equation for the exact solution $u(x)$, we want to approximate it by finding $U(x) \approx u(x)$. To do this, we define a finite element mesh X_h on Ω, by

$$X_h := \{x_i; x_i = ih, i = 0, 1, \ldots, N\}, \tag{9}$$

where $h = 1/N$. We limit our search space to $S_h(0,1) \subset H^1_{(0}(\Omega)$, the space of piecewise continuous linear functions with nodal values at the points in X_h, satisfying (2). This function space is spanned by the set of hat functions defined as

$$\chi_j(x) = \begin{cases} \frac{1}{h}(x - x_{j-1}), & x \in [x_{j-1}, x_j], \\ \frac{1}{h}(x_{j+1} - x), & x \in [x_j, x_{j+1}], \\ 0 & \text{elsewhere,} \end{cases} \tag{10}$$

for $j = 1, 2, \ldots, N-1$ and for $j = N$ by

$$\chi_N(x) = \begin{cases} \frac{1}{h}(x - x_{j-1}), & x \in [x_{j-1}, 1], \\ 0 & \text{elsewhere.} \end{cases} \tag{11}$$

Now, if we require that both U and v are in $S_h(0,1)$, then we can write

$$U(x) = \sum_{j=1}^{N} u_j \chi_j(x), \tag{12}$$

and (8) transforms to

$$\sum_{j=1}^{N} \left[\int_0^1 \chi_j'(x)\chi_m'(x)dx - k^2 \int_0^1 \chi_j(x)\chi_m(x)dx \right] u_j - iku_N\chi_m(1) = \int_0^1 f(x)\chi_m(x)dx, \tag{13}$$

for $m = 1, 2, \ldots, N$.

1.3 Linear system

Since each χ_m is known, we can easily compute the integrals in (13). If we substitute these integrals back into the equation, the result is a system of linear equations with unknowns u_i, $i = 1, 2, \ldots, N$. The linear system is given by

$$(A - k^2 B - ikC)\mathbf{u} = \mathbf{f}. \tag{14}$$

where the matrices are of the form

$$A = \begin{pmatrix} \frac{2}{h} & -\frac{1}{h} & 0 & & \\ -\frac{1}{h} & \frac{2}{h} & -\frac{1}{h} & \ddots & \\ 0 & -\frac{1}{h} & \frac{2}{h} & \ddots & 0 \\ & \ddots & \ddots & \ddots & -\frac{1}{h} \\ & & 0 & -\frac{1}{h} & \frac{1}{h} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{2h}{3} & \frac{h}{6} & 0 & & \\ \frac{h}{6} & \frac{2h}{3} & \frac{h}{6} & \ddots & \\ 0 & \frac{h}{6} & \frac{2h}{3} & \ddots & 0 \\ & \ddots & \ddots & \ddots & \frac{h}{6} \\ & & 0 & \frac{h}{6} & \frac{h}{3} \end{pmatrix},$$

$$C = \begin{pmatrix} 0 & 0 & & \\ 0 & 0 & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & 1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \int_0^1 f(x)\chi_1(x)dx \\ \int_0^1 f(x)\chi_2(x)dx \\ \vdots \\ \int_0^1 f(x)\chi_N(x)dx \end{pmatrix}.$$

Here, the elements of matrices A and B are determined by

$$A_{ij} = \int_0^1 \chi_i'(x)\chi_j'(x)dx, \quad B_{ij} = \int_0^1 \chi_i(x)\chi_j(x)dx. \tag{15}$$

We see that the system matrix $(A - k^2B - ikC)$ is sparce, its entries are mostly zero. Furthermore, it is also *tridiagonal*, which means only its main, lower and upperdiagonal are non-zero. For this type of systems, very efficient solvers are available (see [OSCAR's stuk]). After solving the above system, we plug the values u_i back into (12) and obtain our estimate $U(x)$ to $u(x)$.

1.4 Error estimates

For small h, the Galerkin finite element method as an error estimate of the form

$$\frac{\|u - U\|}{\|u\|} \leq Ch^2, \tag{16}$$

where the constant C is independent of h [][CITE fem-paper]. Although we see that the error goes to zero as h decreaces (i.e. N increases), it is unclear what h is required for a certain level of accuracy. Furthermore, th C in (16) can depend multiple factors. For instance, it could depend on the forcing term f, the exact solution u or the wavenumber k. The last turns out to be problematic in multiple ways.

A higher wavenumber means higher oscillatory behavior of the exact solution. Since the wavelength λ is defined as $\frac{2\pi}{k}$, the wavelength decreases as k incleases. To keep the number of gridpoints per wavelength constant and the approximation of u accurate, the stepsize h has to decrease as k. This leads to bigger systems that have to be solved and a great deal of additional computation costs.

Another problem that accompanies higher wavenumbers is the introduction of pollution errors into the estimate U. Erros caused by the wavelength not being modelled accurately accumilate and reduce the accuracy of our estimation. It was proven [CITE] that the relative error satisfies

$$\frac{\|u - U\|}{\|u\|} \leq C_1 kh + C_2 k^3 h^2, \tag{17}$$

where constants C_1, C_2 are independent of both k and h. We see that, to keep this bound constant, we have to fix k^3h^2. As a result, our linear system rappidly increases in size as k increases.