Introduction to the Finite Element Method

James R. Nagel
Department of Electrical and Computer Engineering
University of Utah, Salt Lake City, Utah
April 4, 2012

1 Introduction

The finite element method (FEM) was originally developed to solve problems related to mechanical engineering in such fields as fluid dynamics and structural analysis. However, it was not long before FEM began to find uses in electromagnetics. Since then, FEM has become an essential tool for simulating complex geometries in electrical devices.

In many respects, FEM is very similar to the finite difference method (FDM). Both methods can generally be used to solve the same physical problems, and both methods eventually lead to the inversion of a matrix-vector equation with the form $\mathbf{Ax} = \mathbf{b}$. However, the advantage of FEM over FDM stems from its ability to sample at arbitrary locations in space rather than adhere to some fixed rectangular grid. This is made possible by the way in which FEM attempts to solve physical problems. While FDM simply applies a direct numerical approximation to the derivatives of partial differential equations, FEM attempts to minimize the total energy contained within a system of discrete volumes, or *elements*. However, if one were to apply FEM along a the same grid locations as those typically used in FDM, the ultimate matrix equations are often times equivalent. It is therefore useful to think of FEM as a more flexible, generalized form of FDM. The trade-off, of course, is that the added complexity with FEM generally makes implementation more difficult.

This tutorial seeks to introduce FEM through a simple, one-dimensional solution to the Laplace equation. Doing so helps to focus strictly on the core principles of FEM without getting bogged down in the excessive book-keeping that tends to occur in higher dimensions.

2 Elements and Shape Functions in 1D

Since the discrete volume element is the core unit of FEM, we begin our discussion by considering two samples of voltage potential, V_1 and V_2, located at the points x_1 and x_2 as illustrated in Figure 1. Whereas FDM tends to deal strictly with discrete samples in space, FEM seeks to interpolate all values of the voltage potential between the samples. For example, one of the most common interpolation methods is called *linear interpolation*, and is indicated by the straght line connecting V_1 and V_2. One could theoretically also resort to various high-order polynomial interpolations for greater accuracy, but we shall not consider such cases here.

To accomplish linear interpolation between the samples, we seek to express the voltage potential inside the element as

$$V_e(x) = V_1\,\alpha_1(x) + V_2\,\alpha_2(x)\ , \qquad x \in [x_1, x_2]\ , \tag{1}$$

where α_1 and α_2 are called *shape functions*. Note how we deliberately expressed this in a form where the inerpolation is achieved by simply weighing each shape function with a corresponding voltage sample.

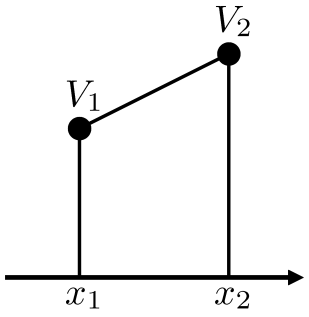

Figure 1: A simple one-dimensional volume element, represented by a linear interpolation between two discrete samples in the voltage potential function $V(x)$.

In order to derive the shape functions, we need to compare them against the general equation for the line connecting V_1 and V_2. Remembering the general form $y = mx + b$, this leads us to

$$\begin{aligned} V_1\,\alpha_1(x) + V_2\,\alpha_2(x) &= \left(\frac{V_2 - V_1}{x_2 - x_1}\right) x + \frac{V_1 x_2 - V_2 x_1}{x_2 - x_1} \\ &= \left(\frac{x_2 - x}{x_2 - x_1}\right) V_1 + \left(\frac{x - x_1}{x_2 - x_1}\right) V_2 \ . \end{aligned} \tag{2}$$

Thus, the 1D shape functions are found to be

$$\alpha_1(x) = \left(\frac{x_2 - x}{x_2 - x_1}\right) \ , \tag{3}$$

$$\alpha_2(x) = \left(\frac{x - x_1}{x_2 - x_1}\right) \ . \tag{4}$$

If we instead consider an arbitrary set of samples in x, then the shape functions for the nth element between the samples x_n and x_{n+1} satisfy

$$\alpha_1^n(x) = \left(\frac{x_{n+1} - x}{x_{n+1} - x_n}\right) = \frac{x_{n+1} - x}{h_n} \ , \tag{5}$$

$$\alpha_2^n(x) = \left(\frac{x - x_n}{x_{n+1} - x_n}\right) = \frac{x - x_n}{h_n} \ . \tag{6}$$

3 Element Energy

The next step is to find an expression for the total energy contained within the element. For a static electric field, this is given by

$$W = \frac{1}{2} \int_{x_1}^{x_2} \epsilon_0 |\mathbf{E}(x)|^2 \, dx \ , \tag{7}$$

where W has units of $\mathrm{J/m^2}$ in one dimension. Because we have assumed linear interpolation between each voltage sample, we have effectively enforced the condition that $\mathbf{E}$ is a constant value inside all

elements. The total energy of the the given element is therefore

$$W = \frac{1}{2} h \epsilon_0 |\mathbf{E}|^2 \ . \tag{8}$$

The next step is to rewrite the electic field in terms of the voltage samples which define the element. This is a straightforward process that produces

$$\begin{aligned} |\mathbf{E}|^2 &= \left| \frac{\partial V}{\partial x} \right|^2 \\ &= \left[\alpha_1'(x) V_1 + \alpha_2'(x) V_2 \right]^2 \\ &= \left[-\left(\frac{1}{h}\right) V_1 + \left(\frac{1}{h}\right) V_2 \right]^2 \\ &= \frac{1}{h^2} \left(V_1^2 - 2 V_1 V_2 + V_2^2 \right) \ . \end{aligned}$$

Although it may not seem immediately obvious at first, we can rewrite the above expression into a matrix-vector equation with the form

$$|\mathbf{E}|^2 = \left(\frac{1}{h}\right)^2 \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \ . \tag{9}$$

Let us now define the matrix-vector parameters using

$$\mathbf{v} = \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T \tag{10}$$

$$\mathbf{C} = \left(\frac{1}{h}\right) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ . \tag{11}$$

The total energy in the element is now written as

$$W = \frac{1}{2} \epsilon_0 \, \mathbf{v}^T \mathbf{C} \mathbf{v} \ . \tag{12}$$

The matrix $\mathbf{C}$ is traditionally called the *stiffness matrix* due to its original use in mechanical engineering applications. Another common name is the *element coefficient matrix*. If we instead generalize the energy expression for the nth arbitrary element in the system, we may simply write the result as

$$W_n = \frac{1}{2} \epsilon_0 \, \mathbf{v}_n^T \mathbf{C}_n \mathbf{v}_n \ , \tag{13}$$

where

$$\mathbf{v}_n = \begin{bmatrix} V_n & V_{n+1} \end{bmatrix}^T \ , \qquad \text{and} \tag{14}$$

$$\mathbf{C}_n = \left(\frac{1}{h_n}\right) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \ . \tag{15}$$

4 Global Assembly

Equipped now with a general expression for the total energy contained within the nth element, let us consider the total energy contained within a series of elements sampled arbitrarily along the

interval $x \in [a, b]$. This is naturally just the summation of all the energies contained within the individual elements:

$$W = \sum_{n=1}^{N} W_n \ . \tag{16}$$

Writing this out more explicitly, we see that

$$W = \frac{1}{2}\epsilon_0 \Big[\mathbf{v}_1^T \mathbf{C}_1 \mathbf{v}_1 + \mathbf{v}_2^T \mathbf{C}_2 \mathbf{v}_2 + \cdots + \mathbf{v}_N^T \mathbf{C}_N \mathbf{v}_N \Big] \ . \tag{17}$$

The next step is somewhat counterintuitive, but it allows us to combine all of the individual summation terms into a single matrix-vector expression. We begin by rewriting the elemental voltage vectors as M-dimensional vectors with only two nonzero elements, where $M = N + 1$ (remember that for N elements in 1D there will be $N + 1$ nodes). For example, the first two element vectors can be written as

$$\mathbf{v}_1 = \begin{bmatrix} V_1 & V_2 & 0 & 0 & \cdots & 0 \end{bmatrix}^T \ , \tag{18}$$

$$\mathbf{v}_2 = \begin{bmatrix} 0 & V_2 & V_3 & 0 & \cdots & 0 \end{bmatrix}^T \ . \tag{19}$$

By analogy, we can also write the element coefficient matrices as $M \times M$ matrices with only four nonzero elements. As an example, $\mathbf{C}_1$ and $\mathbf{C}_2$ would be written as

$$\mathbf{C}_1 = \left(\frac{1}{h_1}\right) \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \ , \tag{20}$$

and

$$\mathbf{C}_2 = \left(\frac{1}{h_2}\right) \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \ . \tag{21}$$

Notice how incrementing in n simply pushes the block of nonzero elements down along the diagonal while the scalar constant in front reflects the spacing between samples. Also notice how writing out the matrices in this way does nothing to change the ultimate scalar value obtained by the elemental energy $W_n = \frac{1}{2}\epsilon_0 \, \mathbf{v}_n^T \mathbf{C}_n \mathbf{v}_n$. We may therefore rewrite Equation (17) as a single matrix-vector operation given by

$$W = \frac{1}{2}\epsilon_0 \, \mathbf{v}^T \mathbf{C} \mathbf{v} \ , \tag{22}$$

where

$$\mathbf{v} = \begin{bmatrix} V_1 & V_2 & V_3 & \cdots & V_N & V_{N+1} \end{bmatrix}^T \tag{23}$$

and

$$\mathbf{C} = \mathbf{C}_1 + \mathbf{C}_2 + \mathbf{C}_3 + \cdots + \mathbf{C}_N + \mathbf{C}_{N+1} \ . \tag{24}$$

The matrix $\mathbf{C}$ is now called the *global coefficient matrix* because it accounts for the entire energy of the system. Writing it out explicitly for the 1D Laplace equation, it is easy to see that

$$\mathbf{C} = \begin{bmatrix} u_1 & -u_1 & 0 & 0 & \cdots & 0 & 0 \\ -u_1 & (u_1+u_2) & -u_2 & 0 & \cdots & 0 & 0 \\ 0 & -u_2 & (u_2+u_3) & -u_3 & \cdots & 0 & 0 \\ 0 & 0 & -u_3 & (u_3+u_4) & \cdots & 0 & 0 \\ & \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & (u_{N-1}+u_N) & -u_N \\ 0 & 0 & 0 & 0 & \cdots & -u_N & u_N \end{bmatrix} , \tag{25}$$

where $u_n = 1/h_n$ is the inverse of the distance between samples.

5 Global Solution

Once we have assembled the global coefficient matrix, our next goal is to *minimize the total energy of the system*. The justification for this step arises from the fact that the solution to Laplace's equation is also the very same solution that happens minimizes the total energy contained within the static electric fields. Since $\mathbf{C}$ is a symmetric matrix (i.e., $\mathbf{C}^T = \mathbf{C}$), we may take advantage of a matrix-vector property which states that

$$\frac{d}{d\mathbf{v}}\mathbf{v}^T\mathbf{C}\mathbf{v} = 2\mathbf{C}\mathbf{v} , \tag{26}$$

which is essentially just the power rule for differentiation when acting on a matrix-vector expression. Setting the derivative to zero therefore leads us to

$$\mathbf{C}\mathbf{v} = \mathbf{0} . \tag{27}$$

Note how the solution to this equation is just the trivial solution $\mathbf{v} = 0$, since this obviously minimizes energy. The reason for this is because we did not specify any sources or boundary conditions on $\mathbf{v}$.

If we instead assume that $\mathbf{v}$ is a combination of both *free* nodes $\mathbf{v}_f$ and *prescribed* nodes $\mathbf{v}_p$, we can rewrite the global coefficient matrix as a block matrix with a mixture of fixed and prescribed segments. The total system energy is therefore

$$W = \frac{1}{2}\epsilon_0 \begin{bmatrix} \mathbf{v}_f & \mathbf{v}_p \end{bmatrix} \begin{bmatrix} \mathbf{C}_{ff} & \mathbf{C}_{fp} \\ \mathbf{C}_{pf} & \mathbf{C}_{pp} \end{bmatrix} \begin{bmatrix} \mathbf{v}_f \\ \mathbf{v}_p \end{bmatrix} . \tag{28}$$

Minimizing the energy with respect to the free (i.e., variable) nodes $\mathbf{v}_f$ then gives us

$$\mathbf{C}_{ff}\mathbf{v}_f + \mathbf{C}_{fp}\mathbf{v}_p = 0 , \tag{29}$$

or

$$\mathbf{C}_{ff}\mathbf{v}_f = -\mathbf{C}_{fp}\mathbf{v}_p . \tag{30}$$

The solution for the free voltage samples along entire domain is therefore

$$\mathbf{v}_f = -\mathbf{C}_{ff}^{-1}\mathbf{C}_{fp}\mathbf{v}_p , \tag{31}$$

which we can see is no longer trivial.