Diffusion Equation with Finite Element Method
TF4062

Iwan Prasetyo
Fadjar Fathurrahman

1 Steady heat equation

As a model problem, we will consider determining the conduction of heat on a slender homogeneous metal wire of length L with a constant cross section. Assume that the left end is exposed to a prescribed heat flux, q, the right end is held at a constant temperature, $T = T_L$, and the length of the rod is surrounded by insulating material. The situation is shown in Figure 1.

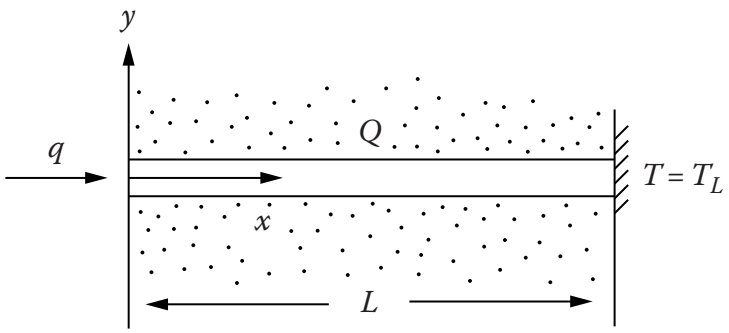

Figure 1: Conduction of heat in a rod of length L (from [1])

The differential equation that governs the distribution of temperature across the rod can be written as

$$-k\frac{\mathrm{d}^2T}{\mathrm{d}x^2} = Q \tag{1}$$

for $0 < x < L$, where k is the thermal conductivity of the material which is assumed constant and Q is the internal heat source. The boundary conditions for this problem are

$$-k\frac{\mathrm{d}T}{\mathrm{d}x} = q \quad \text{at } x = 0 \tag{2}$$

and

$$T = T_L \quad \text{at } x = L \tag{3}$$

The analytical solution to this problem can be expressed as:

$$T(x) = T_L + \frac{q}{k}(L - x) + \frac{1}{k}\int_x^L \left(\int_0^y Q(z)\,\mathrm{d}z \right) \mathrm{d}y \tag{4}$$

For constant Q, this equation reduces to

$$T(x) = T_L + \frac{q}{k}(L - x) + \frac{Q}{2k}(L^2 - x^2) \tag{5}$$

1.1 Discretization and approximation

The FEM procedure start by discretizing the spatial domain into elements e_i where:

$$e_i : \{x_i \leq x \leq x_{i+1}\}, \quad i = 1, 2, \ldots, n \tag{6}$$

The points x_i are also called *nodes*. If we have n elements then we have $n + 1$ nodes (or nodal points). The solution to the PDE can be approximated as:

$$T(x) \approx \sum_{i}^{n+1} T_i N_i(x) \tag{7}$$

where T_i are (unknown) nodal values of $T(x)$ and $N_i(x)$ are *shape functions* (also called *trial functions* or *basis functions*) associated with each node.

1.2 Linear shape functions

The shape functions are defined such that $N_i(x_i) = 1$ and $N_i(x_j) = 0$ if $j \neq i$. In FEM, they are usually chosen to be low order, piecewise polynomials. As an example, we will consider the case of one element, $n = 1$, or two nodes using the first order polynomial. The shape functions can be defined as:

$$\begin{cases} N_1(x) & = 1 - \dfrac{x}{L} \\ N_2(x) & = \dfrac{x}{L} \end{cases} \tag{8}$$

An illustration of this case is shows in Figure 2.

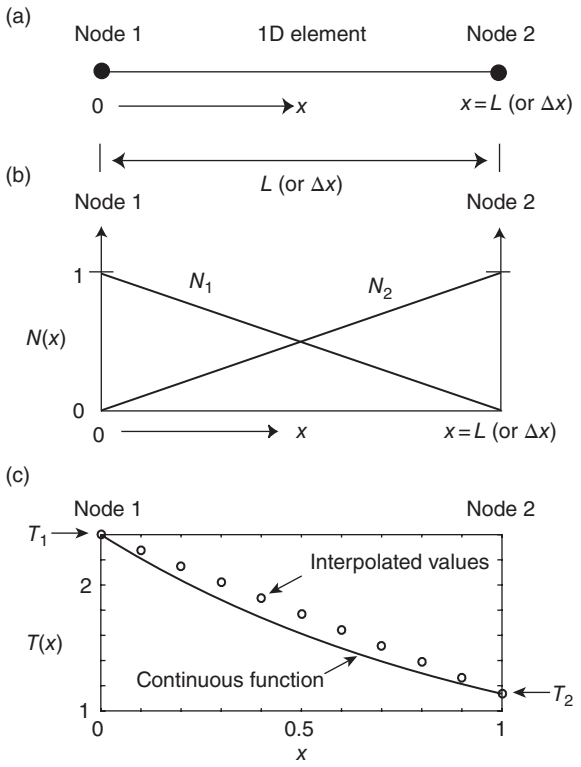

Figure 2: The case of one element with linear shape functions. (from [2])

For general case of linear elements we have the following shape functions. For left boundary points:

$$N_1(x) = \begin{cases} \dfrac{x_2 - x}{h_1} & x_1 \le x \le x_2 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

For interior points:

$$N_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{h_{i-1}} & x_{i-1} \le x \le x_i \\ \dfrac{x_{i+1} - x}{h_i} & x_i \le x \le x_{i+1}, \quad i = 2, 3, \ldots, n \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

And for right boundary points:

$$N_{n+1}(x) = \begin{cases} \dfrac{x - x_n}{h_n} & x_n \le x \le x_{n+1} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where h_i are the spacings between nodal points.

$$h_i = x_{i+1} - x_i \tag{12}$$

In FEM, the size or the shape of elements and thus their spacings do not have to be homogeneous, i.e. they can have arbitrary spacing in the case of 1d. This feature of FEM is more apparent in 2d and 3d case. It is one of its main strength: it can handle complex geometries by allowing arbitary choices of mesh.

1.3 Weighted residuals

When we approximate the solution as in the Equation (7), in general, we cannot obtain the true solution to the differential equation. So, we will not get an exact equality but some *residual* associated with the error in the approximation. This residual can be defined as:

$$R(T, x) \equiv -k\frac{\mathrm{d}^2T}{\mathrm{d}x^2} - Q \tag{13}$$

where T in the above equation is the approximation to the true solution T^* for which we have:

$$R(T^*, x) \equiv 0 \tag{14}$$

However, for any $T \neq T^*$, we cannot force the residual to vanish at every point x, no matter how small we make the grid or long the series expansion. The idea of the *weighted residuals method* is that we can multiply the residual by a *weighting function* and force the integral of the weighted expression to vanish:

$$\int_0^L w(x)R(T, x)\mathrm{d}x = 0 \tag{15}$$

where $w(x)$ is the weighting function. Choosing different weighting functions and replacing each of them in (15), we can then generate a system of linear equations in the unknown parameters T_i that will determine an approximation T of the form of the finite series given in Equation (7). This will satisfy the differential equation in an "average" or "integral" sense. The type of weighting function chosen depends on the type of weighted residual technique selected. In the *Galerkin procedure*, the weights are set equal to the shape functions N_i, that is

$$w_i(x) = N_i(x) \tag{16}$$

So we have:

$$\int_0^L N_i \left[-k\frac{\mathrm{d}^2T}{\mathrm{d}x^2} - Q\right] \mathrm{d}x = 0 \tag{17}$$

Since the temperature distribution must be a continuous function of x, the simplest way to approximate it would be to use piecewise polynomial interpolation over each element, in particular, piecewise linear approximation (by using linear shape functions as in (8)) provides the simplest approximation with a continuous function. Unfortunately, the first derivatives of such functions are not continuous at the elements' ends and, hence, second derivatives do not exist there; furthermore, the second derivative of T would vanish inside each element. However, to require the second-order derivatives to exist everywhere is too restrictive. We can weaken this requirement by application of integration by parts to the second derivative: ($\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u$):

$$\int_0^L N_i \left[-k\frac{\mathrm{d}^2T}{\mathrm{d}x^2}\right] \mathrm{d}x = \int_0^L k\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}T}{\mathrm{d}x}\mathrm{d}x - \left[kN_i\frac{\mathrm{d}T}{\mathrm{d}x}\right]_0^L \tag{18}$$

Substituting the Equation (7) to the integral term at the RHS of equation (18):

$$\int_0^L k\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{j=1}^{n+1} T_j N_j\right)\mathrm{d}x = \int_0^L k\frac{\mathrm{d}N_i}{\mathrm{d}x}\left(\sum_{j=1}^{n+1}\frac{\mathrm{d}N_j}{\mathrm{d}x}T_j\right)\mathrm{d}x \tag{19}$$

$$= \int_0^L k\left(\sum_{j=1}^{n+1}\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}N_j}{\mathrm{d}x}T_j\right)\mathrm{d}x \tag{20}$$

The Equation (18) can be rewritten as

$$\int_0^L k\left(\sum_{j=1}^{n+1}\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}N_j}{\mathrm{d}x}T_j\right)\mathrm{d}x - \int_0^L N_i Q\,\mathrm{d}x - \left[kN_i\frac{\mathrm{d}T}{\mathrm{d}x}\right]_0^L = 0 \tag{21}$$

for $i = 1, 2, \ldots, n+1$. The quantities k, Q, N_i are known. The terms containing T_i can be isolated to the LHS and the the known quantities can be moved to the RHS. This is a system of linear equations with T_i as the unknown variables. In matrix form:

$$\mathbf{Ku} = \mathbf{f} \tag{22}$$

where $\mathbf{K}$, usually called stiffness matrix, is a matrix arising from the integral terms, $\mathbf{f}$, usually called load vector, is column vector whose elements arise from the integrals of source term and boundary terms, and $\mathbf{u}$ is column vector of the the unknown T_i. These linear equations can be solved by standard method such as Gaussian elimination and LU decomposition.

In practice, the integral in Equation (21) can be calculated analytically or by numerical methods such as Gaussian quadrature. In the present case, the integral can be calculated analytically. The integral over all spatial domain can be divided into integrals over elements:

$$\int_0^L = \int_{x_1=0}^{x_2} + \int_{x_2}^{x_3} + \cdots + \int_{x_n}^{x_{n+1}=L} \tag{23}$$

For simplicity, let's consider two elements with equal spacing $h_1 = h_2$ with nodes

$$x_1 = 0, x_2 = L/2, x_3 = L \tag{24}$$

The first integral is done over the first element. In the first element, only N_1 and N_2 contributes to the integrals because N_3 is zero in this interval. So the index i and j are limited to 1 and 2 only. This also applies to other elements.

These integrals can be written as a matrix:

$$\mathbf{K}^{e_1} = \int_0^{L/2} k\begin{bmatrix}\frac{\mathrm{d}N_1}{\mathrm{d}x}\frac{\mathrm{d}N_1}{\mathrm{d}x} & \frac{\mathrm{d}N_1}{\mathrm{d}x}\frac{\mathrm{d}N_2}{\mathrm{d}x} \\ \frac{\mathrm{d}N_2}{\mathrm{d}x}\frac{\mathrm{d}N_1}{\mathrm{d}x} & \frac{\mathrm{d}N_2}{\mathrm{d}x}\frac{\mathrm{d}N_2}{\mathrm{d}x}\end{bmatrix}\mathrm{d}x \tag{25}$$

where $\mathbf{K}^{e_1}$ is called the local stiffness matrix.

The needed integrals are:

$$\begin{aligned}\int_0^{L/2} \frac{\mathrm{d}N_1}{\mathrm{d}x}\frac{\mathrm{d}N_1}{\mathrm{d}x}\,\mathrm{d}x &= \int_0^{L/2} \left(-\frac{1}{h_1}\right)\left(-\frac{1}{h_1}\right)\,\mathrm{d}x \\ &= \int_0^{L/2} \frac{1}{(L/2)^2}\,\mathrm{d}x \\ &= \left[\frac{4}{L^2}x\right]_0^{L/2} \\ &= \frac{2}{L}\end{aligned}$$

and

$$\begin{aligned}\int_0^{L/2} \frac{\mathrm{d}N_1}{\mathrm{d}x}\frac{\mathrm{d}N_2}{\mathrm{d}x}\,\mathrm{d}x &= \int_0^{L/2} \left(-\frac{1}{h_1}\right)\left(\frac{1}{h_1}\right)\,\mathrm{d}x \\ &= -\frac{2}{L}\end{aligned}$$

So the matrix elements can be written as

$$\mathbf{K}^{e_1} = \frac{k}{L/2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{26}$$

For more general case where $h_1 \neq h_2$, we obtain:

$$\mathbf{K}^{e_1} = \frac{k}{h_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{27}$$

Source term for the first element:

$$\mathbf{f}^{e_1,s} = \begin{bmatrix} f_1^{e_1,s} \\ f_2^{e_1,s} \end{bmatrix} \tag{28}$$

$$f_1^{e_1,s} = Q\int_0^{L/2} N_1(x)\,\mathrm{d}x = Q\int_0^{L/2}\left(1-\frac{2x}{L}\right)\,\mathrm{d}x = Q\left[x-\frac{1}{L}x^2\right]_0^{L/2} = Q\frac{L}{4} = Q\frac{h_1}{2} \tag{29}$$

$$f_2^{e_1,s} = Q\int_0^{L/2} N_2(x)\,\mathrm{d}x = Q\int_0^{L/2}\left(\frac{2x}{L}\right)\,\mathrm{d}x = Q\left[\frac{1}{L}x^2\right]_0^{L/2} = Q\frac{L}{4} = Q\frac{h_1}{2} \tag{30}$$

Boundary terms for the first element

$$\mathbf{f}^{e_1,b} = \begin{bmatrix} f_1^{e_1,b} \\ f_2^{e_1,b} \end{bmatrix} \tag{31}$$

By using boundary condition at $x = 0$ and using the fact that $N_1(x = 0) = 1$ and $N_1(x = L/2) = 0$:

$$f_1^{e_1,b} = \left[N_1(x)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\right]_0^{L/2} = N_1(L/2)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\Big|_{x=L/2} - N_1(0)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\Big|_{x=0} = q \tag{32}$$

Meanwhile using using the fact that $N_2(x=0)=0$ and $N_2(x=L/2)=1$:

$$f_2^{e_1,b} = \left[N_2(x)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\right]_0^{L/2} = N_2(L/2)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\Big|_{x=L/2} - N_2(0)\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\Big|_{x=0} = \left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)\Big|_{x=L/2} \tag{33}$$

We can obtain similar expressions for $\mathbf{K}^{e_2}$ and $\mathbf{f}^{e_2}$. These per-element matrices and vector must be used to build the global stiffness and load vector. This process is called matrix assembly process. An illustration of this is given in Figure 3.

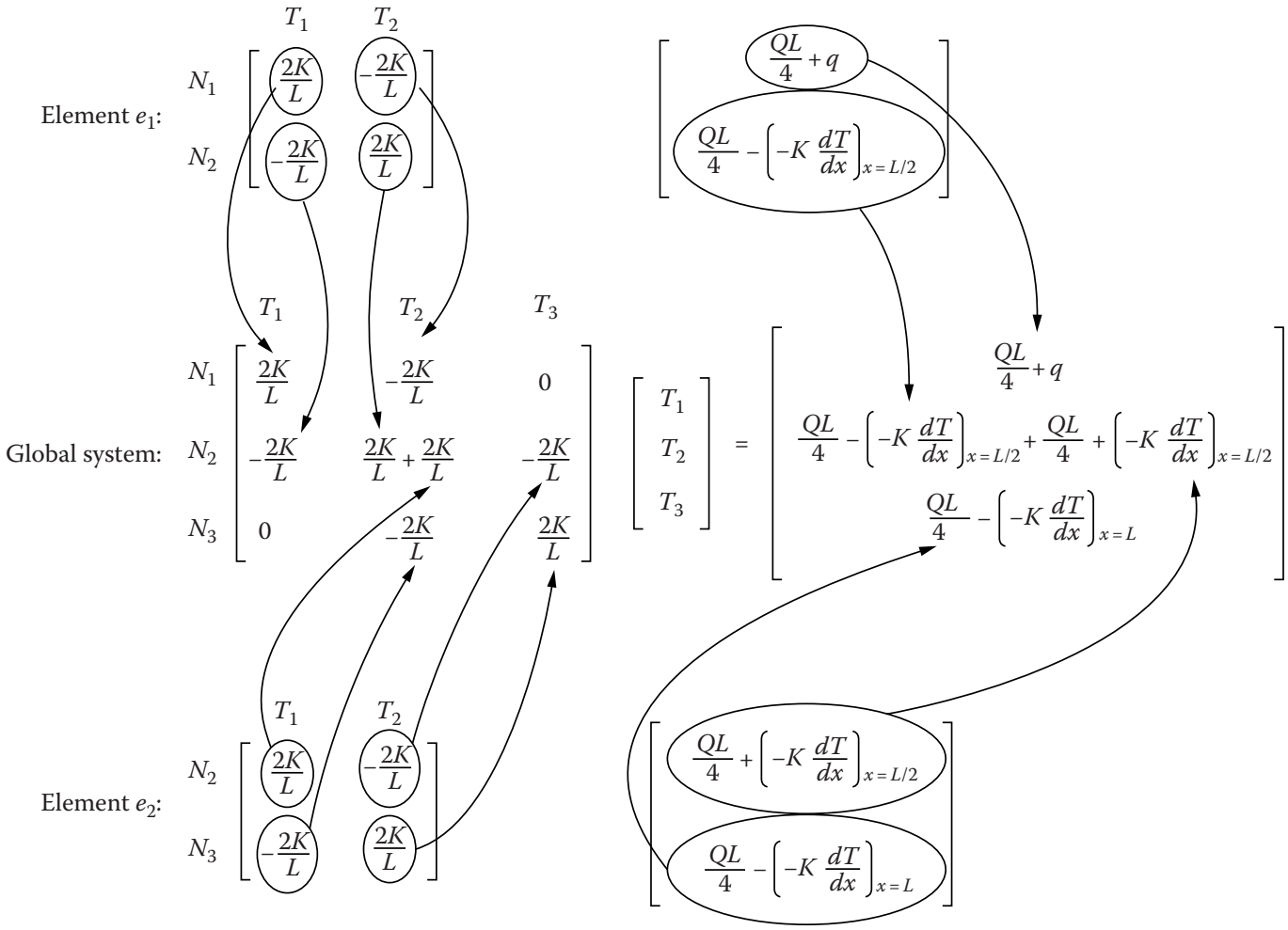

Figure 3: An illustration of matrix assembly for the case of 2 elements.

The global system of equations that we have to solve can be written as:

$$\frac{2k}{L}\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}\begin{bmatrix}T_1\\T_2\\T_3\end{bmatrix} = \frac{QL}{4}\begin{bmatrix}1\\2\\1\end{bmatrix} + \begin{bmatrix}q\\0\\\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)_{x=L}\end{bmatrix} \tag{34}$$

Because $T_3 = T_L$ is known from the boundary conditions, the third equation can be discarded. In the second equation, the term containing $T_3 = T_L$ can be moved to the RHS and the equations are rewritten as

$$\frac{2k}{L}\begin{bmatrix}1 & -1\\ -1 & 2\end{bmatrix}\begin{bmatrix}T_1\\T_2\end{bmatrix} = \frac{QL}{4}\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}q\\0\end{bmatrix} + \frac{2k}{L}T_L\begin{bmatrix}0\\1\end{bmatrix} \tag{35}$$

After T_1 and T_2 has been solved, we can use them to calculate the flux at the right boundary (where Dirichlet boundary condition is specified):

$$\left(k\frac{\mathrm{d}T}{\mathrm{d}x}\right)_{x=L} = \frac{2k}{L}\left(T_2 - T_L\right) + \frac{QL}{4} \tag{36}$$

1.4 Non-constant or variable source term

Now, we will consider the case where the source term Q is not a constant, but a function of spatial dimension: $Q(x)$. To handle this case, we assume that the source is continuous and we use the shape functions to expand the $Q(x)$:

$$Q(x) \approx \sum_{j=1}^{n+1} Q_j N_j(x) \tag{37}$$

The equation (21) becomes:

$$\int_0^L k\left(\sum_{j=1}^{n+1}\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}N_j}{\mathrm{d}x}T_j\right)\mathrm{d}x - \int_0^L N_i\sum_{j=1}^{n+1}Q_jN_j(x)\,\mathrm{d}x - \left[kN_i\frac{\mathrm{d}T}{\mathrm{d}x}\right]_0^L = 0 \tag{38}$$

where $Q_j = Q(x_j)$, i.e. the value of function $Q(x)$ evaluated at node point x_j. Applying this to the two element case, we need to evaluate the integrals which can be written in matrix form:

$$\mathbf{M}^{e_1} = \int_0^{L/2}\begin{bmatrix} N_1(x)N_1(x) & N_1(x)N_2(x) \\ N_2(x)N_1(x) & N_2(x)N_2(x)\end{bmatrix}\begin{bmatrix}Q_1 \\ Q_2\end{bmatrix}\mathrm{d}x \tag{39}$$

and similar expression for $\mathbf{M}^{e_2}$. In several literature this matrix is also known as (local) *mass matrix*. The result after carrying out the integrations is:

$$\mathbf{M}^{e_1} = \begin{bmatrix}\frac{L/2}{3} & \frac{L/2}{6} \\ \frac{L/2}{6} & \frac{L/2}{3}\end{bmatrix}\begin{bmatrix}Q_1 \\ \\ Q_2\end{bmatrix} \tag{40}$$

1.5 Non-constant or variable condution and convection boundary condition

In this case, k is not constant but a function of x, i.e. $k(x)$. Now, assuming $Q = 0$, the steady heat equation can be written as:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(k(x)\frac{\mathrm{d}T}{\mathrm{d}x}\right) = 0 \tag{41}$$

and convection boundary condition at $x = 0$:

$$-k_0\frac{\mathrm{d}T}{\mathrm{d}x} + h(T - T_\infty) \tag{42}$$

where $k_0 = k(x = 0)$, h is convective heat transfer coefficient and T_∞ is an external reference temperature. At $x = L$ we have $T = T_L$. The corresponding weak form is:

$$\int_0^L k(x)\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}T}{\mathrm{d}x} - \left[N_i\left(-K(x)\frac{\mathrm{d}T}{\mathrm{d}x}\right)\right]_{x=0} + \left[N_i\left(-K(x)\frac{\mathrm{d}T}{\mathrm{d}x}\right)\right]_{x=L} = 0 \tag{43}$$

Now, we may drop the boundary term at $x = L$ because the value of T is specified by the boundary condition and replace the boundary term at $x = 0$:

$$\int_0^L k(x)\frac{\mathrm{d}N_i}{\mathrm{d}x}\frac{\mathrm{d}T}{\mathrm{d}x} + N_i h(T - T_\infty)_{x=0} = 0 \tag{44}$$

Using similar expansion for $k(x)$ as we have done for $Q(x)$:

$$k(x) = \sum_{j=1}^{n+1} k_j N_j(x) \tag{45}$$

we can obtain similar linear system. For the present case, using two elements with same spacing between nodes, we have:

$$\frac{1}{L}\begin{bmatrix} K_1 + K_2 + hL & -(K_1 + K_2) & 0 \\ -(K_1 + K_2) & K_1 + 2K_2 + K_3 & -(K_2 + K_3) \\ 0 & -(K_2 + K_3) & K_2 + K_3 \end{bmatrix}\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} hT_\infty \\ 0 \\ 0 \end{bmatrix} \tag{46}$$

2 Time-dependent problem

Consider the unsteady heat PDE in one spatial dimension x:

$$\frac{\partial T}{\partial t} = k\frac{\partial^2 T}{\partial x^2} + Q \tag{47}$$

where k is thermal diffusivity constant and Q is a constant heat source. The initial conditions is

$$T(x, t = 0) = 0, \qquad \forall\, x \in [0, L] \tag{48}$$

and the boundary conditions are

$$T(x = 0, t) = 0 \qquad \text{and} \qquad T(x = L, t) = 0 \tag{49}$$

Following the general steps outlined for the time-independent case we can derive the weak form of this problem as:

$$\int_0^L N_i\frac{\partial T}{\partial t} = -k\left(N_i\frac{\partial T}{\partial x}\right)_{x=0} + k\left(N_i\frac{\partial T}{\partial x}\right)_{x=L} - k\frac{\partial N_i}{\partial x}\frac{\partial T}{\partial x}\,\mathrm{d}x + \int_0^L N_i\, Q\,\mathrm{d}x \tag{50}$$

In the case of specified temperatures at both ends, we can drop the boundary terms at the RHS of equation (50). Substituting the equation (7), we can arrive at the following matrix equation:

$$\mathbf{M}\frac{\partial}{\partial t}\mathbf{u} + \mathbf{K}\mathbf{u} = \mathbf{f} \tag{51}$$

Now, we need to discretize with respect to time variable t. One way to do this is via backward difference

operator[1]:

$$\mathbf{M}\frac{\mathbf{u}^{k+1}-\mathbf{u}^{k}}{\Delta t}+\mathbf{K}\mathbf{u}^{k+1}=\mathbf{f} \tag{52}$$

Rearranging, we obtain

$$\left(\frac{\mathbf{M}}{\Delta t}+\mathbf{K}\right)\mathbf{u}^{k+1}=\frac{\mathbf{M}}{\Delta t}\mathbf{u}^{k}+\mathbf{f} \tag{53}$$

Given $\mathbf{u}^0$, we can find $\mathbf{u}^1$ and at later times by solving the linear system (53).

References

[1] Darrel W. Pepper and Juan C. Heinrich. *The Finite Element Method: Basic Concepts and Applications with MATLAB, MAPLE, and COMSOL*. CRC Press, Boca Raton, 3rd edition, 2017.

[2] Guy Simpson. *Practical Finite Element Modeling in Earth Science using MATLAB*. Wiley Blackwell, Geneva, Switzerland, 2017.

[1]Here k denotes the index of time point.