## Finite Element Analysis of a 1D Convection-Diffusion Problem Formulation, Implementation, and Optimization

## Problem Statement

The governing PDE for the 1D Convection-Diffusion problem is given as:

$$
\begin{equation*}
-\frac{\partial}{\partial x}\left(D \frac{\partial C}{\partial x}\right)+u \frac{\partial C}{\partial x}=0 \quad \text { for } x \in(0, L) \tag{0}
\end{equation*}
$$

With the following boundary conditions.

$$
\begin{cases}C=0 \mathrm{~mol} / \mathrm{m}^{3} & \text { at } x=0  \tag{1}\\ C=1 \mathrm{~mol} / \mathrm{m}^{3} & \text { at } x=L\end{cases}
$$

The parameter values are given as $L=1 \mathrm{~m}, D=2 \mathrm{~m}^{2} / \mathrm{s}$, and $u=10 \mathrm{~m} / \mathrm{s}$.

## Galerkin Formulation

State the Weighed-Residual Form and then obtain the Weak Form via Integration by Parts.

$$
\begin{equation*}
\int_{x_{a}}^{x_{b}} w\left[-\frac{\partial}{\partial x}\left(D \frac{\partial C}{\partial x}\right)+u \frac{\partial C}{\partial x}\right] \mathrm{d} x=0 \Longrightarrow \int_{x_{a}}^{x_{b}}\left[D \frac{\partial w}{\partial x} \frac{\partial C}{\partial x}+u w \frac{\partial C}{\partial x}\right] \mathrm{d} x-\left[w D \frac{\partial C}{\partial x}\right]_{x_{a}}^{x_{b}}=0 \tag{2}
\end{equation*}
$$

The operation is carried out as follows $\left(\int w v^{\prime} d x=-\int v w^{\prime} d x+[w v]\right)$ with $(v=-D \partial C / \partial x)$.
In local coordinates, take $\left(x \longmapsto \bar{x}+x_{a}\right),\left(x_{a} \longmapsto 0\right)$, and $\left(x_{b} \longmapsto h_{e}\right)$. Here, $D$ and $u$ are constants and thus remain unchanged. Note that $\partial x=\partial \bar{x}$. Therefore, (2) becomes:

$$
\begin{equation*}
\int_{0}^{h_{e}}\left[D \frac{\partial w}{\partial \bar{x}} \frac{\partial C}{\partial \bar{x}}+u w \frac{\partial C}{\partial \bar{x}}\right] \mathrm{d} \bar{x}-\sum_{j=1}^{n} w\left(\bar{x}_{j}\right) Q_{j}=0 \tag{3}
\end{equation*}
$$

With $Q_{1}=(-D \partial C / \partial \bar{x})_{x_{1}^{e}}, Q_{i}=(D \partial C / \partial \bar{x})_{x_{i}^{e-}}-(-D \partial C / \partial \bar{x})_{x_{i}^{e+}}$, and $Q_{n}=(D \partial C / \partial \bar{x})_{x_{n}^{e}}$.
Using Galerkin, apply $w=\psi_{i}^{e}$, and $C=\sum_{j=1}^{n} C_{j} \psi_{j}^{e}$ to obtain $\sum_{j=1}^{n} k_{i j}^{e} C_{j}-Q_{i}=0$, with

$$
\begin{equation*}
k_{i j}^{e}=\int_{0}^{h_{e}}\left[D \frac{\partial \psi_{i}^{e}}{\partial \bar{x}} \frac{\partial \psi_{j}^{e}}{\partial \bar{x}}+u \psi_{i}^{e} \frac{\partial \psi_{j}^{e}}{\partial \bar{x}}\right] \mathrm{d} \bar{x} \tag{4}
\end{equation*}
$$

The mesh is composed of $N$ equally-sized centered quadratic elements. Thus $h_{e}=h=L / N$ and the shape functions are:

$$
\psi_{1}^{e}=\psi_{1}=\frac{2}{h^{2}} \bar{x}^{2}-\frac{3}{h} \bar{x}+1 \quad \psi_{2}^{e}=\psi_{2}=-\frac{4}{h^{2}} \bar{x}^{2}+\frac{4}{h} \bar{x} \quad \psi_{3}^{e}=\psi_{3}=\frac{2}{h^{2}} \bar{x}^{2}-\frac{1}{h} \bar{x}
$$

Note that equation (4) is now independent of element number. In other words, $k_{i j}^{e}=k_{i j}$, for all $e \in\{1, \ldots, N\}$.

## Gauss-Legendre Quadrature

An $n$-point Gauss-Legendre quadrature is exact for polynomials of degree $(2 n-1)$. ${ }^{[1]}$
Notice that the integrand $g(\bar{x})=D\left(\partial \psi_{i} / \partial \bar{x}\right)\left(\partial \psi_{j} / \partial \bar{x}\right)+u \psi_{i}\left(\partial \psi_{j} / \partial \bar{x}\right)$ in (4) is cubic for all elements. Thus, the degree of exactness $(2 n-1=3)$ requires only $(n=2)$ Gauss points.

The Gauss point abscissa and weights are $\bar{x}_{i}=\{1 / \sqrt{3},-1 / \sqrt{3}\}$ and $w_{i}=\{1,1\}$. The exact result is obtained:

$$
\begin{equation*}
\int_{a}^{b} g(\bar{x}) \mathrm{d} \bar{x}=\frac{b-a}{2} \sum_{k=1}^{n=2} w_{i} g\left(\frac{b-a}{2} \bar{x}_{i}+\frac{a+b}{2}\right) \tag{5}
\end{equation*}
$$

With $a=0$ and $b=h$, this simplifies to:

$$
\begin{equation*}
\int_{0}^{h} g(\bar{x}) \mathrm{d} \bar{x}=\frac{h}{2}\left[g\left(\frac{h}{2}\left(1+\frac{1}{\sqrt{3}}\right)\right)+g\left(\frac{h}{2}\left(1-\frac{1}{\sqrt{3}}\right)\right)\right] \tag{6}
\end{equation*}
$$

## Mesh and Node Numbering

The node numbering of the $N$ quadratic elements is carried out as follows:


The connectivity matrix for this mesh is thus obtained.

$$
B=\left[\begin{array}{ccc}
1 & N+2 & 2  \tag{7}\\
\vdots & \vdots & \vdots \\
i & N+i+1 & i+1 \\
\vdots & \vdots & \vdots \\
N & 2 N+1 & N+1
\end{array}\right]
$$

From this particular connectivity matrix emerges a noticeable pattern in the global coefficient matrix. For example, with $N=2$ elements, the coefficient matrix is:

$$
K=\left[\begin{array}{ccccc}
k_{11}^{1} & k_{13}^{1} & & k_{12}^{1} & \\
k_{31}^{1} & k_{33}^{1}+k_{11}^{2} & k_{13}^{2} & k_{32}^{1} & k_{12}^{2} \\
& k_{31}^{2} & k_{33}^{2} & & k_{32}^{2} \\
k_{21}^{1} & k_{23}^{1} & & k_{22}^{1} & \\
& k_{21}^{2} & k_{23}^{2} & & k_{22}^{2}
\end{array}\right]
$$

Note that $k^{e}=k$ for all $e$ and thus the superscripts can be omitted.
There is an assembly pattern that can be extracted from the connectivity matrix.

[^0]

## Assembly Procedure

More generally, for any $N$ elements, the $(2 N+1) \times(2 N+1)$ global coefficient matrix can be proven inductively to be of the form:

$$
K=\left[\begin{array}{ccccccccc}
k_{11} & k_{13} & & & & k_{12} & & &  \tag{8}\\
k_{31} & k_{11}+k_{33} & \ddots & & & k_{32} & \ddots & & \\
& \ddots & \ddots & \ddots & & & \ddots & \ddots & \\
& & \ddots & k_{11}+k_{33} & k_{13} & & & \ddots & k_{12} \\
& & & k_{31} & k_{33} & & & & k_{32} \\
k_{21} & k_{23} & & & & k_{22} & & & \\
& \ddots & \ddots & & & & \ddots & & \\
& & \ddots & \ddots & & & & \ddots & \\
& & & k_{21} & k_{23} & & & & k_{22}
\end{array}\right]
$$

The following assembly rules hold:

$$
\begin{gathered}
K_{i, i}=\left\{\begin{array}{cl}
k_{11} & \text { for } i=1 \\
k_{11}+k_{33} & \text { for } i \in\{2, \ldots, N\} \\
k_{33} & \text { for } i=N+1 \\
k_{22} & \text { for } i \in\{N+2, \ldots, 2 N+1\}
\end{array}\right. \\
K_{i, i+1}=\left\{\begin{array}{cl}
k_{13} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array} \quad K_{i+1, i}=\left\{\begin{array}{cl}
k_{31} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array}\right.\right. \\
K_{i, N+i+1}=\left\{\begin{array}{cl}
k_{12} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array} \quad K_{N+i+1, i}=\left\{\begin{array}{cl}
k_{21} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array}\right.\right. \\
K_{i+1, N+i+1}=\left\{\begin{array}{cl}
k_{32} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array} \quad K_{N+i+1, i+1}=\left\{\begin{array}{cl}
k_{23} & \text { for } i \in\{1, \ldots, N\} \\
0 & \text { for } i \in\{N+1, \ldots, 2 N+1\}
\end{array}\right.\right.
\end{gathered}
$$

Thus, in the MATLAB implementation, rather than using the connectivity matrix directly:

```
>> for k = 1:N for I = 1:3 for J = 1:3
    K(B(k,I),B(k,J)) = K(B(k,I),B(k,J)) + ke(I,J);
    end end end;
```

It is possible to exploit the derived sparsity pattern of $K$ for several advantages. First, triplet storage format can be easily implemented using the built-in MATLAB sparse( $i, j, v, m, n$ ) function which greatly enhances performance scaling by eliminating dynamic allocation. Second, avoiding the nested for-loops reduces computational time complexity. Note that, the connectivity matrix is still used during assembly, just indirectly.

## Convergence Analysis

Using Main.m obtain the FEM solution for increasing element numbers $N$. Overlay the result plots in a single graphic to clearly verify the convergence of the FEM model. The model converges at around $N=10$, which will be adopted for further analysis.


## Solution Validation

Three solutions will be obtained and compared: (1) MATLAB finite element solution, (2) COMSOL finite element solution, and (3) the exact solution $C_{\mathrm{ex}}(x)=\frac{e^{5 x}-1}{e^{5}-1}$.

Obtain the exact solution of the problem using the following query in Mathematica 11.0.
DSolve $[\{-2 D[C[x],\{x, 2\}]+1 \theta D[C[x],\{x, 1\}]==\theta, C[\theta]==\theta, C[1]==1\}, C[x], x]$

COMSOL 5.1 is used to generate the solution in FEM_Project.mph to be compared with the MATLAB solution given by the Main.m script. For comparative purposes, use $N=10$ quadratic elements (equivalent to Main(10) in MATLAB and to he $=0.05$ in COMSOL).

All numerical results are listed in the following table.


| $\boldsymbol{x}$ | $\mathrm{C}_{\text {EXACt }}$ | $\mathrm{C}_{\text {COMSOL }}$ | $\mathrm{C}_{\text {MATLAB }}$ | $\boldsymbol{x}$ | $\mathrm{C}_{\text {EXACt }}$ | $\mathrm{C}_{\text {COMSOL }}$ | $\mathrm{C}_{\text {MATLAB }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0.00 | 0.0000000 | 0.0000000 | 0.0000000 | 0.55 | 0.0993306 | 0.0993317 | 0.0993313 |
| 0.05 | 0.0019267 | 0.0019268 | 0.0019259 | 0.60 | 0.1294700 | 0.1294710 | 0.1294911 |
| 0.10 | 0.0044007 | 0.0044008 | 0.0044022 | 0.65 | 0.1681691 | 0.1681706 | 0.1681645 |
| 0.15 | 0.0075773 | 0.0075775 | 0.0075774 | 0.70 | 0.2178601 | 0.2178618 | 0.2178875 |
| 0.20 | 0.0116562 | 0.0116565 | 0.0116598 | 0.75 | 0.2816647 | 0.2816665 | 0.2816464 |
| 0.25 | 0.0168936 | 0.0168939 | 0.0168946 | 0.80 | 0.3635914 | 0.3635933 | 0.3636221 |
| 0.30 | 0.0236186 | 0.0236190 | 0.0236250 | 0.85 | 0.4687873 | 0.4687891 | 0.4687381 |
| 0.35 | 0.0322536 | 0.0322541 | 0.0322554 | 0.90 | 0.6038615 | 0.6038631 | 0.6038872 |
| 0.40 | 0.0433412 | 0.0433418 | 0.0433515 | 0.95 | 0.7773002 | 0.7773013 | 0.7771866 |
| 0.45 | 0.0575779 | 0.0575787 | 0.0575800 | 1.00 | 1.0000000 | 1.0000000 | 1.0000000 |
| 0.50 | 0.0758582 | 0.0758591 | 0.0758736 | c | C | C | C |

Define a function to quantify the error between solution vectors using the 2 -norm:
$\operatorname{Rel} \operatorname{Err}\left(\mathbf{v}_{\mathbf{r e f}}, \mathbf{v}\right)=\left\|\mathbf{v}_{\mathbf{r e f}}-\mathbf{v}\right\|_{2} /\left\|\mathbf{v}_{\text {ref }}\right\|_{2}$ and obtain the following relative errors:

$$
\operatorname{RelErr}\left\{\left(\mathbf{C}_{\mathbf{E X} .}, \mathbf{C}_{\mathbf{C O M} .}\right) \quad\left(\mathbf{C}_{\mathbf{E X} .}, \mathbf{C}_{\text {MAT } .}\right) \quad\left(\mathbf{C}_{\mathbf{C O M} .}, \mathbf{C}_{\text {MAT }} .\right)\right\}=10^{-5} \cdot\left\{\begin{array}{llll}
0.3067 & 8.666 & 8.698
\end{array}\right\}
$$

Using quadratic interpolation order between the nodal values, it is possible to overlay the three solutions and notice that they satisfyingly overlap in the plot below. The COMSOL and MATLAB solutions are deemed valid when compared to the exact solution. More so, they are consistent amongst each other having a miniscule relative error.



[^0]:    ${ }^{[1]}$ Yew, A. Numerical integration: Gaussian quadrature rules. APMA 0160. Spring 2011. www.dam.brown.edu/people/alcyew/handouts/GLquad.pdf

